Journal of Educational Psychology
1982, Vol. 74, No. 6, 868-873

Copyright 1982 by the American Psychological Association, Inc.
0022-0663/82/7406-0868$00.75

Level I/Level II: Factors or Categories?

Arthur R. Jensen
Institute of Human Learning, University of California, Berkeley

An explanation is sought for the striking apparent failure of the interactions of intelligence and memory factors with socioeconomic status predicted by Jensen's Level I/Level II theory, in a study by Stankov, Horn, and Roy (1980). It is suggested that Level I ability may not be a higher-order factor, like Spearman's *g* or Level II but rather a category of narrower abilities involving rote learning and primary memory, with little transformation between input and output. Orthogonalized hierarchical factor analysis of the abilities domain is recommended as potentially the most fruitful method for discovering the nature of Ability × SES (or Race) interactions.

The recent article in this journal "On the Relationship Between Gf/Gc Theory and Jensen's Level I/Level II Theory" by Stankov, Horn, and Roy (1980) merits a critical comment. These authors (henceforth signified as SHR) investigated Level I/II in a factor analytic framework that seems to highlight certain aspects of the Level I/II theory that have been neglected in previous studies. This is indeed valuable. But what I see as the main contribution of their study is apt to be lost in the article's clutter of side issues, including some confusions about the Level I/II theory and all the previous empirical research on it.

Level I/II is more a set of empirical generalizations than it is a theory, although the attempt to explain these generalizations, which so far has not been uppermost, can loosely be termed a theory. I regard theories merely as tools and scaffolding for the discovery of previously unnoticed phenomena. A theory is useful so long as it serves this purpose, and it is defensible on this ground alone. As the phenomena that a theory reveals increase in extent and complexity, the theory must undergo revision or be discarded and replaced by some other theory, if our aim is to go on probing reality. The danger of a theory is not that the theory is wrong or inadequate in light of further discovery, for that is inevitable and necessary. The danger is that proving the theory to be deficient may be misconstrued as justification for ignoring the phenomena that it has helped to reveal. When the theoretical scaffolding is torn down, are the established empirical findings that are left standing of interest, scientifically or practically, and worthy of attempts at better theoretical formulation? If so, the theory has served its legitimate purpose.

I believe this is the case with the Level I/II theory. A large number of studies, of which my own studies of Level I/II are only a minor fraction, have now established certain phenomena that must be understood by differential psychologists. A recent comprehensive review of this research literature concludes:

> The majority of studies ... provide support for the major hypotheses of Jensen's [Level I/II] theory. Different socioeconomic status (SES) and racial groups tend to obtain equal average scores on Level I, whereas middle SES and white groups obtain higher scores on tests of Level II, on average, than groups of low SES and black individuals. Among the mentally retarded, low SES children consistently outperform middle SES children on measures of Level I. (Vernon, 1981)

Also, the results of other large-scale studies of Level I/II (Hall & Kaye, 1980; Scarr, 1981) too recent for inclusion in Vernon's review are consistent with the major hypotheses derived from Level I/II theory—for example, much smaller race and SES differences on Level I (memory) than on Level II (intelligence).

That the SHR study apparently did not yield results consistent with expectations from Level I/II, as so many other much larger studies have done, is itself in need of explanation. SHR's findings do not "disprove," or even bring into question, the main

Requests for reprints should be sent to Arthur R. Jensen, Institute of Human Learning, University of California, Berkeley, California 94720.

empirical generalizations of Level I/II that are well supported by a host of other studies. But a more careful examination of how and why the SHR results differ from those of other studies might well throw important new light on the Level I/II formulation.

