INTELLIGENCE **15**, 437–448 (1991)

Unitary *g*: Unquestioned Postulate or Empirical Fact?

JOHN H. KRANZLER
University of Florida

ARTHUR R. JENSEN
University of California, Berkeley

Carroll (1991) has argued that our empirical test of the hypothesis that psychometric *g* is a unitary factor fails methodologically to prove that *g* is not unitary, and that our finding could have resulted from some impurity in the *g* extracted from 11 psychometric tests. The gist of this argument is that our multiple regression method for testing the unity of *g* and the outcome of this test would be valid only if it were certain that we had a "perfect estimate" of *g* as the dependent variable. We argue that the hypothetical ideal of a perfectly pure *g* is empirically unattainable, but that such purity is an unnecessary condition for testing the hypothesis by the method we used. Our analyses suggest that one would have to assume an improbably large amount of "impure" variance in our *g* factor to make Carroll's argument compelling. Finally, we are most grateful for Carroll's elegant hierarchical factor analysis of our psychometric and chronometric variables. The unity of *g* cannot be proved or disproved by factor analytic methods per se and the unitary *g* hypothesis has only the status of a parsimonious assumption within that framework. But Carroll's factor analysis of our data indeed beautifully represents the relationship between conventional psychometric tests and elementary cognitive tasks based on chronometric techniques and further highlights the central role of efficiency (= speed and consistency) of information processing in *g*.

Carroll's (1991) commentary has two distinct parts: (1) a criticism of our claim that our data and analyses seem to contradict the notion that psychometric *g* is unitary; and (2) a hierarchical factor analysis of the psychometric and chronometric variables in our study.

As for Part 1, we only wish we could believe that Carroll is right, for a truly unitary *g* would be so nice for *g* theory. How beautiful would be an authentic proof of the unity of *g*: a consummation devoutly to be wished indeed! Alas, our study has shaken our hope. But apparently not Carroll's—yet.

As for Part 2, bravissimo! Our article (Kranzler & Jensen, 1991) was aimed solely at testing the hypothesis of unitary *g*, and it was not at all intended to do what Carroll has done with our data in his admirably expert application of the

Correspondence and requests for reprints should be sent to John H. Kranzler, Foundations of Education, 1403 Norman Hall, University of Florida, Gainesville, FL 32611.

most appropriate method of factor analysis. His factor analysis of 38 variables in the combined batteries of psychometric tests and chronometric tests (elementary cognitive tasks, ECTS) displays, probably better than in any previous study, the structural relationship between psychometric and chronometric measures of mental abilities. Also, it beautifully highlights the role of speed and consistency of information processing in general mental ability, or *g*. We are most grateful to Carroll for this contribution to understanding the nature of *g*.

PART 1: CARROLL'S CRITIQUE ON THE UNITY OF PSYCHOMETRIC *g*

Carroll (1991) regards the idea of a unitary *g* as a parsimonious assumption and points out that this assumption has prevailed throughout the history of research and discussion about the *g* factor. There is the notable exception, of course, of Sir Godfrey Thomson's (1951) so-called "sampling theory" of *g*, which he put forth to challenge Spearman's notion of *g* as having a unitary basis. By "unitary" it is meant, in Carroll's (1991) words, that *g* "represents a single entity or constellation in the constitution of the individual that influences a great variety of behaviors and performances, including speed and efficiency of information processing" (p. 434). Thomson's sampling theory, however, never really offered an empirical challenge to the theory of a unitary *g*, because the nature of the population of sampled "elements" was not specified in such a way as to suggest any operations by which the theory could be empirically tested (Jensen, 1987a, pp. 126–128).

