INTELLIGENCE **17**, 47–77 (1993)

Spearman's Hypothesis Tested With Chronometric Information-Processing Tasks

ARTHUR R. JENSEN
University of California, Berkeley

To account for the highly variable size of the mean difference between representative samples of the white and black populations on various psychometric tests of cognitive performance, Spearman (1927) suggested the hypothesis that the relative size of the mean white–black difference (in standardized scores) on various mental tests is a direct function of the tests' different loadings on the general factor, psychometric *g*, the highest order common factor in all complex tests of cognitive ability. This hypothesis has been borne out by numerous studies based on conventional psychometric tests. The present studies, based on large groups of elementary schoolchildren, extend the test of Spearman's hypothesis to performance on reaction time variables in simple information-processing tasks intended to minimize intellectual and cultural content. The variables' *g* loadings were estimated by their correlations with Raven's Matrices. Spearman's hypothesis was borne out significantly and at least as strongly as in previous studies based on conventional psychometric tests.

On most conventional tests of general mental ability, such as IQ, scholastic aptitude, armed services selection tests, and many personnel tests, representative samples of the white (W) and black (B) populations in the United States differ, on average, about 1 $\pm$ 0.2 standard deviations (Osborne & McGurk, 1982). But it is also an important fact that the size of the W–B difference varies considerably on various tests, ranging from differences close to zero on some tests up to differences greater than 1 standard deviation. This variation in the size of W–B mean differences has not been found to be systematically related to any of the surface features of psychometric tests, such as verbal or nonverbal, spatial or numerical, individual or group, paper-and-pencil or performance tests, specific knowledge content, or objective indices of test bias. It is incumbent on psychometricians to discover specifically which aspects of their diverse mental tests are systematically related to the variable size of the mean W–B difference.

The father of classical test theory, Charles Spearman, was probably the first psychometrician to notice this phenomenon and to suggest the aspect of mental tests most systematically related to the variable size of the W–B difference

Correspondence and requests for reprints should be sent to Arthur R. Jensen, School of Education, University of California, Berkeley, CA 94720.

(Spearman, 1927, p. 379). The crucial aspect surmised by Spearman is not a superficial feature of test type or information content, but a latent trait, namely general ability, or *g*, the highest common factor in all measures of cognitive ability. In reviewing an early study of W–B differences on a number of diverse tests, Spearman conjectured that the relative sizes of the W–B differences on the various tests are directly related to the tests' *g* loadings. Spearman himself never formalized this conjecture or tested it empirically, and it remained in obscurity for more than half a century.

In the last decade, however, at least 13 independent empirical tests of Spearman's hypothesis have been conducted in large representative samples of whites and blacks who were given anywhere from 6 to 25 different tests, totaling more than 70 different tests in the 13 studies (Jensen, 1985a, 1985b, 1987a; Jensen & Reynolds, 1982; Naglieri & Jensen, 1987). Thirty-six experts in relevant fields published commentaries on these studies in *Behavioral and Brain Sciences* ("Open Peer Commentary," 1985, 1987; also see Humphreys, 1985a, 1985b; Jensen, 1985c). There are other studies related to Spearman's hypothesis based on white and black samples that are not typical, or American, population samples (Jensen & Faulstich, 1988; Lynn & Holmshaw, 1990).

The overall conclusion from all these studies is that Spearman's hypothesis has been borne out significantly by every study (i.e., 13 of 13) and no appropriate data set has yet been found that contradicts Spearman's hypothesis. Hence, in the case of conventional psychometric tests, Spearman's hypothesis is really no longer an hypothesis but an empirical fact.

The studies here were intended to test whether Spearman's hypothesis can be generalized beyond the realm of typical psychometric tests to the chronometric assessment of speed and efficiency of information processes. Study 1 measures speed and consistency of information processing in three very simple elementary cognitive tasks (ECTs). Previous studies have shown that on even the most complex of the ECTs used in Study 1 (oddman-out paradigm), elementary schoolchildren have response times averaging less than 1 s. Study 2 measures speed and consistency of retrieval of previously acquired information, in this case, basic scholastic knowledge—addition, subtraction, and multiplication of single digit numbers—problems on which all the subjects demonstrated perfect prior knowledge on a nonspeeded paper-and-pencil test. Previous studies of these and similar chronometric tasks have found them to have some degree of *g* loading, albeit usually small, and their correlation with various conventional psychometric tests seems to be based entirely on their common *g* variance and not on any other common factors found in various psychometric tests such as the subtests of the Wechsler Intelligence Scales and the Armed Services Vocational Aptitude Battery (Vernon, 1983; Vernon & Jensen, 1984). Yet, the chronometric tests used in these studies bear virtually no resemblance to conventional psychometric tests. Their theoretical interest lies in the fact that these measures of speed and consistency of processing information apparently reflect certain elemental components of psychometric *g*.

Several previous studies have reported B–W differences on various chronometric tasks that have no intellectual content, but these data, although generally not inconsistent with the studies here, were not obtained or analyzed with any clear reference to testing a theory-derived hypothesis or broadening an empirical generalization, such as Spearman's hypothesis. Most of these previous chronometric studies involving race differences are reviewed in detail elsewhere (Jensen, 1988). One such study that clearly aimed to test a specific hypothesis concerning the information-processing components of racial difference in mental abilities (Borkowski & Krause, 1983) reported B–W differences in certain RT measures, but the results appeared puzzling and highly questionable because of too-small sample sizes, creating the extreme likelihood of Type II errors (Jensen, 1985d; Borkowski, Krause, & Maxwell, 1985). Moreover, the small samples of 29 white and 20 black children differed so atypically (1.55 *SD*) on Raven's Progressive Matrices as to raise doubts about their representativeness. More recently, Lynn and Holmshaw (1990) found significant differences between large samples of white British and black African schoolchildren on chronometric variables highly similar to those in Study 1 here; they also reported results consistent with Spearman's hypothesis.

The studies here were intended to test Spearman's hypothesis on chronometric variables that are exceedingly simple and easily within the capacity of all the participating subjects. The elementary tasks were selected to minimize intellectual content, yet vary slightly in information-processing demands and reliably reflect individual differences in information-processing speed and efficiency. Hence, if speed of information processing in ECTs is one of the basic sources of variance in Spearman's *g*, the chronometric variables derived from these ECTs should be expected to have some varying degree of relationship to the psychometric *g* factor that countless factor analyses have consistently shown to be most highly loaded in complex tests of reasoning ability. It is on such complex and highly *g*-loaded cognitive tests in the psychometric realm that the largest mean differences between representative samples of American white and black populations are generally observed.

These conditions make the chronometric variables derived from ECTs a useful vehicle for rigorously testing Spearman's hypothesis and to determine whether it is generalizable to the most elementary information processes that theoretically are involved in psychometric *g*.

Methodology of Testing Spearman's Hypothesis

The essential method for testing Spearman's hypothesis is conceptually simple, but there are certain methodological requirements for a bona fide test that need to be recognized.

The essential test of Spearman's hypothesis is the correlation between two sets of variables: (a) the column vector of *g* loadings of a set of mental test variables, and (b) the column vector of W–B mean differences on the same variables expressed in standard score form, that is, the difference (on each variable) between the group means (W–B) divided by the average standard deviation within groups. The standardized mean difference is often referred to as the *effect size* (ES). The variables' *g* loadings in the case of conventional psychometric tests represent the general factor of the tests themselves, derived from a factor analysis of the tests. In this study, based on a number of chronometric variables, the index of the variables' loadings on psychometric *g* is their correlations with a psychometric test that is well known to be highly loaded on Spearman's *g*. Raven's Standard Progressive Matrices (SPM), a nonverbal test of reasoning ability, was selected as the best test for this purpose; it usually is among the most highly *g*-loaded tests when factor analyzed among a variety of other psychometric tests.

A proper test of Spearman's hypothesis has seven essential methodological requirements:

1. The population samples being compared should not be selected on the basis of any highly *g*-loaded criteria that would markedly restrict the *g* variance in the samples.
2. The variables on which the group comparisons are made should have reliable variation in their *g* loadings.
3. The variables on which the group comparisons are made should measure the same latent traits in both groups. This can be determined from the similarity of the factor structure of the variables as indicated by the congruence coefficient.
4. The variables on which the groups are compared should also measure one and the same psychometric *g* factor to approximately the same degree for a given variable in both groups (or be correlated to about the same degree with a highly *g*-loaded test). The congruence coefficient shows the degree of similarity between factors obtained in two independent samples. Unless the congruence coefficient, r_c, indicates a high degree of similarity ($r_c > .95$) between the *g* vectors of the white and black groups, a proper test of Spearman's hypothesis cannot be rendered.
5. The *g* loadings of the variables (or their correlations with a highly *g*-loaded test) must be determined separately in each racial group, because, if the groups differ in *g*, the variables' *g* loadings will, to some extent, reflect the group difference, thereby artificially inflating the correlation between the vector of *g* loadings and the vector of W–B mean differences, and spuriously favoring Spearman's hypothesis. However, the vector of *g* loadings may be made more reliable by *averaging* each variable's *g* loadings that were obtained separately within each racial group. The group mean dif-

ference, of course, cannot affect these average g loadings in any way because the averaged loadings (or correlations) were obtained separately *within* each group.

