estratti dalle pp. 11-49 in Edgar F. Borgatta e David J. Jackson (*eds.*), *Factor Analysis and Measurement in Sociological Research: a Multi-Dimensional Perspective*. London: Sage, 1981

Factor Analysis as an Aid in the Formation and Refinement of Empirically Useful Concepts

by Alberto Marradi
University of Bologna (Italy)

11-12: One starts with one set of variables and finds their linear combination(s) best predicting the variance in the set itself. The particular nature of the technique makes it inappropriate for the measurement of relations between phenomena; on the other hand, it makes it the ideal tool for exploiting relations between lower-level phenomena in order to summarize something they have in common, i.e. to measure a higher-level phenomenon.
13: Factor analysis allows one to use statistical relations between several lower-level variables as empirical evidence for or against the establishment of a semantic relationship of indication between these variables and an abstract concept.
15: Indicators bearing a simple, textual resemblance to the term designating the trait are often fraught with strong suspicions of invalidity, i.e. of measuring something different from what they are suppose to measure. As a consequence, we are called upon to exert our semantic imagination, or informed common sense, in order to find questions and stimuli that may serve as indicators of our trait even though there is no textual resemblance between their wording and the term designating the trait.
The relation of indication between the trait and the question is not warranted by a mechanical correspondence of terms, but by the scientist's judgment as to a more or less vast semantic overlap between trait and indicator. Thanks to this overlap, the question will obtain responses that are — for a majority of subjects — largely controlled by their positions on the trait that is being measured. However, the overlap is far from complete, because:
a) The indicator has a much narrower referent than the general trait...
b) On the other hand, the indicator also measures something other than the trait...
As a consequence, the validity of an item as an indicator of a general trait should always be questioned.
16: A plurality of indicators for the same concept is likely to disturb only those who attribute to the whole of science that semantic definiteness and rigidity which has been reached only by some section of physical science, after centuries of struggle about concepts and their proper operational definitions. In the early phase of a science (and probably in any phase of the social sciences, whose objects vary significantly over time and space) a plurality of indicators for the same concept is a physiological rather than a pathological state, and should be blessed in so far as it alerts the scientist to the autonomy of his concept vis-à-vis his measuring instruments, and to the gap dividing the two.
17: The validity of an indicator may not be proved in a 'objective' way even a posteriori, i.e. by algorithms applied to the data gathered...
High correlation coefficients tell us that our indicators are measuring much the same thing, without formally proving anything as to its nature. While validity is never positively proved, if some items shows systematically lower correlations with others, their validity as indicators of that 'same thing' is — at least within the limit of the research in question — disproved.
19-20: If we could "objectively" locate a concept in the space defined by actually measured indicators, the factor loading would be an exact measure of validity of the corresponding indicator — at least, given our sample and the techniques used in gathering and factoring the data. But neither factor analysis nor any other technique can ever establish an objective link between the contents of a researcher's mind and of a data matrix. Therefore, we must be satisfied with measuring the distance between each indicator and the centroid we ourselves implicitly locate when we choose our indicators, and measure them in the field.
20: These distances are expressed by a vector of factor loadings. Indicator having larger (positive or negative) loadings should be given prevalence in the interpretation of the factor: careful evaluation of their semantic contents will very often lead to acknowledging that we have measured something that differs to a greater or lesser extent from what we wanted to measure...
Indicators having smaller loadings may be discarded from the set. Though this invalidity is not definitively proved, surely the concept cannot be simultaneously (i.e. on the same data set) measured by them and by