An empirically testable challenge to unitary *g*, however, is suggested by modern componential theories of mental ability. Based on concepts from experimental cognitive psychology, componential theory hypothesizes that (1) some quite limited number of basic or elemental information processes enter into performance on any complex cognitive task (such as the items in typical psychometric tests of elemental ability); (2) there are individual differences in these elemental information processes; and (3) the various elementary processes (at least some of them) are independent (i.e., uncorrelated) at the level of individual differences. These processes may consist, for example, of stimulus apprehension, discrimination, stimulus encoding, retrieval or transfer of information between short-term and long-term memory systems, chaining of operations or bits of information, response integration and execution, and the like. Certain elemental processes may be describable in physiological terms: neural conduction velocity, synaptic delay, neural "noise" or errors in the transmission of action potentials, neural redundancy, and so on. It does indeed seem intuitively at least plausible that a number of such processes might underlie performance on complex tests, and furthermore, it seems improbable that just a single factor would account for all of the reliable variance in the various processes.

Considerations such as these help to explain why, in the realm of conventional (i.e., relatively complex) psychometric tests, any method of factor analysis per se has not, and *logically cannot,* either prove or disprove the hypothesis that *g* is unitary, that is, a reflection of individual differences in some single process or property of the brain. The simplest hypothetical example makes this clear. Suppose that among all the various information processes there are three "elementary processes," *A, B, C,* and that every psychometric test (or test item) always involves all three of these processes, whatever other processes or information content they may involve. As a result, even if individual differences in *A, B,* and *C* are independent of one another, the various psychometric tests, in which they are all involved, will be positively correlated with one another to some degree, so that a factor analysis of them will yield a strong *g* factor, whether as a first principal factor or as a second (or other higher-order) factor in a hierarchical analysis. Other factors, too, may emerge, reflecting other processes or information content that is not shared by all of the tests. In terms of this collection of complex tests, then, it would perhaps not seem amiss to speak of its *g* factor as "unitary" as far as the application of any kind of factor analysis per se would allow us to go. But we don't have to stop at that point. We can proceed, in our hypothetical example, to a more fine-grained level of analysis in terms of a variety of ECTs, each of which involves, say, only one or two of the processes *A, B, C* (in addition to any other processes that may or may not also enter into some of the psychometric tests). A common factor analysis (or a principal components analysis) of these ECTs thus could yield two or more uncorrelated components (i.e., sources of variance), reflecting the independent aspects of these processes. If, then, it is found that each of these independent components is correlated with the *g* factor of the psychometric tests, it would mean that the *g* is not really unitary at the level of ECTs, but comprises certain independent components of variance reflecting the uncorrelated aspects of processes *A, B,* and *C.* (Note that this would not be discovered by simply factor analyzing collections of psychometric tests when each test comprises variance attributable to processes *A, B,* and *C.*)

This represents essentially the rationale and method of our study. However, instead of obtaining single, simple correlations between each of the uncorrelated principal components of ECTs and psychometric *g,* we used a statistically more efficient and powerful technique—stepwise multiple regression—to test whether more than one of the ECT components adds a significant increment to the multiple correlation (*R*) with the dependent variable, *g.* It turned out that four independent ECT components each contributes significantly to the prediction of *g.*

So what is the problem exactly? Carroll agrees with the logic of our argument, admitting that our reasoning is valid but makes the assumption that our dependent variable is a "completely adequate and 'pure' measure of *g* (p. 424). To closely paraphrase his argument, if the obtained *g* factor scores are not a perfect estimate of *g,* then any covariance they have with one or more orthogonal measures based on ECT measures can reflect whatever additional variance in the estimated *g* factor scores that make them less than pure, and at least in part overlapping with variance from the ECT variables. Carroll's hypothetical data (in his Tables 1–4) simply illustrate the fact that if two or more orthogonal variables (e.g., ECTs) have some of their variance in common with another variable (e.g., *g*), then if the orthogonal variables (e.g., ECTs) are entered as the independent variables in a stepwise multiple regression, they may add significant increments to the prediction of the dependent variable (e.g., *g*). The difference between Carroll's interpretation and ours is that he attributes this outcome to "impurities" in our psychometric *g* that happen to be measured by two or more of the ECT principal components beyond the first component (which is presumably *g* plus any other variance that the ECT variables do not have in common with the psychometric variables). But is it not also just as plausible that what Carroll views as "impurities" are actually constituents of psychometric *g* attributable to different elementary information processes measured by the ECT variables?

