Multivariate Behavioral Research

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hmbr20

Commentary on Guttman: The Irrelevance of Factor Analysis for the Study of Group Differences

Edward E. Roskam & Jules Ellis

Published online: 10 Jun 2010.

Multivariate Behavioral Research, *27* (2), 205-218

Commentary on Guttman: The Irrelevance of Factor Analysis for the Study of Group Differences

Edward E. Roskam & Jules Ellis
University of Nijmegen

Louis Guttman's last article is intriguing, controversial and argumentative, and stimulates the reader to a critical response. Guttman possessed deep insight and profound knowledge, and the senior author of this commentary is very grateful for what he taught him through his writings and in discussions. The significance of many of his statements are yet to be appreciated in their full depth. In advance, we wish to stipulate that Guttman's insisting on facet design for defining a domain, that is a universe of content, is of great methodological significance, and that domain definition is necessary for the definition of theoretical constructs, notably traits, and for testing of theoretical hypotheses. Yet, we can not conceal that Guttman appears to have grossly misjudged the scientific value of his most cherished intellectual creation: facet theory.

In summary, we find that Guttman's article does not refute *g*, and essentially, does not address itself to the question of group differences, simply because facet theory as presented is not able to address that issue. Our comment will be organized around four issues raised in Guttman's article: (a) What have similarity structure analysis (SSA) and Facet Theory to say about *g*? Answer: nothing; (b) What are the implications of the "first law of intelligence"? Answer: it implies *g*; (c) Can response speed open a way to study the basis of group differences in *g*? The answer is: yes, but neither SSA nor factor analysis (FA) appear appropriate to do so; (d) What is the value of the missing theorem? Answer: it appears a tautological consequence of the assumption that only one factor determines group differences, and Jensens second hypothesis is not a tautology. Guttman obviously ignored that Spearman's (1927) theory was a two factor theory and that the tetrad condition, even if it is satisfied in subpopulations, does not imply that group differences on tests reflect only differences in *g*.

The irrelevance of factor analysis does not refute *g*, but Spearman's (1927) two factor theory refutes factor analysis in the presence of group differences.

The authors may be reached at Nijmegen Institute for Cognition and Information, University of Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands, by phone at (+31) 80-612537, or by Bitnet at U212753@HNYKUN11.

The Definition and Existence of g

Whether or not *g* exists is to be decided empirically. Such a decision is not possible without properly stipulating which empirically testable criteria have to be satisfied to warrant the conclusion that *g* exists. Although theoretical variables can not be defined apriori, a meta-theoretical definition should precede its empirical identification. First, *g* is a trait, and should satisfy the criteria of a trait, that is: it should refer to unidimensional individual differences vis-a-vis a universe of content. Secondly, it should satisfy identifiability and invariance, that is: it should always be possible to assess it, and what is identified as *g* should be the same thing in every population and in every subset of items from the universe of content; if a law applies to it, that law should be a universal law. Spearman (1927, p. 75): "that which it [*g*] measures has not been found by declaring what it is like, but only by pointing out where it can be found". Science is constantly in search of concepts which unify and integrate empirical data, and inappropriate definitions may need to be replaced by more adequate but essentially similar ones.

Evidently, *g* as defined by Spearman's tetrad condition does not exist, but Spearman was aware that the tetrad condition can be spoiled by tests of which the specific components overlap (Spearman, 1927, p. 80, p. 150 ff). Speaking metaphorically: "It [*g*] forms a mighty factor in the state, but not the sole one" (Spearman, 1927, p. 84). This statement of Spearman's does not seem too different from Jensen's (1985, p. 194): "the single largest independent source of individual differences that is common to all mental tests". Defined as the first principal factor of intelligence tests, it may appear trivial that it exists, but that conclusion ignores the requirement that this principal factor should be also the same and assessable in the pertinent subpopulations, and present in all mental tests. Spearman's two-factor theory may be correct, but FA may fail to demonstrate it, as will be shown in our discussion of the "missing theorem".

