DOCUMENT RESUME

ED 108 266 CS 501 076

AUTHOR Hensley, Wayne E.
TITLE Tentative Suggestions on the Use of Factor Analysis in Speech Communication.
PUB DATE Apr 75
NOTE 16p.; Paper presented at the Annual Meeting of the International Communication Association (Chicago, April, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS *Communication (Thought Transfer); *Factor Analysis; *Factor Structure; Higher Education; Item Analysis; *Research Methodology; Research Tools; Speech; *Statistical Analysis

ABSTRACT

The basic considerations that should be examined in performing an exploratory descriptive factor analysis of a communication concept are described in this paper. Factor analysis is a statistical technique designed to identify the fundamental, common elements within a pool of variables. For example, imagine asking students twenty questions all designed to measure communication anxiety. Factor analysis tells which of the twenty statements seem to be measuring the same thing. Following an explanation of the concept of factor analysis, the bulk of this paper describes the application of factor analysis in speech and communication research. The basic considerations involved in this statistical technique are defining an area of concern or a concept to investigate, constructing a number of items to measure the concept, collecting responses to the items, calculating a correlation matrix of the responses, and factor analyzing the correlation matrix. Examples of a factor analysis are provided. (RB)

Tentative Suggestions on the Use of Factor Analysis in Speech Communication

Wayne E. Hensley

Indiana University

Invited paper, presented at the International Communication Convention, Chicago, April, 1975.

Tentative Suggestions on the Use
of
Factor Analysis in Speech Communication

The very fact that this program is being presented at the ICA convention ought to be enough justification for our discussion. But, like the detective in a mystery novel, I hate loose ends. As a consequence, there are several questions I should like to address, at least briefly.

The first question to be raised is: why bother with factor analysis at all? What unique advantages does factor analysis possess which warrent its consideration by those in speech communication? To answer the question one must understand that factor analysis is a statistical technique designed to identify the underlying common elements from a pool of n variables. Suppose we ask students 20 questions all designed to measure their communication anxiety. Factor analysis will tell us which of those 20 statements seem to be measuring the same thing. In the event that every single statement of the original 20 measures a mathematically distinct concept, there will be 20 factors--one for each statement. In the event that all 20 statements measure the same concept, there will be one factor. As a case in point, Cronin and Price (1974) have found that the dimensions upon which judgments of speech teachers are based appear to change as students progress through college. Freshmen seem to have one set of criteria, seniors a slightly different set.

The point is not that no other method could have possibly discovered these differences. The point is, depending on the alternative technique being considered, it might have taken years of research and dozens of researchers to have made the discoveries. Factor analysis provides a statistical means of unveiling the common elements describing a communication concept. In addition, such an investigation may be made very rapidly and with great precision. Viewed in this light, the tool has enormous potential and utility.

The second question is more penetrating. Why should a group of people, trained in speech communication, be here telling anyone else about factor analysis? Why not have a group of mathematical statisticians on the panel? I can best answer that question by relating the experience of a friend of mine. Several years ago my friend wanted to learn more about a highly specific application of a statistic to a communication problem. He took this problem to a mathematics professor to seek the answer. What ensued was a classic case of misunderstanding. Neither individual could communicate with the other. The mathematics professor did not know enough about communication and my friend did not know enough about theoretical mathematics. Each one finally agreed they did not understand the other--that was a far as they got. That is the basic reason we are here today and not a panel of mathematical statisticians.

What I will attempt to describe in this paper are some of the basic considerations one ought to consider in performing an exploratory descriptive factor analysis of a communication concept. It should be understood that this is by no means the only purpose for which factor analysis might be used but it is one of the most common. In chronological order, the factor analytic researcher usually follows a fairly typical pattern. The pattern may be described as:

1. an area of concern/concept is selected for investigation.
2. a number of items/statements are constructed to measure the concept.
3. responses to the items/statements are gathered.
4. a correlation matrix of the responses is calculated.
5. the obtained correlation matrix is factor analyzed making the following decisions:
    a. what kind of factor analysis will be used?
    b. how will the communalities be estimated?
    c. what rotational schema will be used?
    d. how many factors will be extracted?
6. the results are examined and factors named.

