Table 4 (Schönemann). *Minimally correlated factor scores and IQ equivalents*

	X	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	IQ	IQ[a]
Galton	.194	−1.278	−.138	−.515	.789	−1.811	.814	87.7	102.9
Binet	−1.695	.784	−.960	−.154	.853	.899	.864	91.9	74.6
Spearman	.281	−1.221	−.354	−.559	.028	.978	.519	91.6	104.2
Wechsler	−.482	.593	1.328	−1.354	−1.647	−.164	.402	96.8	92.8
Guilford	−1.201	−1.146	.452	.704	−.180	.132	−1.878	82.0	82.0
Cattell	.082	1.394	.968	.563	1.216	−.873	−.348	121.7	101.2
Humphreys	.281	.311	−1.021	1.728	−1.490	−.564	.586	101.2	104.2
Harman	1.369	−.164	1.076	.710	.450	1.352	.444	126.6	120.5
Jensen	1.172	.726	−1.352	−1.123	−.019	.050	−1.404	100.5	117.6

[a]Second set of equivalent factor scores which (i) also reproduce the scores y_{ij} under the factor model but which (ii) are minimally correlated with the factor scores in Table 3. The last two columns are equivalent "true g scores" (first columns in Table 3 and Table 4) converted into IQs.

correlation will be. To facilitate interpretation of the discrepancies between the two equivalent numerical assignments on the common "intelligence factor" g, both sets of equivalent g scores were converted into IQs (by multiplying the standard scores by 15 and adding 100). Given his observed test scores (y_{9j} and Jensen's definition of "intelligence", his "true IQ" might be 100, which is average, or 118, which is more than 1 standard deviation above average. Both IQs are equally "true." In this example, nearly half the discrepancies exceed 15 IQ points, that is, 1 standard deviation.

What all this adds up to is that Spearman's attempt to provide us with an objective definition of "intelligence" as g through factor theory was a failure. There have been some recent efforts to define this problem away with additional assumptions (see, e.g., Mulaik & McDonald 1978; Williams 1978), but it is far from clear how these assumptions can ever be tested, and if they can, whether they will hold. All we know at this point is that they have not been tested.

It would be quite unfair to hold Jensen alone responsible for all shortcomings in the field of intelligence testing, just because we do not like his conclusions. Even if one does not agree with Eysenck's (1973) pessimistic assessment that "it would not be entirely inaccurate to say that there are few areas in psychology where so-called 'research' is poorer in quality than here" (p. 480), Jensen does not stand alone as a victim of the facile promises of automatic science by post-Thurstonian psychometricians, as some of the previous Commentary on Jensen (1980b) shows. The indeterminacy issue was first raised by E. B. Wilson in a book review (1928) of Spearman's (1927) *Abilities of Man*, and it was subsequently debated by numerous competent mathematicians, statisticians, and psychologists. These discussions came to an end with Thurstone's appearance on the factor analytic stage, when this and most other problems surrounding the factor model disappeared from our texts and journals. Jensen, for example, does not mention Wilson or any of the other authors who have written on this subject. For a review of the erratic history of the indeterminacy problem see Steiger and Schönemann (1978).

On the other hand, in view of the immense social harm ill-founded "intelligence theories" may have, and, indeed, already have had (see Kamin 1974), it may be well to put on record that, Jensen's claims notwithstanding, we still do not know whether IQ tests measure "intelligence," because the definition of "intelligence" is just as murky today as it was when Spearman (1904) tried to operationalize it more than 75 years ago.

"In truth, 'intelligence' has become a mere vocal sound, a word with so many meanings that finally it has none. . . . test results and numerical tables are further accumulated; consequent action affecting the welfare of thousands of persons is proposed, and even taken, on the grounds of – nobody knows what!" (Spearman 1927, p. 15).