Furthermore, if all intelligence tests satisfy the condition of a positive manifold, and under the same proviso of invariance and identifiability, there can be no rational objection against identifying *g*, for instance, with the test of which the minimal correlation with any other test from the same domain is maximal, if that test is also homogeneous in the sense that it reflects a single trait with respect to its subuniverse of content. If that definition is adopted, Guttman's rule inference in manual manipulation would be a serious candidate. The point here is that a metatheoretical definition of *g* is conceivable which fits into the cylindrical structure of facet theory, and the regionality theory does therefore not "exorcise" *g* (p. 27) (or, at least, no more and no less than does violation of the tetrad condition).

So, the best conclusion so far can be that *g* in the sense of Spearman's (1927) tetrad condition can not be identified, but other definitions can not be dismissed. The first issue to be addressed should concern the metatheoretical meaningfulness of the definition(s). If that can be satisfied, what remains, then, of course, is how to assess it, and whether or not there exist consistent group differences with respect to this *g*, and what the basis of these differences is (sociological, biological, hereditary or not, etc.). This question is unanswered to a large extent.

Essentially, it seems, Guttman claims that intelligence as defined via its facet definition is intrinsically multidimensional. We do not contest that definition. The question, however, is whether that universe of content of intelligence exhibits a particular empirical structure which refutes *g*. Alternatively, one might focus on the single trait that is common to performance on all intelligence test items (if it exists), despite uniquenesses involved in single items (or tests). The latter was obviously Spearman's (1927) contention, and followed by Jensen (1985).

SSA and g: The Regionality of Intelligence Tests

In the same way as factor analysis (FA) of intelligence tests yields a fairly recurrent pattern of loadings, so does Smallest Space Analysis (SSA) show a fairly recurrent regionality. This can come as no surprise, as both use the same correlation matrix as the basis for data analyses. FA and SSA show the same structure but in a different perspective. Ignoring, for the sake of argument, the fact that SSA uses only the ordinal information in the correlation matrix, and taking some lack of fit in either analysis for granted, a computer can calculate the factor pattern from an SSA pattern, and vice versa. What, then, is the value of SSA over FA? Any such value can only be that the one technique provides more definite answers to substantively more rational and meaningful hypotheses than does the other, and at the same time makes fewer or less doubtful assumptions. We believe this was indeed Guttman's motivation to develop SSA, particularly since he found that ordinal patterns in correlation matrices reveal relations such as the simplex, the circumplex and the radex, which are indeed obscured by factor analysis. Moreover, he was able to link these patterns to psychologically relevant facets of tests. This approach, or methodology, then, allowed a direct partnership between the definition of the domain (or universe of content) and an aspect of its empirical structure which had never before been obtained so clearly by any other method of data analysis.

However, SSA per se has no provision to deal with group differences. Though Guttman (p. 177) promises to "suggest a proper way of studying group differences over the universe of items", he does not do so. It may be argued that if factor scores distributions are different in different populations, the correlation matrices will be different, and hence SSA will show different patterns, and so is indirectly sensitive to group differences. But there is no explicit way for SSA to look into group differences. Since SSA uses only the ordinal information in the correlations matrix, it may even throw away crucial information. The structure of intelligence may be qualitatively the same in different populations, yet the distribution of factor scores may be different in different populations. That may show in the values of the correlations among tests. Furthermore, next to nothing is known about sampling error and it is our impression from reading and hearing reports of SSA analyses that different results in different populations are ignored when the regional structures are roughly the same. And, also, like simple structure according to Thurstone's (1935, 1947) original definition is hardly ever found in its pure form, so a radex or cylindrex is hardly ever found in pure form. This need not, in itself, be disturbing, since pure forms are anyway hardly ever obtained empirically, but it does not seem just to dismiss simple structure for this reason and not judge SSA by the same standards. Moreover, since the issue is group differences, it does not seem just to reject FA in favor of SSA which offers no way of dealing with group differences on latent traits.