Initial item selection

As with all research, there is no substitute for a thoughtful and cautious selection of items to be judged by respondents. In the jargon of the computer programmer, GIGO (garbage in, garbage out) ought to be the byword for the factor analyst. The dangers of item selection or construction arise in several areas and should be discussed separately.

Theoretical item pool

From the mathematical standpoint, factor analysis is simply a tool for the identification of whether communalities exist in the data. Thus, the researcher must include items designed to tap each of the suspected dimensions of the phenomenon and no item(s) which measure anything outside the phenomenon. The construction of an item pool for self-esteem, for example, should contain only those items thought to measure something having to do with self-esteem. If we were to include--erroneously--items measuring hair color, those items would probably emerge as a factor. The vexing part is that considerable time would be lost unless it were realized that hair color had nothing to do with self-esteem.

By the same token, if the theoretical analysis had totally missed an area which was really a part of the concept, then that dimension could not possibly emerge in the factor analysis. We might be led to belive that just because our analysis accounted for 95% of the variance, we had totally described the concept. This could be a serious error. If a significant part of the concept was never measured, it cannot be described by the factor solution.

Item construction

Very often an item will emerge with very high loadings on more than one factor. Unfortunately, such items are extremely difficult to interpret even though they do seem to be measuring something. Normally, such items are discarded as "errors." The researcher is in a position to minimize such errors by correctly designing items in the first place. Edwards (1957, pp. 13-14) has given rather specific suggestions in regard to item construction and these suggestions should be heeded by more researchers. The more an item is open to multiple interpretations, the more likely it will be to load on a variety of dimensions. For example, consider the statement: "If you really want to know what's wrong with this country, just look at our moral standards and business practices today." If a respondent disagreed with the statement, is the disagreement based on the words "moral standards" or on "business practices" or both? Each item submitted for respondent judgment ought to clearly tap one and only one facet of the construct. Failure to carefully construct items may yield a confused factor structure, at best, or a totally uninterpretable factor structure, at worst.

Sample size

Although I have stated my opinions on sample size elsewhere (Hensley, 1974), perhaps they bear repeating and expansion. The techniques of factor analysis may begin with a variety of data inputs but the two most common are either a correlation matrix or individual responses from which a correlation matrix is calculated. It should be emphasized that the correlation matrix will produce the same factor structure regardless of how many observations were used in generating the correlation matrix. However, it would be grossly erroneous to assume that the size of the original data pool was unrelated to the emergent solution.

One way to analyze the issue of sample size is to examine the correlation coefficient per se for stability. It is well known that the standard error of a correlation is inversely related to the sample size. This strongly suggests we ought to use rather large samples. But, like most advice, there is another side to the coin. Somewhere in most statistics books the formula for testing the significance of a correlation coefficient is given as $Z = r\sqrt{n-1}$. In words, the degree of confidence we may place in the coefficient is a function of both the magnitude of the correlation and of the sample size. Normally, because of measurement error and inherent subject variablity, we rightly concentrate on the issue of sample size. However, we might just as easily look at the effect of the correlation magnitude. As a case in point, suppose we had a small sample (N = 20) measuring three variables. I have constructed a sample with correlations of: .98, .92 and .98. The factor which describes these three correlations accounts for 97.5% of the variance with loadings of: .999, .978 and .979. Even with a sample so small no respectable researcher would recommend its use, we have obtained a solution which seems highly revealing. (Of course, I made up all the data; they measure nothing.) The point, which the data illustrate, is that the correlation matrix is a mirror of the underlying factor structure. If that correlation matrix measures an enormously powerful factor structure, then the sample size necessary to reveal the structure is drastically reduced. The exception having been observed, we may now point out that the real world is never as cooperative as a set of problem data. However, it is well to keep in mind that one need not throw away data simply because the number of observations falls below the recommended level.