We think it is even more probable that this is the case, for it seems to us most difficult to imagine what the "impurities" could be that would have such surprising properties as the following: (1) the "impurities" would have had to arise entirely from within the psychometric battery, which scarcely resembles the battery of ECTs; (2) these "impurities" would have to constitute more than half of the significantly predictable variance in the psychometric *g*, that is, variance contained in the three principal components of the ECT variables beyond the first principal component; and (3) this assumed "impurity" in the *psychometric g* is contained in just the first few principal components of the 37 ECT variables. Hence, given these properties, the imputed "impurity" in our *g* seems to us implausible. In other studies in which a psychometric battery, such as the Wechsler Adult Intelligence Scale (WAIS) or the Armed Services Vocational Aptitude Battery, has been given to a large group of subjects along with a variety of ECTs, we have found that none of the first-order factors (independent of *g*) extracted from the psychometric batteries shows any significant correlation with the ECTs; *g* seems to be the only psychometric factor that is correlated with the ECTs. It would seem surprising and puzzling, therefore, that the 11 psychometric tests used in our study would harbor some substantial source of variance besides *g*, which also exists in the ECT variables and which has infiltrated the psychometric *g*, making it "impure."

Sources of Error in *g*

Certainly, like any other statistic, a *g* factor based on a limited number of mental tests and a limited number of subjects will contain error. There are three main sources of such error: (1) *subject sampling error*, because a sample does not perfectly represent the population; (2) *psychometric sampling error*, because a limited number of diverse mental tests does not perfectly represent the total population of mental tests, actual or conceivable; and (3) all factor scores in a common factor analysis, including *g* factor scores, by *any* method of derivation, are only *estimates* of the true factor scores, which remain unknown, in the same sense that obtained scores are estimates of true scores, with some determinable margin of probable error, in classical measurement theory. It has been determined mathematically that the average minimum correlation between *estimated* factor scores and their corresponding hypothetical *true* factor scores rapidly increases as a function of the ratio of the number of tests to the number of first-order factors (Gorsuch, 1983, p. 259). With 11 tests and two first-order factors, as in this psychometric battery, the *minimum* correlation between estimated and true factor scores would be +.84, and the actual correlation could be well above this value.

There are several methods, or models, commonly in use for estimating the *g* factor in a correlation matrix, but in any typical battery of 10 or more diverse mental tests, method differences for estimating *g* would be a negligible source of variation in *g* factor scores. The correlations between *g* factor scores derived from different models for extracting a *g* factor from a given correlation matrix typically range above +.97, with a mean close to +.99 (e.g., Ree & Earles, 1991).

As for psychometric sampling error, our *g*, though based on only 11 tests, is probably very close to the general factor in the population of all conventional mental tests, because our battery is composed of Raven's Advanced Progressive Matrices (APM), a good *g* marker, and the 10 subtests of the Multidimensional Aptitude Battery (MAB), which was devised to provide a measure of general ability and to correlate highly with the WAIS, both in terms of the homologous subtests and the full-scale IQ. In general, the psychometric sampling error of *g* is quite surprisingly small. For example, six nonoverlapping test batteries, each composed of nine tests randomly selected from a large pool of extremely diverse tests (used in the U.S. Air Force) showed highly similar *g* factors. Seventeen highly varied "probe" tests were inserted, 1 at a time, into each of the 6 batteries; the *g* loadings of the "probe" tests had an average correlation of +.85 across the 6 batteries (Thorndike, 1987). Hence, we believe that the error or "impurity" in our *g* factor is relatively small, unlike the large "impurities" that Carroll put into the hypothetical data in his Table 1, in which the "impurities" constitute more than half of the dependent variable's communality in some cases. We have no theoretical or empirical basis for supposing that ECTs should predict any large or significant proportion of the sources of error that commonly exist in factors or factor scores. The componential theory of information processing, however, predicts that some independent components measurable with ECTs should be correlated with psychometric *g* (e.g., Detterman, 1987; Sternberg & Gardner, 1982).