indicators having higher loadings, because it would be measured through divergent, inconsistent means. The threshold beyond which a loading is to be judged small should not be defined in advance, as it depends on the nature of the data, the technique by which they have been gathered, the accuracy in error-checking, etc.
20-21: If we decide to discard one or more indicators from the set, we cannot retain the rest of factor loadings at thier previous values. When we remove one item from the list we also remove a row and a column from the correlation matrix. Though other correlations are left unchanged, the whole configuration of inter-item distances is affected to a certain degree, also depending on the number of indicators that are left untouched. Therefore, the factor loadings must be re-computed from the reduced correlation matrix.
23: Using factor loadings as weights in building factor scores is like using zero-order correlation coefficients as weights in a regression equation: each among a set of closely correlated predictors is individually credited with an influence that it shares with other members of the set, and the solution is largely redundant.
Let us translate this in terms of factor analysis. Ideally, we would like every facet of the general trait to be adequately represented in its measurement, and choose — or should choose — indicators accordingly. Covering the whole concept's semantic span usually entails picking indicators at a certain semantic distance from one another. But the public's perception of a semantic space is seldom the same as the researcher's... As a consequence, indicators too close to one another turn up artificially high loadings, while comparatively isolated indicators have their loading depressed...
Using factor loadings as weight entails magnifying consequences of the unavoidably unequal representation in the basketful of indicators we have submitted to factor analysis. **24**: On the other hand, those consequences are reduced if we use weights obtained by regressing the factor on its indicators, i.e. a perfect equivalent of a regression equation beta weights. Such weights, currently called 'factor score coefficients', only assign to each indicator a contribution to an inferred factor which is not shared by other indicators in the battery. For every indicator, the factor score coefficient is smaller than the factor loading; however, the reduction is usually lesser for comparatively isolated indicators than for tightly clustered ones...
Whenever a factor score coefficient falls below a certain threshold (for which a numerical value cannot and should not be defined in advance), the analyst must not hesitate to drop the corresponding indicator, since this simplifies both the computation of factor scores and the interpretation of the concept.
If one or more indicators are discarded, the vector of factor score coefficients should be recomputed. Even when all coefficients in the vector are satisfactorily high, but the analyst judges that the indicators are too numerous, he may experiment with various subsets of indicators in order to find out whether more economical solution exist that he deems satisfactory from a semantic point of view.
25: The initial selection of indicators entails a choice among theoretically countless alternatives. Even the most "objective" conduct (i.e. worshipping the computer printout instead of using it as a ground for wise decisions) cannot mend the unbounded subjectivity of the first step. As a consequence, we are perfectly justified in reducing the initial list when some technical manipulation has made it clear that some indicators would bring largely inconsistent or redundant contributions to the measurement of our concept. If we hesitate in reducing the list, what we are really worshipping is the most subjectively and uninformed of our choices rather than "objective" mathematical results.
27: A sharp discrimination between exploratory and confirmatory uses of factor analysis... neither portrays the present state of affairs in empirical social research, nor seems to be a desirable target for the near future. Social scientists using factor analysis for their empirical research goals... may be ranked along a continuum of awareness of the concepts they intend to measure...
At one extreme of this continuum stand those who punch all the data (the term is purposely chosen, because they seldom have a clear idea of the concept-indicator relationship, and tend to think in terms of "data") at their disposal, throw them into a computer and "let the results (or: the data) speak for themselves". This approach to factor analysis should be described as "blind" rather than "exploratory".
28: At the other end of the continuum are a few who entertain a very firm idea of what concepts are useful in desribing a certain domain, and, not being prepared to give up their idea even in the face of contrary evidence, claim confirmation by resorting to verbal acrobatics in the interpretation of factors... This may be detected by a critical reading, but it runs the risk of no sanction worse than an occasional bitter review or letter to the editor. For most readers, what goes on record is that the analyst's "hypothesis" has been "confirmed". On the other hand, the actual results are also published, thus making alternative interpretations possible.
Things are worse if confirmation-oriented analyst profits by the manifold opportunities that factor analysis offers to sophisticated users who wish to steer its course in a given direction.
29: Placing emphasis on a confirmatory function of factor analysis may lead many into committing sins for which no punishment is to be feared, because little trace of them, besides the outcome, is likely to appear in