The least that should be demanded from a method of data analysis is that it is capable of confirming an hypothesis if that hypothesis is true. If the method is not capable of doing that, its results can never be used as evidence against that hypothesis. The hypothesis should be formulated in terms of the data analysis to be used. By Guttman's own words (1981b, p 63): "a theory which is not stated in terms of the data analysis to be used, can not be tested". Guttman did not formulate the hypothesis that *g* exists in terms of SSA. So, unless someone as yet does so, SSA has nothing to say about *g*. Guttman might have implicitly said that an hypothesis should be formulated in terms of SSA, or else have no meaning, but that would indeed put the cart before the horse.

The FA versus SSA controversy has little to do with the use and usefulness of the facet design technique for characterizing domains. Domain definitions are always necessary, and we fully agree with Guttman's position that domain definition, theory, and data analysis should be firmly linked. We do not agree, however, that the domain definition should be linked exclusively to one kind of theory or one kind of data analysis. Theories must refer to the domain definition, and posit the structure of the data in terms of the data analysis to be used, but the domain definition can and should neither preempt the theory, nor the data analysis. Domain definition, theory, and data analysis are like a trias politica: each needs the other to function meaningfully, but neither has power over the other.

What Can SSA or FA Show?

Assume, for the sake of argument, that any set of items from the universe of content as defined by Guttman, satisfies a unidimensional latent trait model, possibly with specific factors not affecting the correlations. That trait would then be *g*. What would happen if correlations between these items were calculated and analyzed by SSA or FA, respectively? We will not go into details, but point out a few facts, which should be widely known. If the items satisfy a (unidimensional) scalogram, the product moment correlations with properly estimated communalities will yield $n/2$ factors (Dubois, 1960). If the items satisfy the normal ogive model, the matrix of tetrachoric correlations is unifactorial (Lord, 1980). The phi-correlation matrix for that case will in general not be unifactorial (McDonald & Ahlawat, 1974). In neither case would FA of product moment correlations reveal *g*.

Suppose now that the phi-correlations in case of a perfect scalogram are analyzed by SSA. The result will be a unidimensional simplex, so in that case, SSA will give the correct answer. Suppose that tetrachoric correlations satisfy the tetrad condition, and so confirm *g*. A limiting case is the case that all tetrachoric correlations are equal (which would be the case if the discriminative powers of all items are the same). SSA would have no solution for this case. Either a degenerate solution fits the correlation matrix, or any arbitrary solution, or a solution with all points equidistant in n-1 dimensions.[1] We don't know what will happen in the general case when a correlation matrix satisfying the tetrad condition is analyzed by SSA, but our conjecture is that the outcome is definitely not unidimensional.

The Common Range and g

On page 179, Guttman writes: "definitional commonality of range provides part of the rationale for the positiveness of the regression slope". We fail to see how a definition can justify or provide a rationale (explanation?) for an empirical relation, except in so far as stating the empirical relation requires that its domain is defined so as to render the statement semantically consistent.

On page 180, Guttman writes: "Instead of focussing on the common range, he [Spearman] proposed his *g* common-factor hypothesis as an algebraic rationale for the phenomenon [of positive covariance]. His emphasis was more on algebra than on content". This statement is cryptic, and suggests again that the common range serves as a content-rationale for the phenomenon of positive covariance. If this is indeed what Guttman wanted to say, he appears to be saying that the common range accounts for positive covariance. Was it this that also motivated Guttman to force response speed into the same common range,

[1] The solution depends on the details of the algorithm and on the choice of the primary versus secondary approach to ties. Kruskal's Stress formula (which is different from the one implemented in SSA) with the primary approach will yield the arbitrary solution, and with the secondary approach will yield a solution in n - 1 dimensions.

and so account for monotonicity[2] of correctness with response speed? But how can the fact that the response range is correct versus incorrect entail monotonicity? Is monotonicity not an empirically falsifiable hypothesis, which, being falsifiable, can not be accounted for by the common response range, which is definitional? By Guttman's own words: "a mapping sentence is not a theory" (1981a, p. 34).