The SHR Study

Essentially, what SHR did was as follows: They began with 12 primary (first order) factors that had already been established in earlier studies. Each of these primary factors is measured by a small group of tests. Three of the primaries involving memory tests were identified as characteristic of Level I ability: *Memory Span* (Ms), measured by auditory and visual number and letter span; *Associative Memory* (Ma), measured by low association word paired-associates, word–number pairs, and free recall of uncategorized lists; and *Meaningful Memory* (Mm), measured by high association word pairs, emphasized word recall, and free recall of categorized lists. The 12 primary factors (including Ms, Ma, and Mm) were factor analyzed, and three second-order factors with eigenvalues larger than 1 emerged. These three second-order factors, obliquely rotated to approximate simple structure, were identified as *Fluid Intelligence* (Gf), *Crystalized intelligence* (Gc), and *Short-Term Acquisition and Retention* (SAR). The Gf and Gc factors are both essentially Level II ability, which I have always considered to be much the same ability as Spearman's *g* factor. Gf and Gc are different (although usually highly correlated) phenotypic expressions of the same genotypic Spearman's *g*. The SAR factor, being defined by tests of short-term memory, is presumably a factor analogue of Level I ability. That is to say, the SAR factor represents mainly the common variance among three primary factors derived from nine various tests of short-term memory. At least two of the primary factors—Ms and Ma—would surely appear to qualify as Level I; their defining tests, or something very equivalent, have all been used as Level I measures in previous studies. In terms of Level I/II theory, however, the Mm (Meaningful Memory) primary is not as clearly defined by tests that would be characterized as Level I, and in fact one of the tests (free recall of categorized lists) was used as a Level II test in one study (Jensen & Frederiksen, 1973), and it behaved as a Level II measure should behave in relation to other variables, in terms of the theory—this, despite the fact that the Mm primary is as substantially loaded on the second-order SAR factor as are the other two primaries, Ms and Ma, that also define the SAR factor. SHR thus wisely dropped the free recall tests in their subsequent analyses because of their ambiguity as a SAR factor.

Factor scores were then obtained on the oblique second-order factors Gf, Gc, and SAR. These factor scores were then used to compare the means of three SES groups after the total sample of 201 high school students was divided into high (n = 45), medium (n = 75), and low (n = 81) SES groups on the basis of father's occupation.

The main prediction from Level I/II theory, which many other studies have borne out (Vernon, 1981), is that the SES groups should differ very much less on Level I ability (as here measured by SAR factor scores) than on Level II ability (as measured by Gf and Gc factor scores). The results of the SHR study, however, did not bear out this prediction. The SES groups differ consistently and significantly on all of the factor scores, and the SES differences are of about the same magnitude on Gf, Gc, and SAR. This striking result is virtually unique, so consistently has this particular prediction from Level I/II theory been substantiated in previous studies comparing SES and racial (white/black) groups on measures of Level I and Level II. Unfortunately, SHR's explanation for this surprising finding is inadequate and inaccurate. It requires a closer, more careful look than they gave to it. It is even possible that they have found something more interesting than they had apparently realized.

False Explanations

SHR suggest that their results are a result of their better, less biased sampling. They state that

> when sampling does not produce a contrast between only one extreme group and a midgroup with respect to

the abilities of interest, the differences between SES groups are significant for LI (i.e., SAR) as well as for LII (i.e., Gf or Gc), and the differences between the differences are neither noteworthy nor significant." (p. 807)

And they then go on to explain the statistical effects and possible artifacts that could result from selecting extreme groups, concluding that

the present analyses and results differ from those of Jensen in one notable respect: Groups are selected at both extremes and in the middle with respect to each of the three major factors. This means that selection at one extreme has not occurred.

They fail to note that only three of my earliest studies of Level I/II made use of contrasting groups selected for low (IQ 60–80) and above average intelligence (Jensen, 1961, 1963, 1968). The four largest and more recent studies, comprising a total of about 5,000 subjects are based on random sampling of whole public school populations without selection on any variables, and yet the results were strikingly in accord with predictions from Level I/II theory (Jensen, 1973, 1974; Jensen & Figueroa, 1975; Jensen & Inouye, 1980). Moreover, the results were fully consistent with those of the earlier studies based on selected extreme IQ groups, indicating that the essential findings of these earlier studies were not merely an artifact of the study design, as implied by SHR. Many other studies besides Jensen's (see Vernon, 1981) have not selected their samples from the extremes on any ability, yet have confirmed theoretical predictions.

Clearly, then, the findings of results consistent with Level I/II theory and the anomalous results of the SHR study cannot be attributed to differences in the methods of sample selection.