Approaching the problem from another direction, Guertin and Bailey (1970) have empirically investigated the effect of several different sample sizes on the emergent factor structure. Beginning with a sample of 200 persons who took the Wechsler Adult Intelligence Scale, two random samples of 100 and two of 25 were drawn from the original data pool. The factor structures of the two samples of 100 agreed with the parent sample fairly well; the factor structures of the two samples of 25 were disasters. One of the samples of 25, after yielding one identifiable factor, produced a solution "not clearly recognizable as being from the same study" (Guertin & Bailey, 1970, p. 170). It was concluded that ". . . basing even product-moment r's on an n of 100 is a questionable procedure and is suitable for only a very tentative pilot study" (Guertin & Bailey, 1970, p. 170). Generally, the rule of thumb offered on this topic is that the sample size should number at least 200. This is sound advice for most communication research.

<u>What kind of factor analysis will be used?</u>

Contrary to popular belief, the words "factor analysis" do not refer to a single statistical technique. Rather, there are a host of techniques to which we might be making reference. Among the earliest of these techniques was a method called centroid factor analysis. Basically, centroid was only an approximation solution and has fallen into disuse. The most common of today's techniques is the principle components analysis available at most computer centers. I would estimate that as many as 75% of the reported studies probably use this approach. The fundamental idea of principle components is to seek the solution which maximizes the sum of squares for the first factor. The same procedure is then followed for the second factor using the residual matrix and so on. Thus, we extract the "principle components", one at a time, from the correlation matrix. Principle components can tell us what the factor structure looks like in any given sample but, without repeating the sampling on different populations, we are unable to make inferences about the structure of the concept per se.

Alpha factor analysis (Kaiser & Caffery, 1965), on the other hand, is specifically designed to reveal the dimensionality of the concept under investigation. One of the major strengths of alpha factor analysis is that it yields a reliable solution in the Kuder-Richardson sense of reliable. Alpha is uniquely useful for generalizing to a universe of all logical items measuring the concept under investigation. As such, it is a logical tool in scale construction. However, for all of its promise, alpha may not yield identical results with different samples (Hensley, Hensley & Munro, In press). The answer seems to be that the analysis is always data based. While alpha does generalize to all logical items, it can only generalize from the data base provided. (GIGO!)

Having mentioned these different approaches, and there are many more, perhaps some examples would be useful. If we were investigating communication anxiety among students in fundamentals of speech, I would recommend principle components. The interest here is in fundamentals students and not in some other group. If we were interested in the factor structure of alienation--as a psychological concept--I would recommend alpha. Here the interest is in the concept and not in some group's reaction to the concept.

Before leaving the topic I should note that as the sample size increases, the choice becomes moot. At infinity, all forms of factor analysis theoretically converge. For practical purposes, if you have an enormous sample of persons, it probably doesn't matter which type of analysis you choose.

How will the communalities be estimated?

In most factor analytic studies there is a table presenting a column labeled h^2. This is the communality of the variable in that row of the table. The communality may be calculated directly by squaring all the factor loadings for the variable and summing them; it is that portion of the total variance accounted for by the factors. The communality is comparable to r^2 used with correlation coefficients; the difference is that the communality refers to a factor and a variable and not to an individual correlation. The only reason for talking about communalities at all is because the researcher must decide, for mathematical reasons, what sort of estimates the communalities will be based on in the analysis. The communality estimates are used in the diagonal of the correlation matrix and may either increase or decrease the precision of the solution obtained.

Using a communality estimate of 1.00 would mean that the variable was thought to be a perfect mirror of the phenomenon in question. This estimate is the upper bound of the communality and is logically too high. A communality estimate using a test-retest reliablity coefficient admits that the variable has some measurement error built into it. The reliability coefficient eliminates the error variance in the factoring due to sampling vagaries. Reliability coefficients as estimators are more conservative than unities factoring only the common and specific variance. A third type of estimate is to calculate the multiple R^2 of the variable with all other variables in the problem. The choice of multiple R^2 is generally made because a variable cannot have more in common with the factors than it has in common with the sum of all the other variables; which, of course, make up the factors. Multiple R^2 is the lower bound of the communality estimate and the most conservative choice. Logically, the multiple would seem to make the most sense. Further, empirical studies (Humphreys & Ilgen, 1969) have concluded that R^2 yields the most stable factor structure. I recommend its use.

What rotational schema will be used?