6. Imperfect reliability of the variables would, of course, attenuate both the variables' g loadings and the standardized differences, or ESs, between the racial group means on the same variables. And the variables may well differ from one another in reliability. Therefore, a test of Spearman's hypothesis must rule out the possibility that the correlation between the vector of g loadings and the vector of ESs is solely attributable to the variables' differing reliability coefficients. Ideally, there would be no problem if all the variables had comparable reliability. Since this is not the case, however, control of variability in reliability coefficients can be accomplished in two distinctly different ways: (a) by partialing out estimates of reliability from the correlation between the vector of g loadings and the vector of EDs; and (b) by correcting the g loadings and ESs for attenuation. These two methods are not redundant, but control for two fundamentally different aspects of the possible effect of variable reliability on the test of Spearman's hypothesis.
7. The test of Spearman's hypothesis is the *correlation* between the vector of the variables' g loadings and the vector of the variables' ESs (i.e., standardized W–B differences). But some consideration has to be given to the type of correlation coefficient that is used for this purpose. The Pearson correlation (r) is desirable as a descriptive statistic in this case because it best indicates the degree of rectilinearity of the relationship between two variables. A test of the statistical significance of the observed Pearson r, however, is problematic because r is a parametric statistic that makes assumptions about the form of its sampling distribution, and we lack knowledge of the tenability of this parametric assumption as regards the two vectors being correlated. Therefore, Spearman's nonparametric rank-order correlation (r_s) is more appropriate. Its test of significance makes no assumptions about the scale properties of the variables or the population distribution of r_s. It is strictly a permutation test, indicating the percentile rank of a correlation as large as the observed r_s among all possible $n!$ permutations of the two sets of n variables each. In this study, both r and r_s are always reported, but the test of significance is applied only to r_s.

It should be noted that the test of Spearman's hypothesis is necessarily an extremely stringent one, because the degrees of freedom for the statistical rejection of the null hypothesis is based on the number of pairs of *variables* in the correlated vectors (this is always 12 in the present studies) and *not* on the subject sample size. Because of the limited number of variables, rejection of the null hypothesis at a traditionally acceptable level of confidence depends on quite large values of r_s. The virtue of having large subject samples, as here, is the desirable effect of large sample sizes in decreasing the standard errors of the main variables that enter into the test of the hypothesis, namely, the vector of *g* loadings and the vector of ESs. Sampling error can easily obscure the true rank order of the variables composing these vectors.

A Fallacious Criticism. Finally, a mistaken methodological criticism of the test of Spearman's hypothesis should be pointed out here, although the fallacy has already been clearly explicated elsewhere (Shockley, 1987). In the "Open Peer Commentary" (1985) of *Behavioral and Brain Sciences,* 5 of the 32 commentators argued that the method I proposed for testing Spearman's hypothesis must inevitably result in a purely artifactual or tautological confirmation of the hypothesis—the idea being that there is a necessary mathematical connection between tests' *g* loadings and the size of the mean differences between groups—*any* two groups that happen to differ to some degree on the various tests. The fallacy of this argument becomes immediately apparent when it is noted that in calculating correlation coefficients (and ipso facto, factor loadings), all information about groups' means (and standard deviations) is completely lost. Pearson *r* is always the mean of the cross-products of *standardized scores,* that is, *z* scores, with $M = 0$, $SD = 1$. Consequently, a variable's *g* loading (or its correlation with a *g*-loaded test) cannot have any relation to a group's mean on the variable (or any relation to a difference between group means when the variable's *g* loading is calculated separately *within* each group) for any purely mathematical reason. Under the stated conditions, a significant correlation between *g* loadings and mean group differences is an empirical fact, not a mathematical artifact.

STUDY 1
Method and Procedures

Subjects

Subjects were pupils in regular classes in Grades 4 through 6 (in approximately equal numbers from each grade) in several public elementary schools located in largely middle-class suburban neighborhoods. Urban, inner-city schools are not represented in this sample. All of the schools were racially integrated, with the majority of pupils of European ancestry. All of the pupils in intact classes who received parental permission were tested, but only those pupils for whom there were complete data ($N = 820$) are included in this study. These comprise 585 white children of European ancestry and 235 black children. In each group the sex ratio is within 1% of 50:50. The classes had much smaller numbers of Asian, Hispanic, and other ethnic groups who were also tested, but their data are not included in this study.

The mean age of the white sample is 10.93 years ($SD = 1.09$); the mean age of the black sample is 11.19 years ($SD = 1.02$). The age difference, with blacks averaging about 3 months older than whites, is significant, $t(818) = \pm 3.28, p < .001$. To eliminate whatever effect this small difference might have on the subsequent analyses, age was regressed out of all of the test data.

Elementary Cognitive Tasks (ECTs)

Three RT tasks were used to measure the speed of stimulus apprehension and discrimination. The apparatus consists of the subject's stimulus-response console, shown in Figure 1. The console is connected to an IBM-PC; the testing procedure was computer programmed and the RTs (measured in milliseconds) of the subject's responses and the number of erroneous responses were automatically recorded on diskettes. Three separate RT tests were administered with this

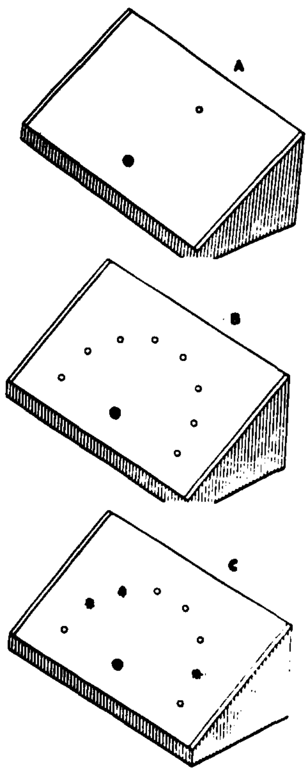

Figure 1. The subject's response console for (**A**) SRT, (**B**) CRT, and (**C**) Oddman-Out RT. The black dot in the lower center of each panel is the home button. The open circles, 15 cm from the home button, are green under-lighted push buttons. In the SRT and CRT conditions (i.e., **A** and **B**) only one green push button lights up on each trial; on the Oddman task, three push buttons light up simultaneously on each trial, with unequal distances between them (shown in **C**), the remotest one from the other two being the oddman-out, which the subject must touch.

apparatus: Simple, Choice, and Discrimination RT. Subjects were tested individually in a quiet room.

Simple Reaction Time (SRT). This task measures the speed of stimulus apprehension when the subject has no uncertainty of the nature and location of the stimulus (a single designated green light going "on," the only uncertainty being the exact moment (within 1–4 s after the "beep" warning signal) that the stimulus will occur. The subject sits in front of the response console, which appears as **A** in Figure 1. The lower push button (i.e., black dot in Figure 1A) is called the "home" button (1 in. [2.54 cm] diameter). After instructions by the tester, the subject begins the test by holding down the home button with the index finger of the preferred hand. In exactly 1 s a beep of 1-s duration sounds as a preparatory signal. Then, after a random interval of 1 to 4 s, a green stimulus light (¾ in. [1.905 cm] diameter and 6 in. [15.24 cm] above the home button), goes on. The subject's task is simply to turn off the light as quickly as possible by reaching up and touching it with his or her index finger. The subject's RT on a single trial is the interval (in ms) between the onset of the green stimulus light and the subject removing his or her finger from the home button. The next trial is initiated by the subject's depressing the home button. Hence the test is self-paced. Each trial takes an average of 6 s (including the intertrial intervals). There are 8 practice trials to familiarize the subject with the procedure, followed by 20 test trials.

Choice Reaction Time (CRT). This task involves the additional uncertainty as to which one of eight alternative lights will go on; hence, it measures speed of stimulus apprehension *plus* the time required to resolve the uncertainty as to which light or button will have to be pressed. In this task, a semicircle of eight green lights is exposed, shown as **B** in Figure 1. The procedure is exactly as before, except that in this task the subject is always uncertain as to which light will go on; the same light never goes on in two successive trials. Only one light goes on in a given trial, and the order of the lights going on is random from trial to trial. For all subjects, this uncertainty as to the location of the stimulus usually has the effect of *increasing* the RT and possibly also of slightly increasing the correlation between RT and IQ, or g. Note that the response requirements involved in CRT are identical to those in the SRT procedure; when any light goes on, the subject simply reaches up and touches it to turn it off as quickly as possible. Subjects are told that this is a test of their speed of reaction and that they should turn out the light as fast as they can. There are 8 practice trials and 30 test trials.