Likely Explanations

Sample Characteristics

Before looking at the main difference between previous confirmatory Level I/II studies and the SHR study, a word needs to be said about the subject sample. Although it is much smaller than that of most of the major studies in this vein, and probably too small to provide a fair test of one of the interactions (see next section) from I/II theory tested in this study, the sample sizes cannot be held responsible for the failure of the main hypothesis, that is, the absence of a significant and appreciable Level I/II × SES interaction. However, it is noteworthy that the SES classification into high, medium, and low SES groups in this Australian sample does not result in as large mean SES differences on the Level II factor scorers (Gf and Gc) as are typically found in American samples that are stratified in much the same way. For example, Jensen (1974) found differences between high and low SES groups on Level II (Lorge-Thorndike Verbal and Nonverbal IQ) of about .90σ within large white and black samples of California school children, as compared with a difference of about .64σ in SHR's Australian sample. In Jensen's study the high–low SES differences were about twice as great on Level II as on Level I within both racial groups. But the fact that the Level I/II × Race interaction was so much stronger than the SES effects (within races), in addition to the fact that the theoretical prediction of the regression of Level I on Level II was so much more clearly borne out for race than for SES, suggests that Level I/II theory applies much more to race (i.e., white/black) differences than to SES differences per se. The total empirical literature on Level I/II generally reveals a better track record with respect to race than with respect to SES (Vernon, 1981). Thus, the relatively small Level II differences between SES groups in the SHR study and the absence of a racial contrast, although strictly not defects of the study design, are unpropitious for Level I/II predictions.

Factor Scores

Here is probably the real gist of the explanation for SHR's discrepant results. First, in order to determine adequately the SAR (Level I) second-order factor by at least three defining primaries, SHR rather reluctantly included the Mm (Meaningful Memory) primary. Mm is the least clear factor in the first-order Procrustes factor analysis, having a high loading only on one test (emphasized-word recall). Moreover, none of the Mm tests is really typical of Level I in terms of its core definition but are more

typical of those tests used in past studies that either behaved as Level II with respect to SES (and race) or were ambiguous, shifting in characteristics depending on the age of the subjects, just as we see is the case in the SHR study (e.g., SHR Table 2). The fact that Mm is pulled into the second-order SAR factor could be due partly to common method variance among the various memory tests intended to measure both Levels I and II, rather than to any true Level I communality between Ms or Ma and Mm. Thus, although SAR clearly emerges as a legitimate second-order factor, it seems suspect as a good representation of the theoretical conception of Level I:

> rote learning and primary memory; . . . the capacity to register and retrieve information with fidelity, . . . characterized essentially by a relative lack of transformation, conceptual coding, or other mental manipulation intervening between information input and output. (Jensen, 1974, p. 99)

This suspicion is reinforced when we look at the low degree of independence of the SAR (Level I) factor from the two Level II factors, Gf and Gc. The Kaiser Little-Jiffy oblique factors that generate the three sets of factor scores on which the SES groups were compared show a correlation between Level I (SAR) and Level II (Gf and Gc) of about .70 (as compared with a correlation of .85 between Gf and Gc). This is a much higher correlation between Level I and Level II measures than has ever been found in other studies, in which correlations are typically in the .40s. This indicates that Level I ability, as measured in the SHR study, is not very different from Level II, a condition that could only obscure the predicted Level I/II interactions with SES.

But an even more fundamental question is raised by the high correlation between the second-order factors SAR and Gf or Gc: Is it likely that Level I is not, or possibly cannot be, a second-order factor? Could it be that the second-order common factor among a number of primary factors, each derived from a variety of proper Level I tests, is really the same *g* as the second-order common factor among a number of primaries, each derived from a variety of proper Level II tests? In other words, higher-order factors are all more *g* than anything else, just so long as all of the tests involved are some kind of mental tests, whether classifiable as Level I, Level II, or anything else. For some time I have suspected that this is a likely possibility. For one thing, factor analyses have had little success in determining any broad learning or memory factors independent of the *g* of general intelligence. An old factor analytic study by Garrett, Bryan, and Perl (1935) nicely illustrates this point. A general factor was extracted from a battery of diverse tests of the type that would characterize Spearman's *g*, that is, tests of reasoning and problem solving (i.e., Level II); and from a battery composed mainly of rote learning and primary memory tests (i.e., Level I), a general factor (although a much smaller factor than in the first battery) was extracted. The correlation between the two general factors turned out to be about as high as reliability would permit.