print. Similarly, a greater methodological awareness should be reached before an exploratory function is singled out at no risk of encouraging blind factor-fishing expeditions.
It is comforting to note, however, that most users of factor analysis are neither unfailing confirmationists nor blind factor-fishers. They select their indicators with reference to some idea they have in mind about what relevant dimensions exist in a given domain. This idea is something different from a criterion of mere availability of indicators, but it is also less rigid and precise than a "hypothesis", whose formal confirmation or disconfirmation is sought. Common sense, if not hampered by misconceived falsificationism, suggests to a good many users an approach that is essentially the same as Thurstone outlined some years ago. He maintained that the theoretically important things in a factor analysis take place in the researcher's head before the analysis, when indicators are selected, and after it, when factors are interpreted.
30: The researcher's image of the concepts he is going to measure need be no more precise than is necessary to permit a meaningful choice of indicators. Any greater degree of formalization may divert the researcher's attention from the meaningful interpretation of whatever results come out, to the search for a yes-no verdict to his favoured "hypothesis".
31: Several aspects of the problem, and the way they are mutually related, should be considered before choosing the indicators to be submitted to a factor analysis:
(a) the extension of the semantic domain we want to chart by singling out and measuring concepts within it;
(b) the level of abstraction of the concepts we want to measure;
(c) the number of indicators we want to factor-analyze;
(d) the a priori strength of the semantic link we are drawing between concepts and indicators;
(e) the degree of textual similarity between the concepts, as we imagine them before the analysis, and their respective indicators;
(f) the degree of association, as measured by a factor loading, sufficient to hold an item a valid indicator of a concept...
If (a) (the domain's semantic extension) is large, and (c) (the number of indicators) is not particularly high, then (b) (the concepts' level of abstraction) will also be high and (f) (the threshold above which a factor loading is considered proof of a semantic relation between concept and indicator) will tend to be low. It is likely that these relations do not hold if (e) (concept-indicator's textual similarity) is high. This means that, for each concept that we imagine as relevant to the domain, we have chosen a set of indicators so worded (or otherwise arranged) as to maximize the probability of their clustering together. Using the same terms, or the same sentence structure, for indicators of the same set is an obvious means of bringing about this outcome. Another device is exploiting response sets by submitting indicators of the same set close together rather than randomizing them throughout the larger battery.
32: The joint factoring of items that we consider indicators of different concepts makes sense only if there is a sizable possibility that they cluster in a way different from what we have imagined. A high (e) (textual similarity within sets) makes this possibility artificially low. However, it is naturally low if (a) is large and (b) is kept low: a wide semantic domain will not be uniformly covered by low-level concepts, which probably have high (d) (close semantic link with their indicators)...
Factor analysis has a real exploratory function when indicators appear to be spread uniformly enough to leave real uncertainty about their actual clustering. This entails that concepts in turn need be semantically contiguous enough to cover the total semantic space they encompass; it also entails that (d) should not to be too high, that (e) should as low as possible, and that we cannot expect to keep (f) too high...
In an exploratory factor analysis the first key decision is to be made either on the concepts' level of abstraction (b) or on the extension of the semantic domain (a). Either of this determines the other, and strongly affect the number of indicators (c) and the strength of the concept-indicator link, both a priori (d) and as measured by a factor loading (f).
34: Kaiser's criterion (1958) states that a factor should be retained only if it gathers a greater share of the general variance than one (standardized) variable contributes. Its critics made the obvious remark that the same cutting point is being used with any number of variables in the matrix...
35: The so-called scree test (Cattell, 1966) seems an easy and very defensible solution, as it takes only the patterning of eigenvalues into account, independently from both subjective semantic appraisal and irrelevant consideration of sample size...
Whatever criterion is followed in determining how many factors to extract, it has little to do with the number of factors that should be interpreted. This decision must be controlled by non-technical considerations, related to the analyst's understanding and aims. Factors having smaller eigenvalues are often formed by one or two indicators plus error variance gathered by the others. The researcher should mention their existence, but he is