Discrimination Reaction Time (Oddman RT). This task involves stimulus apprehension and the resolution of uncertainty, as in SRT and CRT, *plus* the time required for *discrimination* of the oddman button, that is, out of the total array of eight buttons, the one lighted button that is the farthest from the other two lighted buttons. The "oddman-out" procedure was introduced as a new RT paradigm by Frearson and Eysenck (1986). Instead of only one of the eight green lights going on, a set of three lights (among the 8) goes on simultaneously, as shown in **C** of Figure 1. The three lights are always separated by different distances; the remotest light is the oddman-out. The subject's task is to touch the lighted button that is the oddman-out; this immediately shuts off all three lights. With as many as eight possible light buttons, there are 44 different possible combinations of three lights going on simultaneously to produce an oddman-out pattern. The odd light in different patterns varies slightly in perceptual discriminability, depending on its distance from the other two lights. Oddman RT is expected to be considerably slower than in SRT and CRT, and is theoretically predicted to be more highly correlated with IQ because it involves a larger number of elementary information processes. The Oddman task has 8 practice trials and 36 test trials.

The tasks were administered to all subjects in the following order: (a) SRT, (b) CRT, (c) Oddman, and (d) CRT, with half the CRT trials given before and half after the Oddman task. Subjects had several minutes rest between tasks.

It should be noted that the computer administration was programmed so that, on every task, each subject obtained the same number of "good" (i.e., errorless) trials. In order to remove obvious outliers from each subject's RT trials, the trial was invalidated and recycled if a subject released the home button before the onset of the stimulus light, or if the RT was less than 170 ms (which is too fast a response to be a true RT) or more than 2,000 ms (which would be an extreme outlier for normal children in the preset age group). Pressing the wrong button also invalidated the response, and the trial was recycled. (The number of all recycled trials was automatically recorded for each subject.) In CRT and Oddman RT the stimulus pattern for the erroneous trials was not repeated immediately but always occurred at the end of the entire series of trials, in order to maintain randomness and uncertainty as to which button would light up on the next trial.

Chronometric Variables. Each of the three elementary tasks yields *four* chronometric scores, measured in milliseconds, plus an error score.

Reaction Time (RT). The time interval between the onset of the stimulus and the subject's lifting his or her finger from the home button (measured as the *median* RT over all trials).

Movement Time (MT). The interval between lifting the finger from the home button and pressing the lighted stimulus button, which is a movement of 6 in. (15.24 cm; measured as the *median* MT over all trials).

Standard Deviation of RT (SDRT). This is a measure of intraindividual variability in RT, that is, the variability of the subject's RT from trial-to-trial. SDRT is calculated as the standard deviation (*SD*) of the subject's RTs over all trials.

Standard Deviation of Movement Time (SDMT). This is a measure of intraindividual variability in MT, that is, the variability of the subject's MT around his or her own mean MT. It is calculated as the standard deviation of the subject's MTs over all trials.

Errors. Errors consist either of pressing the wrong button or making an anticipatory response before the reaction stimulus appears, or of RT < 170 ms or > 2000 ms.

Standard Progressive Matrices (SPM)
Raven's Standard Progressive Matrices is a 60-item multiple-choice nonverbal test of reasoning based on figural patterns. It is known to be one of the best measures of Spearman's *g*. In factor analyses of a wide variety of mental tests, the SPM usually has the highest loading on the general factor common to all of the tests and has negligible loadings on group factors independent of *g* (Gustafsson, 1988; Snow, Kyllonen, & Marshalek, 1984). This paper-and-pencil test was administered to intact classes, with a 45-min time limit.

Results

Age Control
Because the two racial groups differ slightly but significantly in age, and because the psychometric and chronometric variables are generally correlated with age (shown in Table 3), it was controlled by a multiple-regression procedure applied to the combined groups, removing linear, quadratic, and cubic components of each variable's regression on age (in months). Each of the age-adjusted variables retains its original raw score mean in the total sample of 820 subjects. Each variable's standard deviation in the total sample is slightly shrunken in the age-adjusted data, of course, because the component of variance associated with age has been removed.

All of the variables referred to hereafter are age-adjusted.

Calculation of Racial Effect Size (ES)
The ES of the racial classification in this study is always calculated as the mean difference divided by the square root of the *N*-weighted average of the mean-square variances within the separate groups. The means and standard deviations of the chronometric variables in the white and black groups are listed in Appendix A.

Raven's SPM

The mean SPM raw score of whites is 38.84 (*SD* = 10.23) and 31.61 (*SD* = 9.65) of blacks. Typically, the W–B difference is about 1 *SD*, but it is only 0.72 *SD* in the present sample. This smaller than typical ES of racial classification for the highly *g*-loaded SPM most likely reflects the middle-class status of the majority of subjects, both white and black. The white mean is close to Raven norms for whites in this age group.

Errors on Chronometric Tests

Because the frequency distribution of errors for each chronometric task (SRT, CRT, and Oddman) forms an extreme J curve in each racial group, with the mode at zero errors, a *t* test of the average group difference would be inappropriate. Therefore, chi square was used to test the significance of the difference between the white and black error distributions for 0 to 5 and more than 5 errors on each of the three tasks, shown in Table 1. The obtained chi square (with $df = 6$) indicates that the total frequency distribution of errors is not significantly different between the two groups on any of the tasks. However, if we dichotomize the total distribution of errors into two categories, 0 errors and 1 or more errors, the resulting 2 × 2 contingency table (i.e., Errors × Race) shows a significant ($\chi^2 = 11.71$, $p < .001$) association between errors and race, blacks having relatively more errors on the Oddman task. By the same analysis, race shows a nonsignificant ($p > .20$) association with error rates on both the SRT and the CRT tasks. The fact that even on Oddman, the most difficult task, a majority (68%) of the subjects in both groups made 0 or 1 errors out of 36 trials suggests the extreme simplicity of these tasks as compared to typical psychometric tests, including the SPM. Such low error rates on the chronometric tasks suggest that subjects fully understood the task requirements and complied with them. It is of interest that

TABLE 1
Percentages of White (W) and Black (B) Samples Making 0 to 5 and More Than 5 Response Errors on the Chronometric Tasks

	SRT		CRT		Oddman	
No. of Errors	W	B	W	B	W	B
0	83.1	79.6	94.7	92.3	45.8	32.8
1	8.4	9.4	4.8	5.9	26.0	27.2
2	2.1	3.8	0.3	1.7	13.5	11.9
3	1.5	1.7	0.2	0	6.7	9.4
4	1.4	2.1	0	0	2.7	5.5
5	1.0	0.4	0	0	1.7	3.8
> 5	2.6	3.0	0	0	3.6	9.4
χ^2	3.92, p >.50		3.57, p >.30		7.79, p >.20	

CRT has a lower error rate than SRT, very likely because of the prior practice on SRT.

Similarity of ECT Factors in the White and Black Samples

In testing Spearman's hypothesis it is important to establish that the *n* variables on which the groups are compared measure the same latent traits in both groups. This can be assessed in two ways: (a) by the similarity of factor structure and the congruence of factor loadings in the comparison groups, and (b) by the pattern of correlations of the *n* variables with an external criterion.

A principal components analysis of the 12 ECT variables was performed separately in the white and black groups. In each group (and in the combined groups), three components emerged with eigenvalues greater than 1. Cattell's (1966) *scree* test was also applied and agreed with the Kaiser-Guttman rule (i.e., eigenvalues > 1) in indicating that only the first three factors could be meaningfully rotated. (The first six eigenvalues in the combined groups are 4.53, 2.70, 1.28, 0.59, 0.37, and 0.22). Varimax rotation of the three components reveals three clearly identifiable factors, which are the same in both groups, as shown in Table 2. The three factors can be labeled: (1) RT + RTSD, (2) MT, and (3) MTSD. (Note that in most of the other factor analyses of chronometric variables that I have done, RT and RTSD have their largest loadings on the same factor, because they are not experimentally independent measures and have cor-

TABLE 2
Varimax Factors Extracted From ECT Variables in White (W) and Black (B) Groups

	Varimax Factors					
	1 RT + RTSD		2 MT		3 MTSD	
Variable	**W**	**B**	**W**	**B**	**W**	**B**
SRT	**67**	**60**	32	52	−13	−13
SRTSD	**63**	**57**	04	25	−05	26
SMT	10	13	**88**	**94**	−17	−05
SMTSD	01	11	−07	07	**69**	**73**
CRT	**85**	**74**	15	37	−06	−11
CRTSD	**70**	**68**	11	02	15	31
CMT	11	14	**94**	**95**	−13	−01
CMTSD	−00	14	−11	−06	**72**	**70**
OddRT	**83**	**82**	−01	02	04	−15
OddRTSD	**64**	**73**	−14	−09	30	13
OddMT	06	01	**91**	**92**	01	08
OddMTSD	09	−10	−04	−03	**76**	**70**
% Variance	26.7	25.9	22.1	24.8	14.7	14.6
Congruence Coefficient	.985		.974		.935	

Note. Salient factor loadings appear in boldface.