It is worth keeping in mind that the rotational schema, either orthogonal or oblique, has little to do with the type of factor analysis chosen. Orthogonal rotations assume that each factor is mathematically independent of every other factor. Oblique rotations assume there is some correlation among the factors. Since the publication of an early article by Kaiser (1958), varimax has become the single most commonly used rotational approach for orthogonal factor analysis. It is now so widely accepted, that is is used as the standard to examine the variability of other rotations (Bailey & Guertin, 1970). For research making the orthogonal assumption, varimax is a reasonable and widely accepted choice.

The choice of rotations for an oblique solution does not, by comparison, offer any reduction in researcher dissonance. Rummel (1970, pp. 411-420) reviews no less than eight different types of oblique rotations discussing the merits and demerits of each. Rummel's advice (pg. 411) is to try several rotations to see if they agree. If they do, quit. His comment, as well as my own experience, is that the different rotations will almost always produce highly similar results. If the rotations do not produce similar results, the researcher is forced to examine the assumptions made by each rotational technique for a theoretical choice among them. An investigation in this vein, is that of Gorsuch (1970) in which four different types of rotations were found to produce highly similar results. This empirical finding should tend to defuse any anxiety aroused by the lack of specific recommendations in this area.

Finally, if the factors are truely independent, the oblique and orthogonal solutions should be identical. This feature may be used to examine assumptions of orthogonality. Assuming that the differences between the two solutions are minor ones, it is possible to demonstrate convincingly that the orthogonal solution is the more preferable (see, for example: Hensley & Roberts, 1975).

How many factors will be extracted?

The number of factors to extract is a decision which must be made either by the researcher or by some arbitrary rule. The most popular of the arbitrary rules is the latent root criteria. This rule specifies that whenever the eigenvalue of a factor is less than 1.00, factoring will cease. The criteria is both straightforward and persuasive since, by this rule, it is demanded that any factor must account for at least as much variance as any variable. When the factor does not account for as much variance as a variable, factoring stops. The problem here becomes visible only when the total research picture is taken into account. If we had only 10 variables in the initial problem, then 10% of the variance must be accounted for to meet the latent root criteria. If we had 100 variables, the inclusion of eigenvalues of 1.00, may yield some very trivial factors accounting for only 1% of the variance. In fairness, it must be noted that most of the criticism of the latent root criteria has been because it stops the factoring process much too soon with small samples (Cattell, 1966, p. 207) but the astute researcher should also be aware of the pitfalls possible with large numbers of variables.

Another rule often suggested is the percent of variance accounted for by a factor. For example, the criteria may be that any factor must account for at least 5% of the total variance. Again, there is a compelling logic about this rule. After all, if the factors cannot tell us something about the covariance of the matrix, there is no point in spending time trying to analyze them. But like the latent root criteria, blind application of this rule may either exclude important factors or include banal factors depending on the number of variables.

Before offering any specific counsel, we ought to consider the effects of both overfactoring and underfactoring. In a recent study (Hensley & Roberts, 1975) the Rosenburg scale of self-esteem was shown to be unidimensional. Each of the five positively worded items loaded on one factor and each of the five negatively worded items loaded on the other. The researchers concluded that the two factors were measuring a basic response set and that the Rosenburg self-esteem scale comprised only a single dimension. This factor structure is shown in Table 1.

Insert Table 1 about here.

Next, observe the effect of demanding a one-factor solution on the pattern of factor loadings. This solution is probably not "wrong" or "bad" in the usual sense of those words but the response set so clearly visible in the two-factor solution is now totally obscured.

Insert Table 2 about here.

Finally, if we demand a three-factor solution, the pattern is considerably altered. Factor I emerges intact but Factor II now has only items 1 and

5 remaining. Factor III is now composed of items 3 and 8. Item 4 loads on both factors II and III and might very well be discarded as non-meaningful. Whatever logical sense remains in this solution, it is strained and would require a great deal of explanation on the part of the investigator.

Insert Table 3 about here.