Varying the "Purity" of *g*: An Experiment

It is almost axiomatic that increasing the number and diversity of the mental tests in a battery increases the reliability and validity of the estimate of its general factor. In fact, under the assumption that a given test battery is a random sample from a population of tests containing the same factors, it is possible to calculate the correlation between the *g* of the sample and the hypothetical "true" *g* of the population (Kaiser & Caffrey, 1965). For the *g* in our 11-test battery, this hypothetical validity coefficient is about +.90.

Now, if we extract a *g* from some smaller subset of the 11 tests in our battery, we should expect the proportion of true *g* variance in relation to error or "impurity" variance to decrease; that is, there would be a larger proportion of impurity in the *g* based on a smaller number of tests. As the number of tests increases, the *g* should become more "pure," that is, a more reliable and valid estimate of the general factor in our total battery of 11 tests.

So we proceeded as follows:

1. From the battery of 11 psychometric tests, we made up 3 (nonoverlapping) sets, each composed of 3 tests selected at random; and, in addition, we also made up 3 sets each of 5, 7, and 9 tests, randomly selected with the condition that there be minimum overlap between sets.
2. A *g* factor (represented by the first principal factor) was extracted from each of the 12 sets and also from the total battery of 11 tests.
3. Factor scores for each of the 13 different *g* factors were used as the dependent variables in 13 separate stepwise multiple regressions, in every case using the principal component scores derived from the four principal components of the ECTs that contributed significantly to the prediction of *g* in our original study.
4. The squared multiple correlation coefficients obtained from the regression of *g* on the ECT components in each of the 13 sets of tests then were averaged *within* each of the sets of 3, 5, 7, 9, and 11 tests.

Figure 1 shows the correlations of *g* factor scores based on either 3, 5, 7, or 9 tests with the factor scores based on all 11 tests. As the number (n) of tests that go into the estimate of *g* increases, of course, the larger is the correlation of the *g* based on n tests with the *g* based on all 11 tests; that is, the validity (or factor "purity") of each successive *g* increases, going from $n = 3$ to $n = 9$.

Figure 2 (p. 444) shows the average multiple correlation (R) for the regression of *g* based on either 3, 5, 7, 9, or all 11 tests when the independent variables are either the first principal component (PC:1) of the ECTs; or ECT principal components 3, 4, and 5 (PC: 3, 4, 5), that is, all of the ECT components beyond PC:1 that make significant independent contributions to R; or all four principal components (PC:1, 3, 4, 5). The two most noteworthy features of Figure 2 are:

1. The multiple R based on the 3 ECT components (PC:3, 4, 5) beyond the first component (PC:1) is larger than the R for PC:1 at every battery size (n), indicating that ECT components beyond the first principal component con-

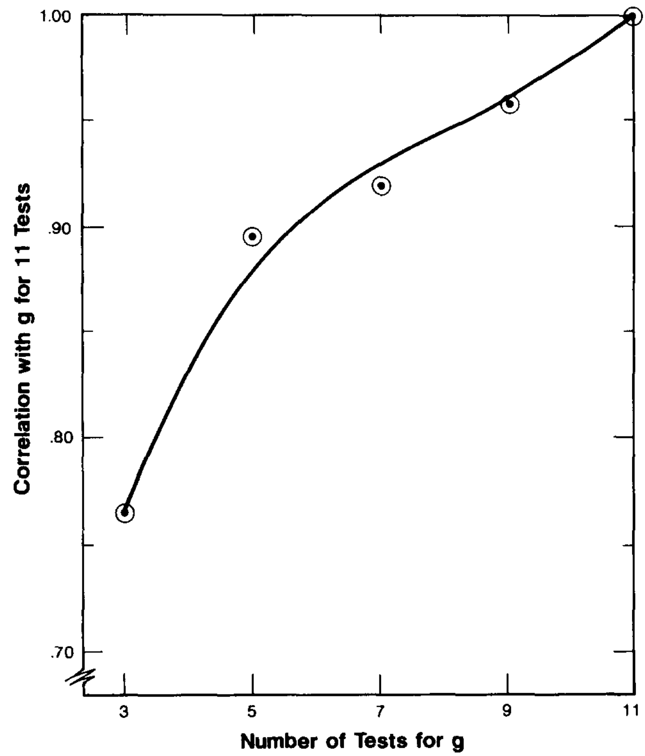

FIG. 1. Correlations between the *g* factor scores based on *g* (first principal factor) extracted from either 3, 5, 7, or 9 tests of the psychometric battery and the *g* factor scores based on all of the 11 psychometric tests.

stitute at least as substantial a proportion of the variance in the psychometric *g* as the first ECT principal component.