TABLE 3

Correlation (r) of ECT Variables With SPM in White (W) and Black (B) Groups, Mean r of W and B Groups, Effect Size (ES) of the Group Mean Difference of the ECT Variables, and Correlation of Variables With Age (in Months) in Combined Groups

	Correlation			ES	
Variable	W	B	Mean r	(W–B)	r_{age}
SRT	−.069	−.039	−.053	−.003	−.115
SRTSD	−.189	−.160	−.174	−.167*	−.076
SMT	−.023	−.061	−.042	.114	−.036
SMTSD	−.129	−.100	−.114	−.097	−.015
CRT	−.147	−.085	−.116	.053	−.094
CRTS	−.245	−.175	−.210	−.235**	−.037
CMT	−.060	−.084	−.072	.063	−.013
CMTSD	−.145	−.119	−.132	−.086	−.011
OddRT	−.269	−.136	−.203	−.189*	−.126
OddRTSD	−.244	−.166	−.205	−.258**	.001
OddMT	−.094	−.087	−.090	−.057	.009
OddMTSD	−.193	−.180	−.187	.009	−.011

Note. Two-tailed tests of significance of correlations: white group, $r > .08$ has $p < .05$, $r > .11$ has $p < .01$; black group, $r > .13$ has $p < .05$, $r > .17$ has $p < .01$; mean r, $r > .07$ has $p < .05$, $r > .09$ has $p < .01$.

*$p < .05$, two-tailed. **$p < .01$, two-tailed.

related errors of measurement. When RT and RTSD are based on experimentally independent measures, their variance is often divided between two factors, one that is predominantly RT, the other predominantly RTSD. For a comprehensive discussion of the relation between RT and RTSD, see Jensen, 1992.) The congruence coefficients between the factors in the two groups indicate that they are highly similar. Congruence coefficients of .90 or above are generally considered indicative that the same factor is measured in both groups.

How similar are the two groups in the pattern of correlations between each of the 12 ECT variables and the SPM? These correlations (shown in Table 3) represent the loadings of the ECT variables on the psychometric *g* factor as measured by the SPM. The similarity of these vectors between the two groups was assessed by three indices: Pearson *r*, Spearman's rank-order correlation (r_s) and the coefficient of congruence (r_c), regarding the vectors as the ECT variables' factor loadings on a psychometric *g* factor; the Pearson $r = .85$, $r_s = .85$, $r_c = .97$. All three indices show that the vector of correlations of the ECT variables with the SPM, or psychometric *g*, is highly similar in the black and white groups. These correlations of each of the ECT variables with the SPM, therefore, may be averaged across the white and black groups, thus providing a more reliable vector of correlations to be used in testing Spearman's hypothesis. It should be noted that the mean differences between the white and black groups on the ECT variables does not enter into the averaged correlations between each of the ECTs and the SPM.

Significance of the W–B Differences on the ECT Variables

Because the ECT variables are correlated with one another, the overall significance of the W–B difference on these variables must be tested by a multivariate method.

A multivariate analysis of variance (MANOVA) was performed on all 15 of the ECT variables, including the error scores on each task. The first canonical variate shows a highly significant difference between the black and white groups, $F(15, 804) = 4.10$, $p < .0001$.

The significance levels of the single ECT variables as tested by the t test are indicated by asterisks in the ES column of Table 3. The significance levels of the ES for the separate variables are unimportant for testing Spearman's hypothesis, which depends on the pattern or rank order of the ESs across the ECT variables. But it is notable that the largest and most significant ESs occur on those variables that have shown the largest correlations with g in many previous studies, namely the RTSD of each ECT. To look at this in more general terms, three factor scores were generated for each subject based on the Varimax factors described in Table 2. The difference between the groups' means on each of the three factor scores was tested for significance and only the first factor (RT + RTSD), with ES = .19, shows a significant difference, $t(818) = 2.41$, $p < .02$. The group differences on the MT and MTSD factors are nonsignificant ($t < 1$ and 1.3, respectively).

Correlation of ECT Variables With Raven's SPM

Table 3 shows all the Pearson correlations between the single ECT variables and the SPM.

Tests of Spearman's Hypothesis

A test of Spearman's hypothesis is essentially the correlation between two column vectors: (a) a number of variables' loadings on psychometric g (or the variables' correlations with a test known to be highly loaded on psychometric g, in this case Raven's SPM); and (b) the standardized differences, or ESs, between the white and black means on the same set of tests. For convenience, these two vectors are henceforth symbolized Vg and VES.

The correlation between Vg and VES is calculated both as a Pearson r and as Spearman's nonparametric rank-order correlation, rs. As previously explained, only the r_s is tested for significance, because the significance test of r_s does not make any assumptions about the population distribution of the vectors or their scale properties. Because Spearman's hypothesis predicts a direct relationship between Vg and VES, and hence a positive correlation, only one-tailed p values accompany the r_s reported in the following text.

Hypothesis Tested With Vg in Separate Groups. The correlation between VES and the white group's Vg is $r = .83$, $r_s = .76$, $p < .01$; between VES and the black group's V*g*, $r = .69$, $r_s = .66$, $p < .05$. The hypothesis is borne out significantly with V*g* based on the separate groups.

Hypothesis Tested With the Mean* Vg *of the Two Groups. Averaging the values of V*g* of the separate groups affords a more reliable measure of the ECT variables' relation to psychometric *g*. The correlation between VES and the *mean* V*g* is $r = .81$, $r_s = .79$, $p < .01$.

Control for ECT Reliability. The size of the correlation between the SPM and an ECT variable, of course, is affected to some degree by the reliability of the variable. The same is true of the standardized difference between the black and white means, or the ES. So we must rule out the possible artifact of confirming Spearman's hypothesis only because of variation in the reliability coefficients of the ECT variables. This can be done in two ways: (a) partialing the vector of ECT reliability coefficients (labeled V*r*) out of the correlation between VES and V*g*, and (b) correcting each of the values in VES and V*g* for attenuation by dividing each value by the square root of its reliability.

The best lower-bound estimates of the reliability of the ECT variables in this study are the estimated communalities of the variables, that is, the squared multiple correlation (SMC) of each of the 12 ECT variables with all the other variables. Other methods for estimating internal consistency reliability, such as odd-even split-half trials and Cronbach's coefficient alpha are not applicable to the RTSD and MTSD measures. The split-half correlation for RTSD could not be boosted by the Spearman-Brown formula (as could RT and MT) because it does not meet the essential assumptions underlying the logic of that formula. It would be improper to estimate reliabilities of different variables by different methods in this case, and so a uniform method of estimation by SMC for all variables, although a lower-bound estimate, seems preferable. SMC estimates of reliability (r_{xx}), shown in Appendix A, were obtained for every variable separately in the black and white groups. The Pearson correlation between the vectors of SMCs (i.e., the estimated reliability coefficients) for blacks and whites is .96 (r_s = .97). Hence, the two vectors can be averaged.

1. *Partial* r. When the mean V*r* is partialed out of the correlation between VSE and mean V*g*, the partial $r = .74$, partial $r_s = .73$, $p < .01$.
2. *Correlation With Disattenuated Variables.* Each value in VES and the mean V*g* was corrected for attenuation and the correlation between the disattenuated vectors was computed: $r = .74$, $r_s = .67$, $p < .05$.

The results of these two analyses, partialing and disattenuating, are consistent in demonstrating that the substantiation of Spearman's hypothesis is not an artifact of variation in ECT reliability.

Is the Correlation Between VES and Vg Due to g or to Other Factors? To examine the remote possibility that the ES reflects mainly some other component(s) of variance in the chronometric measures besides *g* and that the correlation between VES and V*g* is merely coincidental, the following analysis was performed. The linear component of the SPM variance was regressed out of each of the ECT variables in each racial group, and the ES was calculated on the resulting W–B mean difference. The ES from which the SPM has been removed is labeled. ES-*g*. The question then is how the ES is affected by the removal of *g*, or how does ES-*g* differ from ES? The overall mean of ES = −.071; the mean of ES-*g* = +.030; the difference (mean ES − mean ES-*g*) = −.101. But especially noteworthy is the change in sign of the ES. The W–B difference is not just eliminated by removing *g*, it is *reversed,* with the blacks showing the faster response times. There is no really satisfactory statistical test of the significance of this overall difference. But the telling point is the distinctive pattern of the differences on the various ECT variables. The variables with the largest negative ESs are decreased the most by removing *g*. The larger negative ESs (i.e., longer response times for blacks than for whites) occur mainly on the RT and RTSD variables. The mean ES for the RT + RTSD variables = −.133; on the same variables the mean ES-*g* = −.007. That is, removing *g* markedly reduces the W–B mean difference on those variables with the larger negative ES, which are those that mostly reflect decision and discrimination processes. Removing *g* shows an opposite effect on the ECT variables that involve movement time. The mean ES of the MT + MTSD variables = −.009; on the same variables the mean ES-*g* = +.067. That is, removing *g* increases the black advantage in response times on those varieties that most involve movement speed.

This pattern of ES − ES-g differences suggests that, in the observed ESs, *g* acts as a suppressor variable of the purely motor aspects or response time. Although there are only four ESs with positive signs (i.e., blacks faster than whites), there are seven ES-*g*s with positive signs, with the changes from negative to positive signs occurring on variables SRT, CMTSD, and OddMT. Conversely, the motor component of the ECT variables acts, to some degree, as a suppressor of their *g* component, so that if the motor aspect could be removed from all the variables, VES would be almost perfectly correlated with V*g*.

There is only one apparent anomaly in the whole picture, namely, the puzzling fact that CRT shows a positive (though nonsignificant) ES; that is, the black group was slightly faster than the white group on CRT (435.3 ms vs. 439.2 ms). Yet, CRT is very significantly greater than SRT, as should be expected, and it is more negatively correlated with the SPM than is SRT, also as expected. And the ES of CRTSD is quite in line with expectation and with the rest of the data. Attempts to explain why CRT itself is out of line would be purely speculative at present.