Author's Response

The definition of intelligence and factor-score indeterminacy

Arthur R. Jensen

Institute of Human Learning, University of California, Berkeley, Calif. 94720

Peter H. Schönemann, one of our foremost experts on the mathematical and logical underpinnings of factor analysis, finds cause in my admittedly elementary and nontechnical treatment of this subject, in Chapter 6 of my *Bias in Mental Testing*, to once more trot out his favorite hobbyhorse, known as the indeterminacy of factor scores. He has extensively expounded on this topic now for more than a decade (e.g. Schönemann 1971; Schönemann & Steiger 1976; 1978; Schönemann & Wang 1972; Steiger & Schönemann 1978). It is a mathematical fact that factor scores derived from a common factor analysis are indeterminate in the sense that they are imperfectly correlated with the hypothetical factors. This limitation of factor scores has been recognized by most other factor analysts, but it has not been generally regarded as being as momentous as Schönemann seems to think it is. The indeterminacy of factor scores makes it possible, mathematically, to contrive examples that highlight this property of factor scores, as Schönemann has done here and elsewhere. The issue is not whether Schönemann's demonstration is mathematically correct; we may take it for granted that it is. The real question is its relevance to the current use of factor analysis as an operational tool in research on the nature of intelligence. In this context, I believe, it is largely irrelevant. Also, I believe that most psychometricians would regard as wholly gratuitous and sophistic, at best, Schönemann's attempted implication that admitting factor-score indeterminacy is tantamount to saying that we cannot know what IQ tests measure.

For the sake of brevity, I will couch my response to Schönemann's commentary in a series of points, the order of which is not highly important.

1. Intelligence is not a thing. It is not an "it" in a denotative sense. Intelligence, at this point in the history of psychology, is best viewed as a hypothetical construct or theoretical entity that attempts to comprehend or explain two indisputable empirical observations: (a) the fact of individual differences in "goodness" or efficiency of performance on a wide variety of *cognitive*, or mental, tasks (i.e. tasks on which the main source of variance is not attributable to individual differences in sensory or motor capacities per se), and (b) the fact that performance measures on virtually all such cognitive tasks are positively intercorrelated. A complete explanation of these phenomena would constitute a theory of intelligence. The fact that intelligence is not precisely defined at present does not mean that it cannot be a subject of scientific investigation. Its definition, in any case, will never resemble the definition of an object. A proper definition of intelligence is not the beginning point of research, but will gradually take shape as a product of creative research on the empirical phenomena that suggested the need for such a construct in the first place. The fact that the concept of intelligence is at present not clearly defined is not troublesome. A philosopher of science, Elkana (1974), in tracing the history of the law of the conservation of energy, makes an important point of the view expressed by H. A. Kramers: "In the world of human thought generally and in physical science particularly, the most fruitful concepts are those to which it is impossible to attach a well-defined meaning" (quoted in Elkana 1974, p. 52).

2. The construction of a large number of diverse cognitive tests, each with a wide range of item difficulty and a high degree of item homogeneity (i.e. interitem correlations within each test), and administered to a large representative sample of the general population, will demonstrate (a) reliable individual differences in performance (i.e. test scores), and (b) all-positive intercorrelations among all the diverse tests, known as a *positive manifold*, which is a central empirical fact for research on human mental abilities.

3. The first (unrotated) factor extracted from a matrix of intercorrelations among a number of diverse cognitive tests by means of common factor analysis (or principal factor analysis) is a way of quantitatively summarizing the phenomenon of the positive manifold. The loadings of each of the various tests on this first principal factor indicate the degree to which each test measures the largest source of variance common to the various tests in the correlation matrix. Hence, a principal factor analysis helps researchers to select those particular tests that best measure whatever all of the various cognitive tests measure in common. The first principal factor is the largest common factor, and the tests that are most highly loaded on it come closest to representing the phenomenon of most central interest in the study of human abilities, namely, the common factor, which Spearman labeled g (for "general") factor, that "accounts for" the positive manifold in the domain of ability tests.

In this respect, the first principal factor may be regarded as a good "working definition" of intelligence. By "working definition" I simply mean that this definition helps us to focus our research on those kinds of psychometric tests or laboratory measurements that are most representative of the common source of variance in any large collection of diverse cognitive tests.

For this purpose, which does not require the computation (or estimation) of factor scores, but simply the observation of the pattern of factor loadings on the various tests, there is no real problem with common factor analysis. As Schönemann and Wang (1972) note in discussing the problems of factor-score indeterminacy, "The point can be made that none of these problems bear on the practical utility of factor analysis as a research tool as long as it is used to study the structure of variables without attempting to estimate a person's factor scores" (p. 89). In *Bias*, I specifically stated that "if we wish to obtain factor scores on a number of factors, principal components analysis is preferable" (p. 258). The factor-score indeterminacy problem does not extend to "factor scores" (more precisely, "component scores") computed from a principal components analysis, which are completely determinate.