2. The multiple correlation coefficient, *R*, for PC:3, 4, 5 *increases* with the number of psychometric tests and hence with the increasing validity (or "purity") of the estimated *g*. The opposite trend should be expected if the ECT PCs beyond the first PC reflected only impurities in the psychometric *g*, which, if such impurities existed, should be a decreasing proportion of the *g* variance as the number of tests increases.

The actual outcome (shown in Figure 2), therefore, would seem more consistent with the hypothesis that the variance in psychometric *g* reflects a number of independent components of information processing, as measured by ECTs, than with the hypothesis that the *g* of our psychometric battery is really unitary but our

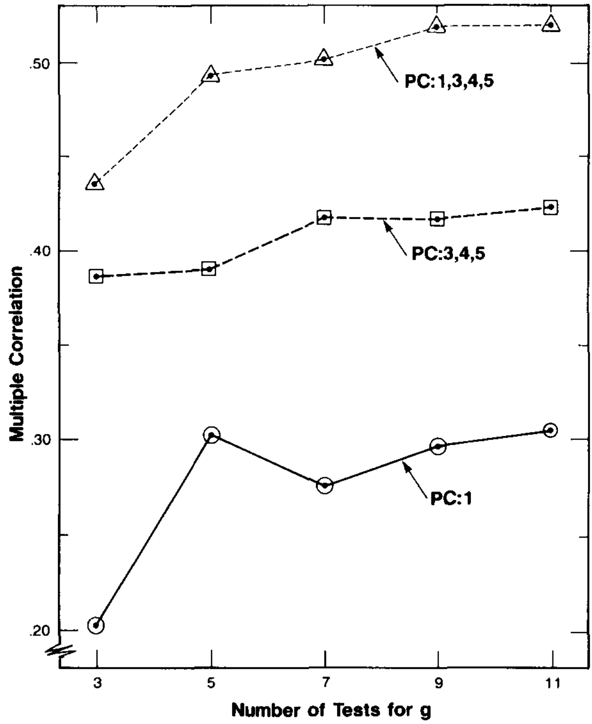

FIG. 2. Multiple correlation (*R*) of ECT principal components with psychometric *g*, when *g* is extracted from either 3, 5, 7, 9, or all 11 of the psychometric tests. The *r* for just the first principal component (PC:1) is contrasted with the *R* for the three significant components excluding the first (PC:3, 4, 5) and with all four components (PC:1, 3, 4, 5) that make significant independent contributions to *R*.

estimate of it contains some error or impurity that just happens to be significantly predicted by three principal components (beyond the first PC) of the ECT battery.

PART 2: CARROLL'S FACTOR ANALYSIS OF THE COMBINED PSYCHOMETRIC AND ECT BATTERIES

This, we believe, is surely the major contribution of Carroll's article, although it has no direct bearing on the main purpose of our study, which was to examine the hypothesis of unitary *g*. The result, presented in Carroll's Table 6, is probably the clearest and theoretically most interpretable factor-analytic representation of the relationship between psychometric and chronometric tests that we have seen anywhere in the published literature. This is largely a result of applying the most appropriate hierarchical factor model to an unusually large and diverse battery of chronometric (ECT) and psychometric measures.

Carroll's insightful discussion of the theoretically most important features that emerge from the hierarchical factor analysis of these variables is here supplemented with a few additional observations. (Carroll's numbering of the factors in his Table 6 are indicated in boldface print.)