But the fact that the ES is positive (i.e., blacks have shorter response times) on 4 of the 12 chronometric variables is contrary to Spearman's hypothesis and

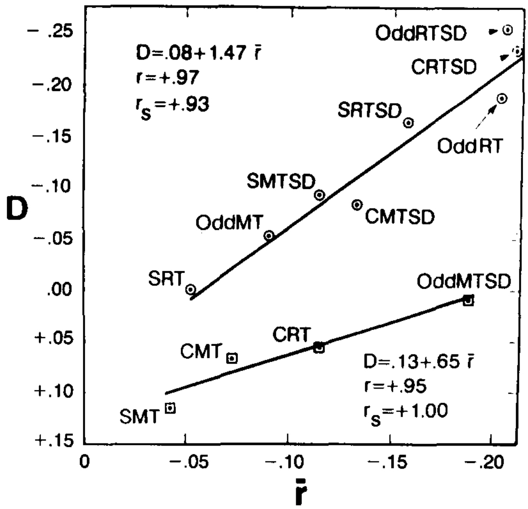

Figure 2. Regressions of the standardized mean W–B difference [**D** = (white mean − black mean)/(average within-group *SD*)] on the average correlation, $\bar{r}$, of the ECT variables with the Raven SPM. Two separate regressions are shown, one for *negative* values of **D**, and one for *positive* values of **D**. If *g* were the sole determinant of B–W differences on all of the chronometric variables, all of the data points should fall on only one and the same regression line.

indicates that some other factor in addition to *g* is involved; as this factor is negatively correlated with *g*, it suppresses its effect on performance. This is seen more easily by plotting graphically the relation between ES and the mean correlation of the ECT variables with the SPM. Such a plot is shown in Figure 2, in which the ordinate is the ES, or **D** (for the W–B mean difference), and the abscissa is the average correlation, $\bar{r}$, between each ECT variable and the SPM. From inspection of the scatter diagram, it can be seen that the data points are best fitted by two separate regression lines, one line for the 4 variables on which **D** (or ES) is positive (i.e., blacks average shorter response times or smaller MTSD) and the other line for the 8 variables on which **D** is negative (i.e., whites average shorter response time and smaller RTSD). Although the use of separate regressions here is itself discordant with Spearman's hypothesis, it is most interesting that for each of the two regression lines, the fit of the data points to Spearman's hypothesis is very close, as indicated by the large values of r and r_s in each case. The fact that 4 of the 12 variables show positive W–B differences, probably because of a motor component acting in an opposite direction as a suppressor variable to the *g* factor in these four ECT variables to an extent that counteracts the negative difference, still does not obscure the manifestation of Spearman's hypothesis: As the ECT variables increase in *g* loading, the W–B difference in response time trends linearly from a positive value toward zero. It appears that this phenomenon reflects the interaction of two distinct factors: a *g*-related decision speed and consistency factor, on the one hand, and a non-*g* movement speed factor, on the other, with the mean W–B difference going in opposite directions on the two factors. (CRT is inexplicably the one variable that is discordant with this interpretation.) A quite similar phenomenon has been noted in the General Aptitude Test Battery (GATB); the one aptitude (motor coordination) on which the W–B mean difference is zero despite having a *g* loading of .48 (Jensen, 1980, p. 734, Figure 15.2). Evidently, blacks excel whites on the motor component in this test, and this large component suppresses the effect of the test's *g* loading in determining the mean W–B difference; the other eight aptitudes measured by the GATB all show significantly positive W–B mean differences, with ESs ranging from .20 to 1.4.

Observed ES Compared With Theoretically Predicted ES. The standardized mean W–B differences, or ESs, on the very simple ECTs in this study are, of course, individually quite small, as would be expected for variables with generally small *g* loadings. But the important question is: How do the magnitudes of the ESs compare with theoretical expectation from Spearman's hypothesis, given the correlation of each ECT with the SPM? In accord with Spearman's hypothesis that the size of the W–B difference on any mental test is directly related to the test's *g* loading, we should expect that the SPM, as a good though imperfect measure of *g*, would give an approximate prediction of the ES on each ECT variable to the extent that it is *g* loaded, as estimated by the ECT variable's correlation with the SPM. The ES for the SPM itself is .72 and so the predicted ES for any ECT variables would be the product of .72 and the average within-group correlation of the variable with the SPM. The overall mean of the *observed* ESs is −.071; the mean ES for just the RT + RTSD variables is −.133; and for just the MT + MTSD variables the mean is −.009. The means of the corresponding *predicted* values are −.095, −.115, and −.071, respectively. The differences between the mean observed and predicted (O-P) ESs are: Overall = +.024; RT + RTSD = −.018; and MT + MTSD = +.062. It can be seen that the O-P difference in ES is smaller for the RT + RTSD variables and there is slight *under*prediction of the ESs, whereas the MT + MTSD ES is *over*-predicted. Hence, the deviations of the observed from the predicted values seem to be associated with the motor speed aspect of response time, on which the black group averages faster than the white group, and which has considerably greater effect in the MT variables than in the RT variables, which more strongly reflect the speed of information processing.

STUDY 2
Method and Procedures

Subjects

The subjects were 73 white and 118 black children in Grades 4 to 6 in three of the schools where Study 1 was conducted, which afforded more time for individual testing. The unequal sample sizes in the two groups are not by design but resulted from time constraints in scheduling pupils for individual testing near the end of the spring term. To obviate obtaining further parental permission slips and scheduling additional classroom time for administering Raven's SPM, only those children who took part in Study 1 were included in Study 2. The only planned basis for selection of subjects from the total pool available for this study was a brief screening test for knowledge of elementary number facts (described in the following).

Mean ages of the white and black groups were 10.8 (SD = 1.0) and 10.7 (SD = 1.1) years, respectively.

Techniques and Procedures

Because the chronometric tasks are intended to measure individual differences in information-processing speed rather than in knowledge per se, subjects in Study 2 were specially selected on the basis of their obtaining a perfect score on a short, nonspeeded, written, group-administered test of elementary number facts—simple addition, subtraction, and multiplication, all based on single digits—that pupils typically have learned before fourth grade. The chronometric tasks in Study 2 comprised the same number facts as the written test and were intended to measure only the speed of retrieval of this information from long-term memory.

Chronometric Apparatus. The subject's binary response console, shown in Figure 3, consists of a 6½ × 10 in. (16.51 × 25.40 cm) panel of three push buttons, each of 1 in. (2.54 cm) diameters, spaced equidistantly 2½ in. (6.35 cm) apart in the form of a triangle, with the black home button at the lower point nearest the subject. The two green binary response buttons are labeled **Yes** and **No.** The console is interfaced with an IBM-PCXT computer, which controlled the presentation of stimuli and recorded the subjects' response time and errors. The response stimuli were presented on an IBM monochrome monitor located directly behind the subjects' response console.

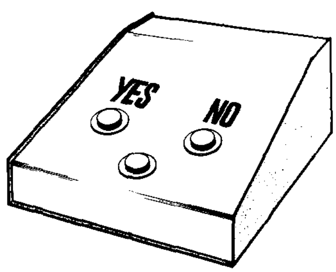

Figure 3. Binary response console (6½ × 10 in. [16.51 × 25.40 cm]) used in the Math Verification Test (MVT). The lower button in the equilateral triangle formed by the 1 in. (2.54 cm) diameter push buttons is the "home" button, with the response buttons (**Yes** or **No**) 2½ in. (6.35 cm) above the home button. The reaction stimuli (math problems in this study) are presented on an IBM-PC display placed behind the response console.

Math Verification Test (MVT). The MVT consists of single-digit *addition, subtraction,* and *multiplication* problems, presented on the monitor in the form **A * B = C,** where **A** and **B** are digits 1 through 9 and * is either +, −, or ×. On a random half of the trials, **C** is the *correct* answer (in response to which the subject lifts the index finger holding down the home button and touches the button labelled **Yes**) and on the other random half of the trials **C** is an incorrect answer (to which the subject presses the button labelled **No**). Three types of incorrect answers were programmed: (a) the correct answer + 1 (e.g., 3 + 4 = 8), (b) the correct answer for a different kind of computation (e.g., 2 + 4 = 8), and (c) a figure much larger or smaller than the correct answer (e.g., 2 + 4 = 24). In the addition and multiplication problems, the larger digit in each equation appeared first or second on alternate trials. In the subtraction problems the larger digit was consistently the minuend.

Each type of problem was presented in a separate block of trials, in the order: addition, subtraction, multiplication. Each set was prefaced by a title to inform the subject of the type of problem in the following set of problems. Fifteen practice trials comprising equal numbers of each type of problem were given before the first set of test trials. There were 20 test trials in each set. The trials were subject paced. One second after the subject pressed the home button, a preparatory stimulus appeared in the center of the screen to orient the subject's attention; it consisted of a square located at the exact spot where the mathematical operation sign (+, −, or ×) of the upcoming problem would appear. After a continuous random interval of 1 to 4 s, the problem appeared on the screen. The subject's response was immediately followed by the word **Correct** and a 1-s beep. An erroneous response was followed by the word **Incorrect** without any beep. As in Study 1, every subject obtained the same number of errorless trials, because trials with incorrect responses were repeated at the end of each set. Also, as in Study 1, trials on which RTs were shorter than 170 ms or longer than 2,000 ms were eliminated and recycled.