4. The problem of factor-score indeterminacy that concerns Schönemann applies only to common factor analysis (or principal factor analysis), and not to principal components analysis at all. Although principal factor analysis and principal components analysis represent different mathematical models, practically speaking, their results are in fact usually (but not necessarily) highly similar, especially in their representation of the most general, or g, factor of the matrix of intercorrelations among a collection of various mental ability tests. The patterns of test loadings on the first principal factor and on the first principal component are virtually always very highly correlated for real data, although one could probably contrive artificial data in which the correlation would be appreciably lower. In dozens of test batteries that I have subjected to both principal factor analysis and principal components analysis, the simple correlation (Pearson r) between the loadings on the first principal factor and first principal component has always been about .95 or higher, and congruence coefficients (an index of factor similarity) are usually close to .99. (It is true, however, that successive unrotated factors and components beyond the first show decreasing resemblance for each successive factor.) Thus, if our "working definition" of intelligence is the first principal factor, the estimated factor scores, despite their theoretical indeterminacy, are correlated something close to .99 with the component scores based on the first principal component, and principal components do not share the indeterminacy of principal factors. So, if all one is concerned with is the general factor, in practice it makes an utterly negligible difference whether one uses principal factor analysis or principal components analysis, or whether one estimates factor scores on computed, perfectly determinate component scores. To satisfy the purists who, from one theoretical standpoint insist on principal factor analysis, and to satisfy those who, for a different set of reasons (mostly mathematical) favor principal components analysis, I have often performed both types of analysis on the same data. Never is there more than an iota of difference in the results, either in the pattern of factor (or component) loadings, or in the factor (or component) scores, or in any theoretical or practical conclusions one would draw from the analysis. To me, as

an empirical researcher, this illustrates the practical triviality of Schönemann's complaint, whatever mathematical rectitude may be claimed for it.

5. For some years, researchers in the field of human abilities have been generally agreed that while factor analysis is most useful for identifying the largest common factor, or *g*, in the diversity of cognitive tests, and indicates those tests that are the most loaded on *g*, it cannot, by itself, reveal or explain the *nature* of *g*. To accomplish that, investigation must proceed beyond factor analysis to the experimental analysis of cognitive tasks that have been specially selected or devised to test specific hypotheses about the nature of the mental processes elicited by the tasks. The particular tasks are intended to be more amenable to experimental analysis than are most of the items in ordinary IQ tests, and the degree to which the experimental tasks reflect the largest common factor in diverse cognitive tests can be shown by its *g* loading in a factor analysis with standard reference tests. The shift in emphasis from the factor analysis of tests to the experimental analysis of cognitive processes is the main thrust of current research on intelligence. This modern information-processing approach to the study of intelligence is well exemplified in recent books edited by H. J. Eysenck (1982) and R. J. Sternberg (1982). [See also Sternberg: "Sketch of a Componential Subtheory of Human Intelligence" *BBS* 3(4) 1980.]

The *g* factor also gains in importance as a theoretical construct through findings of its relationship to phenomena outside the realm of psychometrics, such as the finding that psychometric *g* is correlated with brain-evoked potentials and with choice reaction time (Hendrickson & Hendrickson 1980; Jensen, Schafer & Crinella 1981) and the finding that the genetic phenomenon known as inbreeding depression affects various mental test scores in varying degrees for different tests, depending largely on the size of the test's *g* loading, there being a direct relationship between inbreeding depression and *g* loadings (Jensen 1983).

6. Schönemann's dragging IQ tests into his discussion of factor-score indeterminacy is quite gratuitous and merely reflects his evident antipathy toward IQ testing. The fact is that none of the most widely used individual or group IQ tests today is based on factor analysis and the IQ scores they yield are not factor scores. The intended implication that, because of factor-score indeterminacy, whatever IQ tests measure is arbitrary or meaningless, is false. The fact that a wide variety of different IQ tests measure a common factor is attested to by the high correlation between the tests, mostly in the range from .70 to .80, which is what would be expected if the tests' *g* loadings were in the range from .80 to .90 (Jensen 1980a, pp. 314–16). What IQ tests measure gains its meaning not from lexical definitions or from factor analysis, but from the substantial correlations between IQ and numerous nonpsychometric variables, which I have reviewed in extenso elsewhere (Jensen 1980a, Chapter 8). As Miles (1957) has noted, "The important point is not whether what we measure can appropriately be labeled 'intelligence,' but whether we have discovered something worth measuring" (p. 157). The field of applied psychometrics provides ample evidence that we have indeed discovered something worth measuring (Jensen 1981).