1. All of the reaction time (RT) measures and intraindividual consistency of RT measures (RTSD, i.e., the standard deviation of RT across trials) have loadings on the second-order factor [**1**], which may be characterized as "fluid" *g*. The RT and RTSD of the more complex ECTs have higher loadings on this factor than do the less complex ECTs, although even the most complex ECTs have RTs that average less than 1 s.
2. All of the psychometric tests are loaded on the same fluid *g* factor [**1**] as the RT and RTSD measures. The *g* factor extracted from just the psychometric battery is similar to the general factor [**1**] of the combined psychometric and chronometric measures, with a congruence coefficient of .98 between the two sets of *g* loadings. On this fluid *g* factor [**1**], as one would expect, the six nonverbal tests have slightly larger loadings than the five verbal tests.

   Carroll suggests that the correlation of the psychometric tests with the RT measures could have resulted from the fact that the MAB was administered under timed conditions. This conjecture can be discounted by the results of other studies in which psychometric tests administered without a time limit showed highly similar correlations with the same RT variables as in this study. Typically, the most speeded psychometric tests have slightly *lower* correlations with RT measures (as well as lower *g* loadings) than tests given with more liberal time limits (e.g., Jensen, Larson, & Paul, 1988). Studies done specifically with the MAB have shown its correlation with RT measures to be virtually unaffected by whether the MAB is administered under speeded or nonspeeded conditions (Vernon & Kantor, 1986; Vernon, Nador, & Kantor, 1985).
3. The spatial visualization [**4**] and crystallized–verbal [**5**] factors of the psychometric battery come out as separate (orthogonal) first-order factors. It is noteworthy that all of the ECT variables have absolutely negligible loadings on these two factors [**4** and **5**]. This finding is consistent with the finding that the RT in ECTs is correlated with only the *g* factor of the WAIS and has near-zero correlations with all other factors or subtests of the WAIS independent of *g* (Vernon, 1983).
4. Besides having substantial loadings on fluid *g* [**1**], a considerable part of the RT variance of every ECT involving either short-term or long-term memory search and visual search ECT (but *not* RTSD, i.e., intertrial consistency of RT) comes out on a separate first-order factor [**3**], which, of course, is orthogonal to the second-order fluid *g* factor, and the psychometric tests have small, nondescript loadings on this memory search factor [**3**]. It appears that ECTs measuring the speed of visual search and search for information in both long-term and short-term memory involve not only the

general speed-of-information-processing factor, represented here by fluid *g* [**1**], but also involve a separate search–speed factor. This seems to be strong evidence for independent cognitive processes involving speed of processing (cf. Jensen, 1987b).

Similarly, RT and RTSD of simple and choice RT (Hick 0-bit and Hick 3-bit), besides having substantial loadings on the second-order fluid *g* factor [**1**], have part of their variance on a separate first-order factor [**6**], which the Hick paradigm has in common with inspection time (IT). Thus, it appears that a limiting factor to the correlation that the RT of any particular ECT has with psychometric *g* is the fact that some fairly large part of the RT variance on a given ECT is associated with certain features of the measurement paradigm itself, whereas only some part of the RT variance in any particular paradigm represents general information-processing speed, or fluid *g*.

5. Movement time (MT) on all of the ECT variables comes out separately on two first-order factors [**8** and **9**] and a second-order factor, as general MT [**7**]. The general MT factor [**7**] is very clearly distinguished from RT, a fact that justifies the separate measurement of RT and MT in chronometric studies of individual differences. An ECT that amalgamates both RT and MT in a single measure combines two quite different and poorly correlated sources of variance. The MT on all of the ECTs has an average loading of only .17 on the second-order fluid *g* factor [**1**]. Odd-man RT, for example, is loaded .62 on the fluid *g* factor, while Odd-man MT has a *g* loading of only .17. Inexplicably, 3 of the 11 psychometric tests (Information, Comprehension, and Picture Completion) have loadings on the general MT factor (**7**] slightly over .30, which is close to their loadings on the fluid *g* factor [**1**]. Any interpretation of this enigma would be wholly speculative at present.