Chronometric Scores on the MVT. The measures derived from the MVT procedure are defined in precisely the same way as in Study 1, except that on the MVT, errors are defined only as incorrect responses. Median RT, RTSD, median MT, and MTSD (in milliseconds) over 20 test trials were obtained for each type of arithmetic problem.

Results

As the following analyses are essentially the same as those of Study 1, they need not be spelled out again in detail.

Age Control

As in Study 1, the linear, quadratic, and cubic components of the regression of the chronometric variables on chronological age were regressed out of all of the measurements in the combined groups prior to all of the following analyses. The

Pearson r of each of the variables with age (in months) in the combined groups is shown in Table 5 (p. 68). Descriptive statistics on the age-adjusted MVT variables are given in Appendix B.

Raven SPM Scores

The white group's mean = 38.56 (SD = 8.86); the black group's mean = 31.70 (SD = 9.86). The difference is significant, $t(189) = 4.83, p < .001$; ES (W–B) = .72.

Errors

As was expected, error responses were few, with a mode of 0 on every problem type and with means ranging from 1.7 to 2.2 in the white group and 1.7 to 2.6 in the black group. The groups do not differ significantly in number of errors on any of the problem types.

Similarity of MVT Factors in the White and Black Samples

A principal factor analysis of the 12 MVT variables reveals two large and clear-cut factors in both racial groups. The Varimax rotated factors are shown in Table 4. The two factors in both groups clearly reflect the RT and MT variables. The

TABLE 4
Varimax Factors Extracted From MVT Variables in White (W) and Black (B) Groups

	Varimax Factors			
	I RT + RTSD		II MT + MTSD	
Variable	**W**	**B**	**W**	**B**
Addition				
RT	**90**	**85**	−14	−12
RTSD	**78**	**63**	09	22
MT	−07	−01	**83**	**80**
MTSD	−01	04	**67**	**64**
Subtraction				
RT	**91**	**89**	−15	−11
RTSD	**79**	63	05	24
MT	−15	−15	**82**	**79**
MTSD	06	08	**69**	**75**
Multiplication				
RT	**77**	**93**	−08	−14
RTSD	**81**	**87**	00	06
MT	−09	−09	**83**	**83**
MTSD	−02	04	**72**	**77**
% Variance	34.5	33.0	29.6	30.5
Congruence Coefficient	.989		.991	

Note. Salient factor loadings appear in boldface.

congruence coefficients for both factors are so high as to indicate virtual identify of these factors in the white and black groups. It should be noted that the factor structure of the MVT variables does not in the least reflect the types of arithmetic operations, but clearly reflects only the different response time measures, mainly RT and MT. On the chance that type of arithmetic problem might be registered on a third factor, albeit a nonsignificant one (i.e., eigenvalue < 1), three factors were extracted and Varimax rotated. The third factor is relatively small (variance < 3% in both groups). and is nondescript in its pattern of loadings, which does not evince the slightest relation to problem types. Evidently, as regards individual differences within groups, the chronometric measures reflect only variance associated with basic cognitive and motor processes, rather than the interaction of individual differences with the various types of arithmetic problems.

The column vector (symbolized V*g*) of correlations of each of the 12 MVT variables with the SPM, shown in Table 5, is also quite similar in the white and black groups, as indicated by a Pearson $r = .87$, $r_s = .74$, $p < .01$.

TABLE 5
Correlation (*r*) of MVT Variables With SPM in White (W) and Black (B) Groups, Mean *r* of W and B Groups, Effect Size (ES) of the Group Mean Difference on the MVT Variables, and Correlation of Variable With Age (in Months) in Combined Groups

Variable	Correlation W	Correlation B	Mean *r*	ES (W–B)	r_{age}
Addition					
RT	−.314	−.181	−.247	−.376*	−.236
RTSD	−.247	−.205	−.226	−.194	−.213
MT	.063	.191	.127	−.056	.019
MTSD	.011	.053	.032	.007	−.003
Subtraction					
RT	−.310	−.238	−.274	−.430**	−.204
RTSD	−.184	−.218	−.201	−.220	−.183
MT	.095	.219	.157	−.032	−.022
MTSD	−.070	.122	.026	.007	−.060
Multiplication					
RT	−.234	−.109	−.171	−.443**	−.197
RTSD	−.095	−.046	−.071	−.343**	−.140
MT	.219	.217	.218	−.178	−.051
MTSD	.145	−.003	.071	−.070	−.114

Note. Two-tailed tests of significance of correlations: white group, $r > .23$ has $p < .05$, $r > .31$ has $p < .01$; black group, $r > .18$ has $p < .05$, $r > .24$ has $p < .01$; mean r, $r > .143$ has $p < .05$, $r > .189$ has $p < .01$.

*$p < .05$, two-tailed. **$p < .01$, two-tailed.

Correlation of MVT Variables With Raven SPM

The Pearson correlations of each of the MTV variables with the SPM are shown in Table 5. The multiple correlation of all 12 MVT variables with the SPM in the white group is $R = .40$ ($p < .05$); in the black group, $R = .44$ ($p < .01$).

Hypothesis Tested With Vg in Separate Groups. Spearman's hypothesis predicts a positive correlation between VES and Vg. The correlation between VES and Vg in the white group is $r = .75$, $r_s = .65$ ($p < .05$); in the black group, $r = .70$, $r_s = .71$, $p < .01$.

Hypothesis Tested With the Mean Vg of the Two Groups. The correlation between VES and the mean Vg is $r = .73$, $r_s = .68$, $p < .03$.

Control for Variation in Reliability of MVT Variables. A lower-bound estimate of the reliability of each of the MVT variables is provided by the SMC of each variable with all of the other 11 variables, shown in Appendix B. The vector of SMCs, averaged for blacks and whites, was partialed out of the correlation between the VES and the mean Vg. The partial correlation is $r = .70$, $r_s = .70$. The fact that these correlations, with the vector of estimated reliabilities of the MVT variables partialed out, are of about the same magnitude as the zero-order correlations indicates that the confirmation of Spearman's hypothesis is not an artifact of the differences among the MVT variables' reliability coefficients.

Another way to examine the effect of variable reliability on the test of Spearman's hypothesis is to correct both VES and Vg for attenuation, using the SMCs. The correlation between the disattenuated VES and disattenuated Vg is $r = .73$, $r_s = .65$, $p < .05$. Again, it is evident that the substantiation of Spearman's hypothesis is not a result of the differing reliabilities of the MVT variables.

Comparison of Observed and Theoretically Predicted ESs. According to Spearman's hypothesis, the ES (i.e., W–B) on cognitive tests should directly reflect their g loading, which in this case is estimated by their correlation with the SPM. Hence, the predicted ES for any given test variable is the product of this correlation and the ES on the SPM, which for the present groups is .72. The overall mean observed ES = −.194; the predicted mean ES = −.034; the difference between observed and predicted (O-P) = −.160. For just the RT + RTSD variables, the observed mean ES = −.334; the predicted mean ES = −.143; O-P = −.191. for just the MT + MTSD variables, the observed mean ES = −.053; the predicted mean ES = +.076; O-P = −.129.

On every one of the MVT variables, we see that the ES is *under*predicted, that is, the black group actually has longer response times (both RT and MT) and greater intraindividual variability, relative to the white group, than is predicted by the g loadings of the MVT variables. There is clearly some other influence in addition to g (and not correlated with g) that enters into the W–B differences on all of the MVT variables. This non-*g* factor appears to be specific to the type of arithmetic problem, as shown by the following: The mean under prediction (O-P) of ES for Addition is −.098, for Subtraction, −.116, and for Multiplication, −.267. The theoretically predicted ESs for Addition, Subtraction, and Multiplication are −.057, −.053, and +.009, respectively, indicating that *g* has scarcely any differential relation to the type of arithmetic operation, but is differentially related to the type of processing measure, that is, RTs, MTs, and their intraindividual standard deviations. Because tasks and their order of presentation are confounded in this study, these detailed comparisons between the chronometric variables on the three types of arithmetic operations are a weak basis for causal inferences and must be viewed as speculative hypotheses.

It seems a reasonable hypothesis, however, that these arithmetic operations on single-digit numbers or the retrieval of simple number facts from long-term memory may be less completely automatized in the black group than in the white. Addition facts are the earliest learned and would have received more practice, hence greater automatization, than the more recently learned multiplication tables, whereas subtraction would be intermediate in this respect. This conjecture is consistent with the systematic increase in the degree of underprediction of the ESs, going from addition to subtraction to multiplication problems.

Spearman's Hypothesis Tested With the Combined Variables of Study 1 and Study 2

Finally, the ESs over all 24 chronometric variables (12 ECT variables and 12 MVT variables) can be correlated with the 24 variables' *g* loadings (i.e., correlations with the SPM). This correlation between VES and V*g*, with each vector containing 24 values, is $r = .82$, $r_s = .70$, $p < .001$. Thus, Spearman's hypothesis is strongly borne out by this test based on the combined variables from both studies.