The profound virtue of the mapping sentence for intelligence tests items is that it defines the universe of content. It thereby constitutes the empirical reference for theoretical statements. "A theory is an hypothesis of a correspondence between a definitional system for a universe of content and an aspect of its empirical structure, together with a rationale for such an hypothesis." (1981b, p. 50). The rationale for the hypothesis of positive covariance may very well be that a common trait (intelligence, or *g*) is involved in the correctness of the responses to any item from that universe of content. It may even be that whenever the common response range is correctness, the covariance will be positive. But the referent of a statement can not be the rationale for that statement. If that would be essential to Guttman's facet theory, then facet theory is unscientific. (It is possible that Guttman conceived of the mapping sentence as an analytical statement; his use of the term a priori is suggestive of this interpretation. However, that would confound definition with hypothesis, and make facet theory tautological).

The "First Law" and g

Although Guttman has always been extremely critical in analyzing the mathematical consequences of a mathematical statement, he seems never to have analyzed the implications of the first law of intelligence mathematically. Recently, Jules Ellis (1992; Ellis & VandenWollenberg, 1992), did so, and his conclusion has unexpected implications.

We assume that the first law is meant to be an inductive generalization in the first place, which is then held to be a universal truth and is therefore called a law. As inductive generalization, the first law seems hardly contested: indeed, all intelligence tests appear to correlate positively, and any exception can be easily attributed to sampling error and/or adverse measurement error. Its merit is that it stipulates the fact of positive monotonicity with reference to a well defined universe of content. However, there is more to it.

Since statistics must be taken into account, we presume that the first law refers to the statistical expectation of the monotonicity coefficient. Double

[2] More properly formulated: across subjects, expected or average correctness on a (set of) item(s) is monotone with expected or average response time to a (same or different) (set of) item(s).

stochasticity is involved: sampling error, and measurement error. Statistical expectations should be taken over both measurement error and sampling error with respect to the actual population being studied. Taking an entire population, how ever defined, sampling error is not involved. Moreover, if the first law is meant to be a universal truth, it should hold in every population, if not selected artificially. Guttman was obviously aware of the fact that one may artificially construct populations where the regression between two items is negative. However, the term *not selected artificially* is not quite clear, and we take it that any population picked at random is not artificially selected.

Ellis' (1992; Ellis & VandenWollenberg, 1992) argument runs as follows. First, dichotomously scored items would merely be a special case and satisfy the law. For that case, the issue of linearity of slope versus monotonicity is not relevant. Secondly, if the law is a universal truth, it should be true in any randomly selected population, regardless of its size. It may be that that was not meant by Guttman, but then it should not have been stated as a universal law, or should not have been stated for any population.

Now, let X_{vi} be the observed score variable of subject v on item i, and let τ_{vi} be its expectation, or true score. Let τ_v be the sum of the true scores on all items in the universe of items. Ellis first points out that, if the first law holds in every (sub)population regardless of size, and taking two arbitrary subjects v and w, $\tau_v = \tau_w$ implies $\tau_{vi} = \tau_{wi}$ for every i. The proof is simple. If $\tau_v = \tau_w$ but τ_{vi} were unequal to τ_{wi}, for instance $\tau_{vi} < \tau_{wi}$, then there must exist an item j such $\tau_{vj} > \tau_{wj}$. But then $\mathrm{Cov}(X_i, X_j) < 0$ in the population $P = \{v, w\}$, contrary to the first law. Next, if $\tau_v > \tau_w$, there must exist at least one item such that $\tau_{vi} > \tau_{wi}$. Suppose that for some item, k, $\tau_{vk} < \tau_{wk}$, then $\mathrm{Cov}(X_i, X_k) < 0$ for $P = \{v, w\}$, contrary to the first law. Therefore, if $\tau_{vi} > \tau_{wi}$, it is necessarily true that for any other item k, $\tau_{vk} \geq \tau_{wk}$. Therefore[3], for any item i, τ_{vi} is a non-decreasing function of τ_v. In other words, if the first law is a universal law, there exists a score τ for every subject such that the expectation of the score on any item is a non-decreasing function of it. But this is tantamount to unidimensionality, or g.