7. Schönemann's sparse bits of quotations from such prestigious psychologists as Eysenck and Spearman appear to serve Schönemann's own message very well, but do they really express the intended meaning of their authors? The quotation from Eysenck, read in its full context, conveys a considerably different meaning from that suggested by Schönemann's use of Eysenck's words – as does, I believe, the quotation from Spearman, in which, readers should note, the ellipsis represents a deletion of no less than 160 of Spearman's words! Is it just Spearman's name, rather than his intended meaning, that graces Schönemann's own sentiments about IQ testing?

References

Borg, I., ed. (1981) *Multidimensional data representations: When and why*. Mathesis Press. [PHS]

Elkana, Y. (1974) *The discovery of the conservation of energy*. Harvard University Press. [ARJ]

Eysenck, H. J. (1973) *The measurement of intelligence*. Williams and Wilkins Company. [PHS]

ed. (1982) *A model for intelligence*. Springer. [ARJ]

Hendrickson, D. E. & Hendrickson, A. E. (1980) The biological basis of individual differences in intelligence. *Personality and Individual Differences* 1:3–34. [ARJ]

Jensen, A. R. (1980a) *Bias in mental testing*. Free Press. [ARJ, PHS]

(1980b) Multiple book review of *Bias in mental testing*. *Behavioral and Brain Sciences* 3.325–71. [PHS]

(1981) Test validity: *g* versus the specificity doctrine. Invited address at the annual convention of the American Psychological Association, Los Angeles, Calif. [ARJ]

(1983) The effect of inbreeding depression on mental ability factors. *Personality and Individual Differences* 4.71–87. [ARJ]

Jensen, A. R., Schafer, E.W.P. & Crinella, F. M. (1981) Reaction time, evoked brain potentials, and psychometric g in the severely retarded. *Intelligence* 5 179–97. [ARJ]

Kamin, L. J. (1974) *The science and politics of the IQ*. Lawrence Erlbaum Associates. [PHS]

Miles, T. R. (1957) Contributions to intelligence testing and the theory of intelligence. I. On defining intelligence. *British Journal of Educational Psychology* 27.153–65. [ARJ]

Mulaik, S. A. & McDonald, R. P. (1978) The effect of additional variables on factor indeterminacy with a single common factor. *Psychometrika* 43:177–92. [PHS]

Schönemann, P. H. (1981a) Factorial definitions of intelligence: Dubious legacy of dogma in data analysis. In: *Multidimensional data representations: When and why*, ed. I. Borg, pp. 325–74. Mathesis Press. [PHS]

(1981b) Power as a function of communality in factor analysis. *Bulletin of the Psychonomic Society* 17:57–60. [PHS]

Schönemann, P. H. & Steiger, J. H. (1978) On the validity of indeterminate factor scores. *Bulletin of the Psychonomic Society* 12:287–90. [PHS]

Schönemann, P. H. & Wang, M. M. (1972) Some new results on factor indeterminacy. *Psychometrika* 37:61–91. [ARJ, PHS]

Spearman, C. (1904) General ability, objectively determined and measured. *American Journal of Psychology* 15:201–93. [PHS]

(1927) *The abilities of man*. MacMillan. [PHS]

Steiger, J. H. & Schönemann, P. H. (1978) A history of factor indeterminacy. In: *Theory construction and data analysis in the behavioral sciences*, ed. S. Shye, pp. 136–78. Jossey-Bass. [ARJ, PHS]

Sternberg, R. J. (1982) *Advances in the psychology of human intelligence*. Lawrence Erlbaum Associates. [ARJ]

Williams, J. S. (1978) A definition for the common factor analysis model and the elimination of problems of factor score indeterminacy. *Psychometrika* 43:293–306. [PHS]

Wilson, E. B. (1928) Review of "The abilities of man, their nature and measurement," by C. Spearman. *Science* 67:244–48. Reprinted in *Multidimensional data representations: When and why*, ed. I. Borg, pp. 375–79. Mathesis Press, 1981. [PHS]