The point to be made is not that overfactoring is worse than underfactoring but that the effect of either is to cloud the phenomenon and make interpretation difficult. The standard I would recommend for the number of factors to extract is not a statistical one at all. It is the principle of interpretability. I know of no substitute for an intelligent and careful appraisal of the data by an informed researcher. Surely, statistical standards can suggest the approximate area in which the best solution must reside. But, ultimately, there is no substitute for theoretical expertise in the substantive area being investigated. Remembering that a factor solution is, at least metaphorically, like focusing a camera or microscope, we must ask at what point the revealed picture of reality takes on the greatest clarity. Deliberately calling for the surrounding solutions--both over and underfactoring--will most clearly reveal the best, i. e. the most theoretically interpretable, solution for the researcher.

Which items should be retained?

One of the most troublesome problems in factor interpretation is which items ought to be used to describe the factors. Clearly, not all the items should be used but which one should be rejected? Basically, the researcher must juggle several considerations at once to reach a considered judgment. Not the least of these considerations is the absolute magnitude of the factor loading. There are a number of questions to be asked about the size of factor loadings before reaching a decision to retain or reject an item as loading on a factor. First, the factor loading is the correlation between the item and the factor. Understanding this fact, can one justify the retention of an item loading .30 (a relatively common standard in the past) when the item explains less than 10% of the factor variance? Second, is it possible to justify the retention of an item simply because it achieves the necessary magnitude (Parker, 1970) to reach statistical significance? For example, with a very large sample and reasonably robust factor structure, I have seen factor loadings of .17 reach the .05 level of confidence. Third, is the aim of the study to get some "ball park" idea of what the factor looks like or to map the construct more precisely?

Each of the questions and considerations involves matters of personal judgment and, as such, are always subject to criticism. For my own part, I would recommend never accepting a factor loading below .30. Depending on the degree of precision desired, it might be desirable to even demand loadings of .50 or .60 for the investigation.

The other side of the coin involves the degree of factorial purity of the items. Other investigations have considered the issue of factorial purity in detail (Pennell, 1968) and I will not attempt to replicate that effort here. What is meant is the degree to which an item is allowed to load on one and only one factor in order to be considered representative of that factor. Common standards are loadings of at least .50 with no other loadings as high as .30 and so forth. The real question is how one discovers the "gap" between the primary and secondary loadings (in the previous sentence, for example, the gap was .20). I have never seen a careful consideration of the gap issue but it generally turns out to be .20 or .30 in most of the published speech communication research. Here, I am guilty of using the criteria being imposed by other investigations simply because they are used by other investigations. I would recommend a gap of at least .20 but perhaps some interested reader will investigate the issue and discover a more precise standard for gaps taking into account sample size, loading magnitude, rotational criteria and so forth.

Other issues

There are several questions I should like to address which do not fall logically into the previously constructed outline. Of necessity, they must be placed in the "other" category. I do not mean to imply, and it should not be presumed, that these are topics of little importance. Nor are they afterthoughts constructed at the last minute. Rather, these are issues which will be encountered only after considerable use of factor analysis.

Factor scores

Perhaps the best way to understand a factor score is to observe that a factor loading, at least for orthogonal solutions, is the correlation between that item and the factor. Consequently, it is clear that no item ever perfectly measures the factor in question. A logical means of proceeding would be to take the weight provided by the factor loading and multiply it times the raw score obtained for each individual. This should not be done. While such a computation makes intuitive sense, the actual composition of factor scores is much more complex. As a rule of thumb, the most simple method to obtain factor scores is to have the computer program provide them as output onto tape or cards.

As an example of factor score use, suppose the researcher wished to know if males differed from females in terms of communication anxiety. It would be erroneous to simply add the raw scores because each of the items differentially measures the concept. Thus, we should obtain the factor score output and run the t-test on the factor scores and not on the raw data.

Interstudy factor comparisons

At one point in time it was all but impossible for factor structures from one study to be compared with other factor structures. This restriction has been overcome by sophisticated computer programs available at most computer centers. If a computer center does not yet have such a program, a good one is given in Veldman (1967) called RELATE. The basic idea is to take one solution as a criterion and rotate the other solution to the point of greatest correspondence. A matrix of cosines is calculated which, roughly, may be interpreted as correlations. Another approach, is to compare the same persons with cannonicals although the interpretation problem may be a bit more complex.