DISCUSSION

Certainly, the empirical evidence that psychometric *g* is not unitary is far from conclusive. But the assumption that *g* is unitary, we believe, has been brought seriously into question. Carroll has underlined the considerable difficulty in making a strong test of the unitary *g* hypothesis, but without suggesting any better means than the method we tried for seeking an answer to this long-standing and fundamental question. Carroll or someone else might come up with an ingenious method that could yield a more definitive answer. But, so far, we have thought of no other method for approaching it, except to replicate our study with different batteries of psychometric tests, for it seems unlikely that the estimated *g* in every battery would contain the same "impurities" imputed by Carroll, and be predicted by the same ECT components in every case. In fact, there are already two more recent studies that test the unitary *g* hypothesis using our method, with different tests and ECTs but results essentially like ours (Miller & Vernon, in press; Vernon & Weese, 1991).

The very nature of the problem, however, practically precludes actually *proving* that *g* is unitary, for it would be, in effect, a case of trying to prove the null hypothesis. If we had found that only a single principal component of our large collection of ECT variables had a significant correlation with our psychometric *g*, it still could be argued that *g* is *not* unitary, and that our choice of ECTs had not included the right elementary processes or components. If, in a series of such studies based on a wide variety of ECTs, there were repeated failures to find independent components that correlated with *g*, we would have to accept the unity of *g*, not as empirically proved, but as a parsimonious assumption with little chance of empirical contradiction.

Certain assumptions and Occam's razor, of course, have a legitimate and necessary place in theory development. But in science, our assumptions should also be treated, whenever possible, as empirically testable hypotheses, in hope of moving them from their status as assumptions or postulates to the status of facts. We believe it would be an important advance in the theory of mental ability if the nature of *g* as either unitary or componential could, if at all possible, be decided empirically, rather than remain as an arguable assumption.

REFERENCES

Carroll, J.B. (1991). No demonstration that *g* is not unitary, but there's more to the story: Comment on Kranzler and Jensen. *Intelligence, 15*, 423–436.

Detterman, D.K. (1987). What does reaction time tell us about intelligence? In P.A. Vernon (Ed.), *Speed of information processing and intelligence*. Norwood, NJ: Ablex.

Gorsuch, R.L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Jensen, A.R. (1987a). The *g* beyond factor analysis. In J.C. Conoley, J.A. Glover, & R.R. Ronning (Eds.), *The influence of cognitive psychology on testing and measurement*. Hillsdale, NJ: Erlbaum.

Jensen, A.R. (1987b). Process differences and individual differences in some cognitive tasks. *Intelligence, 11*, 107–136.

Jensen, A.R., Larson, G.E., & Paul, S.M. (1988). Psychometric *g* and mental processing speed on a semantic verification test. *Personality and Individual Differences, 9*, 243–255.

Kaiser, H.F., & Caffrey, J. (1965). Alpha factor analysis. *Psychometrika, 30*, 1–14.

Kranzler, J.H., & Jensen, A.R. (1991). The nature of psychometric *g*: Unitary process or a number of independent processes? *Intelligence, 15*, 397–422.

Miller, L.T., & Vernon, P.A. (in press). The general factor in short-term memory, intelligence, and reaction time. *Intelligence*.

Ree, M.J., & Earles, J.A. (1991). The stability of *g* across different methods of estimation. *Intelligence, 15*, 271–278.

Sternberg, R.J., & Gardner, M.K. (1982). A componential interpretation of the general factor in human intelligence. In H.J. Eysenck (Ed.), *A model for intelligence*. Heidelberg: Springer-Verlag.

Thomson, G.H. (1951). *The factorial analysis of human ability* (5th ed.). Boston: Houghton Mifflin.

Thorndike, R.L. (1987). Stability of factor loadings. *Personality and Individual Differences, 8*, 585–586.

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

Vernon, P.A., & Kantor, L. (1986). Reaction time correlations with intelligence test scores obtaine under either timed or untimed conditions. *Intelligence, 10,* 315–330.

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

Vernon, P.A., & Weese, S.E. (1991, April). *The effect of task complexity on the relationship betwee intelligence and speed-of-information processing.* Paper presented at the annual meeting c the American Educational Research Association, Chicago, IL.