Furthermore, Ellis proves that if the first law is true, there exists a (sub)population where SSA of items is not unidimensional.

So, if intelligence is not unidimensional, the first law can not be a universal law, and if the first law is universally true, and hence intelligence is unidimensional, SSA may not show it.

An alternative to the First Law might be that the probability of positive monotonicity among intelligence items increases with the size of the population observed, but the implications of that are not obvious and it may not be easily

[3] A related but weaker theorem was given by Rosenbaum (1984). The proof as given here assumes that the set of items is finite, otherwise no finite value of τ as defined here exists. Ellis has also given a proof which is valid for an infinite but denumerable set of items.

testable. Issues of sampling, and particularly of representative and of random sampling are much more intricate and precarious than many social scientists are aware of. Guttman may have surmised the problem as he added the condition that the population is not artificially selected, but that same condition may very well beg the question. Statistical regularity can not be a universal law if that regularity is population dependent.

Reaction Times and g

Guttman follows Spearman (1927) in his statement that "goodness" and "speed" are positively correlated. We believe this is true, but the issue is by far not as simple as that, and unfortunately, Guttman seems to ignore the vast literature on reaction times (Berger, 1982; Brand & Deary, 1982; Luce, 1986; Pachella, 1974). Precise definitions are needed. Response speed is an inaccurate term; response time is the accurate term. Across experimental conditions, the correlation between mean response time and percentage correct responses to simple mental tasks, is positive. This is called the macro trade-off, or speed-accuracy trade-off function, SATF. Within experimental conditions, the conditional probability of a correct response, given the response time, is called the micro trade-off (or conditional accuracy function, CAF), and both increasing, decreasing, and stationary CAF's have been found. Our tentative hypothesis is that for common intelligence tests, the probability of a correct response increases with mental processing time, that is, the more time the subject invests in solving the item, the larger the probability of a correct response. In one experiment, using experimenter-controlled inspection time (rather than process controlled response time) it was found that a subject parameter (mental speed for the pertinent task), an item-parameter (difficulty), and inspection time, each contribute independently to the probability of a correct response (Roskam, Van Breukelen, & Jansen, 1989; Van Breukelen, 1989).

Putting the pieces of the *g* and response time puzzle together requires a testable theory which states how a number of determinants affect the correctness and the speed of the response. Roskam (1983) has outlined a tentative theory (not unlike Jensen, 1985, p. 196-197) which stipulates, among other things, that mental speed and mental resources determine the time to respond with a given probability of being correct. Mental speed may be defined as the amount of correctness gained per unit of mental processing time, and it can be considered as the number of chunks of mental information processed per unit of time; and mental resources refers to the information per chunk. (For example, a subject may know as one chunk that $(a + b)^2 = a^2 + 2ab + b^2$, or may have to expand the expression element-wise). Acknowledging that both mental speed and mental resources determine response time, that SAT may vary between items and depend on the instruction given to the subject and on his test-taking attitude, it should come as no surprise that correlations between response time and correctness are not easy to interpret.

This little digression served to make it clear that response time is a response facet, alongside correctness. The items do indeed, as Guttman pointed out, satisfy the domain definition of intelligence items, but response time can not be entered as an item facet. What can be entered as an item facet is inspection time, but the (assumed) positive correlation between inspection time and correctness is of a totally different nature than the relation between mental speed (as an explanatory factor) and correctness. A construction as proposed in Guttman's article, namely: "How correct is the answer given by subject *s* within *x* seconds" is not a proper construction to serve partnership between domain definition and theory development. Guttman, apparently, wished to force the response time into the domain definition as an item facet so that the first law would apply to it, and it would also show up in SSA. Precisely what that means is obscure: correlations between correctness of responses to items presented with varying inspection times are likely to be positive, but that is not the issue. The issue is a correlation between response speed and correctness. Stating that reaction time tests are but a further variety of intelligence tests sounds like a willful attempt to ignore differences in mental speed as a potential explanation for differences in performance.