DISCUSSION

In both studies, each one based on chronometric variables derived from tasks that minimize variance attributable to individual differences in prior knowledge, measuring mainly individual differences in speed and consistency of information processing, the outcome predicted by Spearman's hypothesis is significantly borne out. That is, the size of the standardized mean W–B difference (or ES) on the chronometric variables is directly related to the variables' loadings on psychometric *g*, as predicted by Spearman's hypothesis. Moreover, the degree of correlation between VES and V*g* on these chronometric variables is of about the same magnitude as the corresponding correlation found with conventional batteries of mental tests. For example, the correlation between the VES and the mean V*g* obtained with 25 conventional psychometric tests (subtests of the combined WISC-R and K-ABC) administered to groups of 86 white and 86 black fourth- and fifth-grade schoolchildren is $r = .78$, $r_s = .75$, $p < .001$ (Naglieri & Jensen, 1987).

But if Spearman's hypothesis is in fact true, one might ask why the correlation between V*g* and VES is not even larger than the quite substantial correlations we actually find. One explanation, suggested by examination of 11 tests of the hypothesis based on independent samples and different sets of psychometric tests (Jensen, 1985a), is that the size of the observed correlation between V*g* and VES is somewhat constrained by the amount of variation among the values in each vector. If some of the tests used to test Spearman's hypothesis happen to have very similar *g* loadings, the rank order of the *g* loadings is apt to be unreliable, and similarly for the rank order of the W–B differences. Hence, small differences among the elements in the vector of *g* loadings or the vector of ESs, being unreliable, would bias the test of Spearman's hypothesis to favor the null hypothesis.

Another possible cause of the obtained correlation between V*g* and VES being smaller than theoretically expected is the presence of other factors besides *g* in some of the tests on which the groups may also differ. These other factors, independent of *g*, create perturbations with respect to the test of Spearman's hypothesis. When other racially differential factors besides *g* exist in a given test, they may either increase or decrease the mean W–B difference, hence creating a perturbation in the ESs predicted by Spearman's hypothesis. Because such perturbations have been discovered and their causes identified (Jensen, 1985a), we must distinguish between a *strong* and a *weak* form of Spearman's hypothesis. The strong form states that *g* is the *only* factor determining the relative sizes of the W–B differences on various tests. The weak form states only that *g* is the *predominant* factor, but not the sole factor, determining size of W–B differences. Among conventional psychometric tests, two distinct factors have been identified that, for certain tests in which they are loaded, systematically cause a slight discrepancy between the observed ES and the ES predicted from the test's *g* loading. The two perturbation factors with respect to Spearman's hypothesis are *short-term memory,* (STM, as in forward digit span) and *spatial visualization* (as in block designs, object assembly, paper folding, and the like; Jensen & Reynolds, 1982). Blacks exceed whites, on average, on the STM factor independent of *g*, and hence tests that involve STM show a *smaller* W–B difference than is predicted by the test's *g* loading. Spatial ability has just the opposite effect. Whites exceed blacks, on average, on the spatial factor independent of *g*, and so the presence of this factor in a test makes the size of the W–B difference somewhat *larger* than is predicted by the test's *g* loading. Thus, when the variables that enter into the test of Spearman's hypothesis do not differ much from one another in their *g* loadings, these other non-*g* factors can cause the rank order of the ESs to deviate somewhat from the order that is predicted strictly from the tests' *g* loadings.

Although only the weak rather than the strong form of Spearman's hypothesis strictly accords with all of the evidence, *g* is clearly the predominant factor determining the relative sizes of the W–B differences on various psychometric tests. For example, when the W–B mean differences on various tests were expressed as point-biserial correlations and were factor analyzed among a wide variety of psychometric tests in 18 large white and black samples, it was found that the loadings of the W–B variable on the *g* factor averaged more than twice the size of the average of *all* the other significant factors (i.e., those with eigenvalues > 1; Jensen, 1987a).

In the present studies based on chronometric tasks, there is evidence of some perturbation involving the movement speed aspects of responding, especially manifested in the measurements of MT. In Study 1, based on ECTs, the MT, like RT, consistently has a negative correlation with *g* (i.e., Raven's SPM) within each racial group. That is, *faster* response times are *positively* related to higher SPM scores for both RT and MT. Yet, both on SMT and on CMT, blacks, who have *lower* average SPM scores than whites, have *faster* MTs than whites, on average. Why? It is apparent that RT generally involves more cognitive and less motor activity and agility than the MT variables; also, the absolute ESs on the RT variables conform more closely to the ESs predicted by the variables' *g* loadings, in accord with Spearman's hypothesis.

There is a marked difference between the ESs of RT and MT. The overall mean (W–B) ES in the combined Studies 1 and 2 are: ES(RT) $= -.231$; ES(MT) $= -.024$. Entirely aside from the issue of Spearman's hypothesis, this empirical finding itself would seem to render as implausible any interpretation of the W–B differences on these chronometric variables in terms of a W–B difference in motivation that affects performance. One would have to argue that whites are more highly motivated than blacks on RT but not on MT, at the same time recognizing that the responses measured as RT and MT occur within a brief period of 2 s or less after the onset of the reaction stimulus. Also to be noted is that most people report the subjective experience of their response as a being a single ballistic movement to hit the MT button as quickly as they possibly can. One of the referees of this article suggested that "RT and MT might be systematically related from trial to trial on the basis of a motivational factor (e.g., impulsivity) that might also differ between groups." However, it seems to me an implausible and purely ad hoc hypothesis that the disparity between the ESs of RT and MT could be attributable to a racial group difference in motivation that so rapidly and systematically fluctuates between the RT and MT stages of the total response time on each trial. If there were trial-to-trial fluctuations in motivation (or strategies) that affected RT and MT differently, there should be significant within-subjects correlations (positive or negative) between these variables. In three studies with a total of 267 subjects in which this question was examined, however, no evidence of such within-subject correlations between RT and MT was found (Jensen, 1987b, pp. 124–126). The frequency distribution of all these correlation coefficients, randomly dispersed about a mean near zero, is what would be expected for perfectly uncorrelated variables.

On the other hand, Spearman's hypothesis predicts the ES difference between RT and MT as being due to their different *g* loadings. (The average *g* loadings of RT and MT for the two studies combined are $g(\text{RT}) = -.18$ and $g(\text{MT}) = +.05$.)

An interpretation of the RT–MT interaction with race in terms of response strategies, whereby there is supposedly a systematic racial difference in tradeoff between RT and MT, also seems untenable. One might suppose the reason that blacks have absolutely faster MTs than whites (as is actually the case for SRT and CRT) is because their total response time is allocated differently to the RT and MT stages of responding, such that if their RT is longer, their MT will be correspondingly shorter. The use of such a strategy by a majority of subjects would result, of course, in a *negative* correlation between RT and MT. But in fact, the *r* is *positive:* for the correlation of SRT × SMT, $r = +.36$ for whites and +.53 for blacks; for CRT × CMT, $r = +.24$ for whites and +.44 for blacks.

Another way to examine the possibility of a racial difference in RT–MT tradeoff strategy is to compare the mean MT:RT ratios in each group. The use of different strategies of RT–MT tradeoff in the two groups should result in the groups' showing distinctly different MT:RT ratios, indicating that the total response times are divided differently between RT and MT in the two groups. These MT:RT ratios for Studies 1 and 2 are shown in Table 6. In Study 1 (based on ECTs), the differences between the ratios for whites and blacks are negligible. But Study 2 (MVT) shows a small yet consistent group difference, with blacks having relatively shorter MTs on each problem type. Although this could simply reflect superior motor speed and agility in the black group, a possible strategy effect cannot be ruled out in the case of the MVT. However, it would be puzzling why the groups should differ in strategies, if indeed that is what these data signify. But the effect of differing strategies would, if anything, only tend to becloud the test of Spearman's hypothesis; strategy differences would not "explain away" the empirical findings of numerous studies that repeatedly substanti-

TABLE 6
MT : RT Ratios of White (W) and Black (B) Groups in Studies 1 and 2

Study 1 (ECTs)			Study 2 (MVT)		
Ratio	**W**	**B**	**Ratio**	**W**	**B**
SMT : SRT	.85	.82	Add MT : RT	.29	.25
CMT : CRT	.70	.70	Sub MT : RT	.30	.25
OddMT : OddRT	.50	.49	Mult MT : RT	.28	.25

ate the hypothesis with a wide variety of cognitive measures, both psychometric and chronometric.