Second-order factors

Second-order factor analysis refers to those higher level concepts which account for the factors found in the initial factor analysis. After calling for an oblique solution, the answer is subjected to another factor analysis subject to all the constraints and demands discussed in this and the other papers. The purpose of the second-order analysis is to ascertain if there is a "super" factor or factors which underly the whole theoretical concept.

Technically, the factoring process may go on indefinitely but as the second, third and higher level factors emerge, they become increasingly difficult to interprete. Personally, I cannot remember having ever seen an example of any third-order factors and even second order factors may defy explanation (Hensley, Hensley & Munro, In press).

Overview

What I have attempted to outline is a list of basic considerations to which one ought to give thought when embarking on a factor analytic investigation. It is not an overstatement to say that factor analysis is a powerful tool for the communication researcher. But, like any power tool, it should be use properly and with due caution.

References

Bailey, J. P. & Guertin, W. H. Test item dependence of several oblique factor solutions. Educational & Psychological Measurement, 1970, 30, 611-619.

Cattell, R. B. (Ed.) Handbook of multivariate experimental psychology. Chicago: Rand McNally, 1966.

Cronen, V. E. & Price, W. K. Class year, dimensions of student judgment, and the use of course evaluation instruments. Speech Teacher, 1974, 23, 34-39.

Edwards, A. L. Techniques of attitude scale construction. New York: Appleton-Century-Crofts, 1957.

Gorsuch, R. L. A comparison of biquartimin, maxplane, promax and varimax. Educational & Psychological Measurement, 1970, 30, 861-872.

Guertin, W. H. & Bailey, J. P. Introduction to modern factor analysis. Ann Arbor: Edwards Brothers, 1970.

Hensley, D. R., Hensley, W. E., & Munro, H. P. The factor structure of Dean's alienation scale among college students. Psychological Reports (In press).

Hensley, W. E. A criticism of 'Dimensions of source credibility: A test for reproducibility.' Speech Monographs, 1974, 41, 293-294.

Hensley, W. E. & Roberts, M. K. Dimensions of Rosenburg's self-esteem scale. (Submitted for publication, 1975).

Humphreys, L. G. & Ilgen, D. R. Note on a criterion for the number of common factors. Educational & Psychological Measurement, 1969, 29, 571-578.

Kaiser, H. F. & Caffery, J. Alpha factor analysis. Psychometrika, 1965, 30, 1-14.

Parker, R. M. A subroutine to estimate the critical value of factor loadings. Educational & Psychological Measurement, 1970, 30, 713.

Pennell, R. The influence of communality and N on the sampling distributions of factor loadings. Psychometrika, 1968, 33, 423-439.

Rummel, R. J. Applied factor analysis. Evanston: Northwestern University Press, 1970.

Veldman, D. J. Fortran programming for the behavioral sciences. New York: Holt, Rinehart & Winston, 1967.

Table 1

Factor Structure of
Rosenburg Self-Esteem Scale
Principle Components Analysis
Varimax Rotation,
N = 479

		Factor I	Factor II
1.	At times I think I am no good at all.	-.11	.71
2.	I take a positive attitude toward myself.	.53	-.19
3.	All in all, I am inclined to feel that I am a failure.	-.12	.67
4.	I wish I could have more respect for myself.	-.18	.63
5.	I certainly feel useless at times.	-.12	.57
6.	I feel that I am a person of worth, at least on an equal plan with others.	.61	-.17
7.	On the whole, I am satisfied with myself.	.59	-.17
8.	I feel I do not have much to be proud of.	-.18	.52
9.	I feel that I have a number of good qualities.	.67	-.09
10.	I am able to do things as well as most other people.	.71	-.10

Table 2

One-Factor Solution

Item	Factor Loading
1	-.53
2	.50
3	-.53
4	-.53
5	-.47
6	.55
7	.54
8	-.47
9	.52
10	.52

Table 3

Three Factor Solution

Item	Factor I	Factor II	Factor III
1	-.11	-.57	-.34
2	.52	.13	.13
3	-.11	-.33	-.58
4	-.16	-.41	-.43
5	-.12	-.57	-.21
6	.60	.17	.06
7	.59	.09	.16
8	-.16	-.19	-.53
9	.63	.05	.10
10	.65	.05	.07