Explanations have to go beyond the data in the sense that they look for underlying processes, and mental speed is a serious candidate to explain individual differences in performance on intelligence tests. A necessary condition to make this a potential fruitful venue, is to look into the black box, or split the atom (Eysenck, 1982), by developing and testing theories about the relation between response time and correctness which goes beyond mere monotonicity and beyond regionality.

There is clear evidence that mental speed is related to intelligence (even though, for the time being, we define intelligence as whatever it is that determines correctness of responses to intelligence items as defined by Guttman). The question is, what precisely is that relation. That question is not answered by SSA.

Jensen's "Second Hypothesis", and the Missing Conditions of the "Missing Theorem"

The Second Hypothesis and g

If Spearman's (1927) tetrad condition is satisfied in each of several populations, and if indeed the single common factor in each of these populations is the same (which might be concluded from invariance of factor loadings if that were sufficient — see the following), than the second hypothesis appears to be a mathematical consequence of g, and there would be no need to test it empirically. However, Spearman's theory was a two factor theory, that is: each test is determined by a single common factor, and a specific factor. Group differences with respect to the factor scores on either or both of the general and the specific factor of a given test will determine mean group differences on a test.

The Missing Conditions of the Missing Theorem

One missing condition in the missing theorem is the assumption that there are no specific factors, and group differences depend only on g; we come back to that. But there is more. Suppose tests are determined by two common uncorrelated factors, with equal loadings within tests, but different across tests. Suppose that the factor score variance of the one factor is zero in one population, and the factor score variance of the other factor is zero in the other population. Let the factor with non-zero variance have the same variance in each population. Then, in both populations, the tetrad condition is satisfied, the (single) common factor loadings and correlations are the same, and yet two uncorrelated factors are involved, one in the one population, the other in the other population. Moreover, FA of the total population will show a single common factor with loadings equal to those in each subpopulation. This shows that g factors found in different populations may not be the same, and group differences may refer to different factors. However, in his proof of the missing theorem, Guttman assumed that the same g is involved in each (sub)population.

Factor Loadings, Correlations and Variances

Factor loadings should be invariant across subpopulations: the factor pattern describes the relations between tests and factors. In order to arrive at meaningful conclusions in comparing groups, it must be ascertained first that that relation is the same across groups. If not, the tests are psychologically different in different groups and no comparative conclusions can be drawn. The need for factor pattern invariance follows from the simple requirement that assuming a person's factor scores known, one must be able to predict the test score(s) within the limit of measurement error. Using different formulas for different subpopulations is tantamount to explaining group differences by group membership, which is an idem-per-idem explanation. Yet, group differences may exist in distributions of factor scores, and hence correlations among tests and factors may be different in different groups.

Even if the correlation matrix in each subpopulation, p, satisfies the tetrad condition, the factor correlations need not be the same, the variances, $\mathrm{Var}(g_p)$, need not be the same, and the test communalities need not be the same (even though the residual variances may be the same). It appears that Guttman ignored the difference between factor loadings and factor correlations, which is an incredible mistake. Equations 15 and 16 are derived in terms of correlations, and are correct. But, following Equation 16, Guttman now calls the *R* and *r*'s involved: loadings, and that makes Equation 16 incorrect unless $\sigma_g = \sigma_{g1} = \sigma_{g2} = 1$. Loadings are regression coefficients. Usually, tests and factors are assumed to be standardized (i.e. scaled such that their variances are equal to one), and for that standardization, regression coefficients are equal to correlation coefficients if the predictors (here: the factors) are uncorrelated. In case of a single common factor, that orthogonality condition is satisfied by definition. But in the case of different subpopulations, the standardization of the (single) factor and of the variables can be done only once (e.g., the pooled within-group variances are set equal to one).

Guttman was obviously aware of the question of units for the factor scores, and of the frailty that factor invariance can only be tested if factor loadings are invariant. Then he stipulates that the factor correlations are the same in each group, which is wrong unless the factor variances and the residual variances are also the same in the two groups. The loadings should be the same, the units must be equal, but the variances and the correlations can be different between subpopulations.