Intraindividual trial-to-trial variability in RT (i.e., RTSD) has been of considerable interest in the chronometric study of individual differences, mainly because, in many studies, RTSD has shown as large or larger correlations with *g* measures than does median RT itself, despite the usually lower reliability of RTSD (Jensen, 1987b, 1992; Larson & Alderton, 1990). Yet, RTSD is not a measure of information-processing *speed* per se, but indicates the degree of fluctuation (or conversely, the *consistency*) in the speed with which information is processed at different points in time. RTSD has been hypothesized to reflect the amount of errors or "noise" in the neural transmission of information (Eysenck, 1987). MTSD, on the other hand, is quite different from RTSD; previous studies show much smaller correlations between MTSD and *g*. The mean correlation of RTSD and MTSD with psychometric *g* in 26 independent samples (with total $N > 1{,}100$) are −.275 and −.103, respectively (Jensen, 1987b). In view of this well-established difference in the *g* loadings of RTSD and MTSD, it should be of interest to look at these variables in relation to Spearman's hypothesis. Table 7 shows the overall mean correlations of all the RTSD and MTSD variables with Raven's SPM and their corresponding mean W–B ESs. The correlations of SDRT with the SPM are larger than for MTSD, and the ESs are correspondingly (and significantly) larger for RTSD in both studies, as Spearman's hypothesis should predict.

Finally, it should be noted that there is no obvious, self-evident explanation for the empirical phenomena subsumed under Spearman's hypothesis, nor is the study here addressed to that question. A causal explanation would itself call for other hypotheses. What the studies here do substantiate, however, is that Spearman's hypothesis applies not only to conventional psychometric tests, as many previous studies have amply proved, but it also applies as well to a qualitatively quite different type of mental measurement: chronometric variables based on tasks that involve only the simplest, most elemental aspects of information processing, speed and consistency of stimulus apprehension, decision, discrimi-

TABLE 7
Mean Correlation (*r*) of RTSD and MTSD With Raven's SPM and Their Corresponding Mean W–B Effect Size (ES)

	Study 1 (ECTs)		Study 2 (MVT)	
Variable	***r***	**ES**	***r***	**ES**
RTSD	−.20***	−.22**	−.17**	−.25*
MTSD	−.14***	−.06, n.s.	.04, n.s.	−.02, n.s.
Difference	−.06, n.s.	−.16**	−.21*	−.23*

*$p < .05$, one-tailed. **$p < .01$, one-tailed. ***$p < .001$, one-tailed.

nation, and retrieval of information in long-term memory (LTM). And, as shown in many other studies, individual differences in these chronometric variables are correlated in varying degrees with psychometric *g*. The fact that the relative sizes of these correlations are directly related to the variable sizes of the standardized W–B differences on the chronometric variables means that the racial difference on *g*-loaded tests is not solely attributable to group differences in the specific information content of LTM or to strategies of reasoning and problem solving, or other complex metaprocesses of the kind presumably involved in conventional mental tests.

As I have discussed in detail elsewhere (Jensen, 1991, in press), the practical implications of this study, as of research on Spearman's hypothesis in general, derives from the considerable relationship of psychometric *g* to individual and population subgroup differences in many educationally, economically, and socially significant variables.

REFERENCES

Borkowski, J.G., & Krause, A. (1983). Racial differences in intelligence: The importance of the executive system. *Intelligence, 7,* 379–395.

Borkowski, J.G., Krause, A., & Maxwell, S.E. (1985). On multiple determinants of racial differences in intelligence: A reply to Jensen. *Intelligence, 9,* 41–49.

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245–276.

Eysenck, H.J. (1987). Intelligence and reaction time: The contribution of Arthur Jensen. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy.* New York: Falmer Press.

Frearson, W.M., & Eysenck, H.J. (1986). Intelligence, reaction time (RT) and a new "odd-man-out" RT paradigm. *Personality and Individual Differences, 7,* 808–817.

Gustafsson, J.-E. (1988). Hierarchical models of individual differences in cognitive abilities. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4). Hillsdale, NJ: Erlbaum.

Humphreys, L.G. (1985a). Race differences and the Spearman hypothesis. *Intelligence, 9,* 275–283.

Humphreys, L.G. (1985b). Attenuated hypothesis or attenuated test of hypothesis? *Intelligence, 9,* 291–295.

Jensen, A.R. (1980). *Bias in mental testing.* New York: Free Press.

Jensen, A.R. (1985a). The nature of the black–white difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences, 8,* 193–219.

Jensen, A.R. (1985b). The black–white difference in *g:* A phenomenon in search of a theory. *Behavioral and Brain Sciences, 8,* 246–263.

Jensen, A.R. (1985c). Humphrey's attenuated test of Spearman's hypothesis. *Intelligence, 9,* 285–289.

Jensen, A.R. (1985d). Race differences and Type II errors: A comment on Borkowski and Krause. *Intelligence, 9,* 33–39.

Jensen, A.R. (1987a). Further evidence for Spearman's hypothesis concerning black–white differences on psychometric tests. *Behavioral and Brain Sciences, 10,* 512–519.

Jensen, A.R. (1987b). Individual difference in the Hick paradigm. In P.A. Vernon (Ed.), *Speed of information processing and intelligence.* Norwood, NJ: Ablex.

Jensen, A.R. (1988). Speed of information processing and population differences. In S.H. Irvine &

J.W. Berry (Eds.), *Human abilities in cultural context.* Cambridge: Cambridge University Press.

Jensen, A.R. (1991). Spearman's *g* and the problem of educational equality. *Oxford Review of Education, 17,* 169–187.

Jensen, A.R. (1992). The importance of intraindividual variation in reaction time. *Personality and Individual Differences, 13,* 869–882.

Jensen, A.R. (in press). Psychometric *g* and achievement. In B.R. Gifford (Ed.), *Testing and public policy.* Boston: Klewer.

Jensen, A.R., & Faulstich, M.E. (1988). Difference between prisoners and the general population in psychometric *g*. *Personality and Individual Differences, 9,* 925–928.

Jensen, A.R., & Reynolds, C.R. (1982). Race, social class, and ability patterns on the WISC-R. *Personality and Individual Differences, 3,* 423–438.

Larson, G.E., & Alderton, D.L. (1990). Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence, 14,* 309–325.

Lynn, R., & Holmshaw, M. (1990). Black–white differences in reaction times and intelligence. *Social Behavior and Personality, 18,* 299–308.

Naglieri, J.A., & Jensen, A.R. (1987). Comparison of black–white differences on the WISC-R and the K-ABC: Spearman's hypothesis. *Intelligence, 11,* 21–43.

Open peer commentary. (1985). *Behavioral and Brain Sciences, 8,* 219–246.

Open peer commentary. (1987). *Behavioral and Brain Sciences, 10,* 507–512.

Osborne, R.T., & McGurk, F.C.J. (Eds.). (1982). *The testing of Negro intelligence* (Vol. 2). Athens, GA: Foundation for Human Understanding.

Shockley, W. (1987). Jensen's data on Spearman's hypothesis: No artifact. *Brain and Behavioral Sciences, 10,* 512.

Snow, R.E., Kyllonen, P.C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2). Hillsdale, NJ: Erlbaum.

Spearman, C. (1927). *The abilities of man.* New York: Macmillan.

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

Vernon, P.A., & Jensen, A.R. (1984). Individual and group differences in intelligence and speed of information processing. *Personality and Individual Differences, 5,* 411–423.

APPENDIX A

Means, Standard Deviations (in ms), and Reliability Coefficients (r_{xx})[a] of the ECT Chronometric Variables in the White (n = 585) and Black (n = 235) Groups

	White			Black		
Variable	M	SD	R_{xx}	M	SD	R_{xx}
SRT	347.4	60.7	.575	347.6	72.9	.596
SRTSD	84.4	55.5	.302	94.4	69.6	.400
SMT	295.0	90.0	.711	283.7	119.0	.837
SMTSD	93.1	61.8	.187	99.1	63.1	.225
CRT	439.2	73.4	.692	435.3	71.9	.638
CRTSD	95.6	48.8	.393	108.0	61.1	.427
CMT	309.5	81.7	.850	303.7	115.0	.897
CMTSD	97.3	52.5	.222	101.8	50.6	.269
OddRT	717.2	151.9	.643	745.9	151.9	.558
OddRTSD	179.1	76.5	.492	199.5	85.0	.463
OddMT	360.3	108.1	.750	367.3	154.7	.804
OddMTSD	177.8	83.9	.336	177.1	75.8	.246

[a]The reliability coefficient given here is a lower-bound estimate of reliability provided by the squared multiple correlation of a given variable with all of the other 11 variables.

APPENDIX B

Means, Standard Deviations (in ms), and Reliability Coefficients (r_{xx})[a] of the MVT Chronometric Variables in the White (n = 73) and Black (n = 118) Groups

	White			Black		
Variable	M	SD	R_{xx}	M	SD	R_{xx}
Addition						
RT	1531	889	.856	1838	768	.839
RTSD	735	815	.706	873	637	.653
MT	446	274	.776	462	291	.790
MTSD	290	142	.406	289	129	.495
Subtraction						
RT	1479	888	.867	1803	656	.862
RTSD	670	071	.734	848	590	.584
MT	444	303	.755	453	268	.769
MTSD	278	126	.481	277	142	.591
Multiplication						
RT	1429	691	.739	1761	784	.870
RTSD	567	630	.715	798	699	.813
MT	400	266	.770	446	254	.804
MTSD	246	126	.522	255	131	.637

[a]The reliability coefficient given here is a lower-bound estimate of reliability provided by the squared multiple correlation of a given variable with all of the other 11 variables.