perfectly free to drop them with the corresponding indicators.
36: What is obtained by looking at the plots is an immediate insight into the way indicators cluster in the semantic space. Of course a plot represents that space from a point of view, thus blurring or concealing aspects that are clearly evidenced by other plots. Therefore, pairwise plots of all important (i.e. having sizable eigen-values) factors need to be inspected before the structure of that space is fully understood... the respective position of indicators in the space is the only empirical result of the analysis, while the location of reference axes is arbitrary. Locating the reference axes is only a necessary step to obtain representation of the indicator-points, as is true for any system of axes and points. The former may and must be altered in order to improve representation of the latter.
As a consequence, it is at the actual clustering of indicators, as portrayed in the plots, rather than at reference axes that we should look when singling out factor for interpretation, refinement, and further use. The advantage inherent in so doing is clear. Each reference axis is orthogonal to all previously defined axes, since it is extracted from a matrix of residual correlations. **37**:This constraint confers a high degree of rigidity to the whole system of axes, which encounters obvious difficulties in fitting a configuration of points freely clustering in a p-dimensional space. A technique of orthogonal rotation such as Varimax has been developed in order to improve the fit by maximizing the variance of squared loading for each factor (Kaiser, 1958), which has the effect of drawing the reference axes as near as possible to the indicator-vector. As this simplifies the task of interpreting factors represented by orthogonal reference axes in terms of actual indicators, Varimax has been largely adopted by researchers trying to reconcile the search for meaningful factors with a fidelity to the technical requirement of orthogonality...
The techniques of oblique transformation offer a fallacious alternative, in that they remove the constraint of orthogonality only to replace it by other constraints, less straightforward and less perceived but no less binding. In the various formulas there is an arbitrary parameter which actually determines the degree of obliqueness in the solution: (b) for Promax, (c) for Harris and Kaiser's method, (d) for Direct Oblimin, gamma for Carrol's Indirect Oblimin and so on. Several authors rank oblique transformation techniques in order of increasing, or decreasing, obliqueness of the produced solution; this amounts to denying the assertion that such techniques provide empirical information on the amount of factual correlation between factors.
38: <we should> abandon the idea that the final factors need in any way be connected with the reference axes that are drawn as a result of the first factor analysis of the entire battery of indicators. The drawing of reference axes may be considered as no more than a prerequisite of two-dimensional representation of vectors' configuration. When inspection of two-dimensional representations has induced a satisfactory understanding of the structure of the semantic space, the function of the reference axes may be considered accomplished, and it is the clustering of indicators as it appears in the various plots that should retain the attention.
In other words, whenever a gathering of indicator-vectors in a certain portion of space is confirmed by the relevant plots, there exists something to be interpreted...
The analyst must appeal to his judgement in order to decide what condensation of points is sufficiently tight and consistent across plots to be defined a cluster and call for interpretation. He also has to resort to his judgement in order to assign each indicator to one cluster, or possibly to none, or to more than one. In dubious cases, he may compare distances from the cluster's centroids, but he will nevertheless have to define such centroids subjectively. **39**: it usually happens that one or two clearly defined clusters show up, while a few points appear isolated and many others swarm about in the p-dimensional space, with different plots possibly supplying divergent indications as to their memberships in clusters. Each clearly defined cluster will be submitted to a separate single-factor analysis: interpretation of the factor will be refined in view of the loadings' magnitude; possibly some indicators will be dropped; when a satisfactory solution is found, factor scores will be dropped; when a satisfactory solution is found, factor scores will be computed for every individual.
Indicators that appear spatially isolated from the rest may be dropped. Small clusters may also be dropped, if good reasons are given for so doing; for example, textual similarity.
39-40: All indicators that are neither isolated nor clearly belonging to a cluster should be pooled and resubmitted to a factor analysis. The removal of already defined clusters and of isolated indicators might be described as a decompression of the semantic space, which often has the effect, in my experience, of evidencing some additional clusters that were obscured in previous representations. Each cluster defined through re-pooling will of course be submitted to a separate single-factor analysis, like the originally defined clusters...
40: A researcher may experience greater difficulties in interpreting the results of a factor analysis if he sticks to his original idea of the conceptual dimensions relevant to a certain domain... in single-factor analysis the researcher's understanding of the concept he is trying to measure should be refined and articulated, and sometimes substantially revised, in view of the factor loadings' magnitude. Yet, it is hardly conceivable that the

factor loadings of the chosen indicators could exhibit patterns that would entail a wholesale rejection of the original concept.
41: On the contrary, this may happen, and indeed it rather frequently does, in multiple-factor analysis... Two different degrees of negative consequences may be drawn from this finding:
(i) the concept should have been measured using an at list partially different set of indicators. In this case, factor analysis performs the important function of refining and improving the stock of indicators of a given concept, by eliminating those items that have proved invalid, at least within a given context;
(ii) the concept itself is scarcely useful in that particular context, as it does not favour a meaningful categorization and understanding of attitudes, abilities, acts, behaviours. This consequence should be drawn with considerable caution, and perhaps only after repeated failures with different sets of indicators and in different subgroups of a population...
The possibility of a concept in the researcher's mind finding no empirical counterpart in a cluster of indicators is matched by the possible formation of a cluster having no ready conceptual counterpart in the researcher's mind.
42: If normally we first form a concept and then look for its suitable indicators, what we may have here... are indicators in search of a concept...
we form our concept by drawing from our social and personal fund of experiences. While it would be foolish to think that the analysis of a computer printout can replace this patrimony, it definitely can become a part of it, and therefore give its peculiar, and occasionally decisive (though never exclusive), contribution to concept formation.