However, assuming that the tetrad condition is satisfied in each subpopulation and that the correlations of the tests with the single common factor, *g*, are the same in each group, the theorem implies that the two correlation matrices are the same. Moreover, the theorem assumes that the tetrad condition also holds in the total population, and this makes, as will be shown, the theorem a tautological consequence of the assumption that any group difference is a function of one factor only.

Two Factor Theory and Jensen's Second Hypothesis

Spearman's (1927) was a two factor theory: a general, *g*, factor common to all tests, and *n* specific factors, one for each test. We will show that two-factor theory is incompatible with the assumption that the tetrad condition is satisfied in any (sub)population unless subpopulation differences concern one factor only.

We denote tests by subscripts *j, k*, loadings by Greek letters, factor score variables by *g* and s_j for general and specific factors, respectively, and residual (error) variables by ε. We assume that x_j, *g* and s_j are standardized but the precise way is immaterial as long as it is done only once. The model is:

(1) $$x_j = \alpha_j g + \beta_j s_j + \varepsilon_j,$$

with the usual assumption that the expected subpopulation covariances among g, s_j and ε_j are zero, and the expectation of ε_j is also zero in any (sub)population. At this point, we assume that the covariances among g and s_j in the combined population need not be zero.[4] We come back to this. The loadings ought to be assumed the same in both subpopulations, but the variances of g and s_j need not be the same across subpopulations. Each subpopulation satisfies the tetrad condition, but the entire population need not satisfy the tetrad condition. Let groups be denoted by p, q. Let d denote the expectation of a group difference, for example:

(2) $$d_j = \text{Exp}(x_{j(p)}) - \text{Exp}(x_{j(q)}).$$

It follows immediately that:

(3) $$d_j = \alpha_j d_g + \beta_j d_{sj},$$

and this can be considered as the regression of d_j on α_j, with slope d_g and residual term $\beta_j d_{sj}$. Jensen's (1985) second hypothesis is that group differences on tests $j, k, ..., n$ are determined by g, rather than by s_j, and this hypothesis is both relevant, substantive, and in no way a tautological consequence of the tetrad condition in each group. The hypothesis is properly tested by inspecting the scatter plot of d_j and α_j. The correlation will be equal to 1.0 if there are no group differences except on g. Of course, if more than a single common factor is assumed, a similar hypothesis is meaningful, replacing s_j by the combined effect of all factors other than g. Considering Jensen's results, the correlation of about .6 between d_j and α_j seems to indicate that other factors besides g have a substantial effect on black-white differences, and this is perhaps a more serious criticism of Jensen's article than anything else.

The derivation previously given shows that the second hypothesis is a trivial consequence of the tetrad condition if and only if g is the only factor involved in each subpopulation: Given that we are concerned with linear models, any linear function of the data must be a linear function of the only factor involved. The tetrad condition does not, however, exclude the presence

[4] After completing the version of this commentary which was sent to the other commentators, we found it incorrect to assume non-zero covariances in the subpopulations. The more appropriate model should assume zero covariances and unit variances in some population (e.g., the total population) and allow non-zero covariances and non-unit variances in any other (sub)population. This does not affect our main argument, but it raises the issue how to define g, which we address in our rejoinder section.

of test-specific factors, and Guttman appears to have totally ignored that element of factor theory.

The Tetrad Condition in the Combined Subpopulations

In the presence of group differences on the specific factors, assuming that these factors are uncorrelated in each of the subpopulations, it will not in general be true that the specific factors are also uncorrelated in the combined population. It follows immediately that the correlation matrix of the combined population will not in general satisfy the tetrad condition, contrary to what is assumed in Guttman's missing theorem. It can be shown that the combined population correlation matrix will satisfy the tetrad condition if and only if group differences exist on no more than one of the $n + 1$ factors (g and specific factors). In other words, if the combined population correlation matrix satisfies the tetrad condition, and if there are differences on more than one test, these can be a function of g only. So the implicit assumption of the missing theorem is that there is only g and nothing else.