All of this could mean that Level I consists only of a number of quite narrow primary factors, which are Level I because the defining tests truly meet the theoretical conception of Level I. But they have little if any common variance aside from the common *g* of all mental tests. (At a lower level in the hierarchical factorial structure, *g* may break into Gf and Gc.) Level I, then, could not properly be thought of as a broad higher-order factor like *g* (or Gf and Gc), but as a *category* of tests or performances. These Level I tests, when factor analyzed, might yield a number of fairly narrow first-order factors, any one of which, however, would not qualify as Level I if any of its defining tests were not consistent with the theoretical conception of Level I. But they would certainly not be just "arbitrary" tests, as SHR suggests. The category of Level I tasks comprises rote learning and primary memory, with high correspondence between input and output, involving little transformational encoding, mnemonic elaboration, retrieval from previously encoded information, semantic generalization, or conceptual transfer of training. Whether a variety of mental tests meeting these criteria, when factor analyzed among a number of other types of mental tests, yields a clearly determined second-order factor, is not crucial to the Level I/II theory. A number of studies have already shown that a variety of tests with

Level I characteristics (e.g., forward digit span, serial and paired-associate rote learning, and free recall of unrelated items) behave differently than Level II (intelligence) tests in relation to SES and race. The fact that some other types of tests (or factors) besides Level I tests might also be found that do not discriminate SES or racial groups in the same way as do standard intelligence tests is simply further grist for research and theoretical understanding, not a contradiction of Level I/II. I would be the first to agree that when the nature of SES and racial differences in mental abilities (not just in intelligence or *g*) is much better known, the Level I/II theory will most likely be discarded for some better, more comprehensive formulation. But the phenomena to which Level I/II theory has drawn attention, unless proved an outright methodological artifact, will have to be comprehended by any new formulation.

The failure of the second-order SAR factor to replicate the predicted Level I/Level II × SES interaction should have led SHR to investigate whether any one of the purported Level I primary factors (Ms, Ma, or Mm) separately would conform to the theoretical prediction. If Ms, for example, did not, we would be faced with an inexplicable failure of the SHR study to replicate a finding that has been substantiated in much larger studies. And if none of the primary Level I factors accorded with theoretical prediction, would any of the separate tests do so? If not, then the results of the SHR study would seem even more surprising and puzzling. But these further analyses should have been done by SHR to help locate more precisely the failure of the theoretical prediction.

Another procedure that might enlighten these data is an orthogonalized hierarchical factor analysis, as is provided by the Schmid-Leiman (1957) transformation, in which each factor at each level of the hierarchical structure is uncorrelated with every other factor. Then see if factor scores on Level I and Level II types of factors, whatever their level in the hierarchy (if such factors indeed emerge) interact with SES in the way predicted by Level I/II theory.

Incorrect and Weak Inferences

A theory should not be held accountable if incorrect inferences from it are not borne out empirically, or if the theory allows prediction of certain effects that are only very weak and would require an unusually large sample for an adequate statistical test. Such is the case with two of SHR's predictions. First, they expected (p. 805), supposedly from Level I/II theory, that the variances of Gf and Gc (or Level II) should be positively related to SES level. They found just the opposite trend, although the variances did not differ significantly. But their theoretical expectation was the opposite of what Level I/II theory should predict. What they actually found thus accords with the theory. As noted elsewhere (Jensen, 1974), it was hypothesized that social mobility in an industrialized society is more dependent upon Level II than upon Level I abilities. SES is indexed mainly by attained occupational status. High and low prestige occupations differ in the intelligence threshold below which entry into, or successful performance in, the occupation is impossible or highly improbable. Consequently, the variance in intelligence (Level II) should *decrease*, going from lower to higher occupational levels (and hence from lower to higher SES). This, in fact, is what is reported in a number of large studies (Jensen, 1974; Jensen, 1980, p. 344).