Moreover, in the combined population, the specific factors may be correlated with the common factor, and so FA of the combined populations correlation matrix will show a pattern of loadings which is different from that of the subpopulations, as pointed out by Guttman.

Conversely, if the entire population satisfies the tetrad condition, and there are group differences on more than one test-specific factor, then the subpopulation correlation matrices will not in general satisfy the tetrad condition. Ellis (1992) has shown that if the covariance matrix in each subpopulation is unifactorial, then all tests must be congeneric, that is, only one factor (g) is involved.

Guttman pointed out the crucial weakness of FA in the last section of his article: "can factors be universal?" In the same way as the correlations in a heterogeneous population can be contaminated by group differences, so the correlations in a group can be contaminated by subgroup differences. This contamination can only be excluded if there is only one factor involved, that is, if all tests are congeneric. In the presence of test-specific factors or minor factors, FA is not the appropriate method to assess the principal factor unequivocally, but neither is SSA. If is is true that $x_j = \alpha_j g + \beta_j s_j + \varepsilon_j$, neither FA nor SSA will be able to test that truth unequivocally.

References

Berger, M. (1982). The scientific approach to intelligence: An overview of its history with special reference to mental speed. In H. J. Eysenck (Ed.), *A model for intelligence*. Berlin - New York: Springer.

Brand, C. R., & Dreary, I. J. (1982). Intelligence and inspection time. In H. J. Eysenck (Ed.), *A model for intelligence*. Berlin - New York: Springer.

Dubois, P. H. (1960). An analysis of Guttman's simplex. *Psychometrika, 25*, 173-181.

Ellis, J. L. (1992). Subpopulation invariance of patterns in covariance matrices. Paper submitted for publication.

Ellis, J. L., & VandenWollenberg, A. L. (1992). Local homogeneity in latent trait models; a characterization of the homogeneous monotone IRT model. Paper submitted for publication.

Eysenck, H. J. (1982). Introduction. In H. J. Eysenck (Ed.), *A model for intelligence*. Berlin - New York: Springer.

Guttman, L. (1981a). What is not what in statistics. In I. Borg (Ed.), *Multidimensional data representations, when and why* (pp. 20-46). Ann Arbor, Mathesis Press. Reprinted in R. M. Hauser, D. Mechanic, & A. Haller (Eds.) (1982), *Social structure and behavior* (pp. 331-348). New York: Academic Press.

Guttman, L. (1981b). What is not what in theory construction. In I. Borg (Ed.), *Multidimensional data representations, when and why* (pp. 47-64). Ann Arbor: Mathesis Press. Also in *The Statistician*, 1977, *26*(2), 81-107.

Jensen, A. R. (1985). The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences, 8*, 193-263.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. London: Erlbaum.

Luce, R. D. (1986). *Response times. Their roles in inferring elementary mental organization*. New York - Oxford: Oxford University Press.

McDonald, R. P., & Ahlawat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology, 27*, 82-99.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. Hillsdale, NJ: Erlbaum.

Rosenbaum, P. R. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika, 49*, 425-436.

Roskam, E. (1983). Toward a psychometric theory of intelligence. In E. E. Roskam & R. Suck (Eds.), *Progress in mathematical psychology*. Amsterdam: Elsevier.

Roskam, E. E., Van Breukelen, G. J. P., & Jansen, R. T. W. L. (1989). Concentration, speed and precision in time-limited tasks. In D. Vickers & P. L. Smith (Eds.), *Information processing: measures, mechanism, and models*, (Proceedings of the XXIV International Congress of Psychology) (pp. 291-310). Amsterdam: Elsevier.

Spearman, C. (1927). *The abilities of men*. London: MacMillan.

Thurstone, L. L. (1935). *The vectors of mind*. Chicago: Chicago University Press.

Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: Chicago University Press.

Van Breukelen, G. J. P. (1989). Concentration, speed and precision in mental tests. Unpublished doctoral dissertation, University of Nijmegen, Institute for Cognition Research and Information Technology.