SHR also predicted that the variance of SAR (Level I) should increase going from low to high SES. This is theoretically correct, but the theory would predict only a very weak relationship that would not likely be significant in a sample as small as SHR's. The directional prediction is borne out in much larger samples (Jensen, 1974).

SHR warn against interpreting Level I as "a kind of intelligence" (p. 808). I am not aware that I have ever referred to Level I as a kind of intelligence, which I equate with Level II—in *contrast* to Level I. But both, of course, are surely mental abilities.

Future Research

Level I/II theory has successfully identified a class of mental abilities (whether they qualify as broad factors or not) on which social classes and racial groups generally do not differ nearly as much as they differ on tests of intelligence. What is now needed is a more complete analysis of SES and race

differences over a much broader range of mental abilities. I believe that factor analysis is not only the most useful tool for such exploration but is now indispensable for any economy and clarity in such investigation. What we wish to discover is all of the interactions of race (white–black or other racial contrasts) and SES with all of the main *independent* factors of ability. I emphasize "independent," because, as I have argued elsewhere (Jensen, 1980, pp. 729–736), group mean profiles of abilities based on correlated tests (or oblique factor scores) are virtually meaningless and can only obscure the analysis of population differences. Therefore, orthogonalized hierarchical factor analysis, à la Schmid-Leiman (1957), would seem to be the preferred method. The hope remains, as I stated originally (Jensen, 1969), that we may discover other abilities, besides Spearman's *g*, on which the direction or magnitude of SES and race differences are less disadvantageous to the traditionally unfavored groups in school, and which may be used to their advantage in the instructional process.

References

Garrett, H. E., Bryan, A. I., & Perl, R. E. The age factor in mental organization. *Archives of Psychology*, 1935, No. 176.

Hall, V. C., & Kaye, D. B. Early patterns of cognitive development. *Monographs of the Society for Research in Child Development*, 1980, *45*(2, Serial No. 184).

Jensen, A. R. Learning abilities in Mexican-American and Anglo-American children. *California Journal of Educational Research*, 1961, *12*, 147–159.

Jensen, A. R. Learning abilities in retarded, average, and gifted children. *Merrill-Palmer Quarterly*, 1963, *9*, 123–140.

Jensen, A. R. Patterns of mental ability and socioeconomic status. *Proceedings of the National Academy of Science*, 1968, *60*, 1330–1337.

Jensen, A. R. How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 1969, *39*, 1–123.

Jensen, A. R. Level I and Level II abilities in the three ethnic groups. *American Educational Research Journal*, 1973, *4*, 263–276.

Jensen, A. R. Interaction of Level I and Level II abilities with race and socioeconomic status. *Journal of Educational Psychology*, 1974, *66*, 99–111.

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

Jensen, A. R., & Figueroa, R. A. Forward and backward digit span interaction with race and IQ: Predictions from Jensen's theory. *Journal of Educational Psychology*, 1975, *67*, 882–893.

Jensen, A. R., & Frederikson, J. Free recall of categorized and uncategorized lists: A test of the Jensen hypothesis. *Journal of Educational Psychology*, 1973, *65*, 304–312.

Jensen, A. R., & Inouye, A. R. Level I and Level II abilities in Asian, white, and black children. *Intelligence*, 1980, *4*, 41–49.

Scarr, S. *IQ: Race, social class and individual differences: New studies of old problems*. Hillsdale, N.J.: Erlbaum, 1981.

Schmid, J., & Leiman, J. M. The development of hierarchical factor solutions. *Psychometrika*, 1957, *22*, 53–61.

Stankov, L., Horn, J. L., & Roy, T. On the relationship between Gf/Gc theory and Jensen's Level I/II theory. *Journal of Educational Psychology*, 1980, *72*, 796–809.

Vernon, P. A. Level I and Level II: A review. *Educational Psychologist*, 1981, *16*, 45–64.

Received April 10, 1981 ■