

Research Notes

Analyzing Library Survey Data Using Factor Analysis

Linda L. Phillips and William Lyons

Factor analysis was used to investigate the dimensions of faculty attitudes about library policies. Initially developed in the context of psychology, factor analysis attempts to place variables in groups. This statistical technique is well suited for managing large data sets such as those collected in a survey. Factor analysis of nineteen policy questions from a library survey at The University of Tennessee, Knoxville, revealed that faculty think of library policies in four dimensions—computerized access, circulation, specialized collection/services, and general collection. A characterization of each category is followed by suggested applications.

The purpose of this paper is to illustrate the use of factor analysis to analyze and interpret large data sets. Factor analysis is particularly well suited for the examination of library survey data because surveys typically generate a great amount of raw data that must be interpreted before meaningful conclusions can be reached. Unfortunately, reports of survey research are frequently simplistic and descriptive because the raw data elude interpretation. Appropriately managed survey data, however, reveal relationships among variables, explain differences among responses, and permit understanding of apparent conflict.

Factor analytic procedures attempt to place variables in groups. The groups are

named, permitting variables in the group to be considered together. In effect, grouping of the data makes interpretation possible. For example, in a survey that contained thirty variables, factor analysis might place the variables into three sets. Subsequently, the researcher would draw relationships and reach conclusions by considering the three sets, rather than by comparing each of the thirty variables with all of the others in the survey.

A review article by F. William Summers concludes that surveys are "clearly the oldest and most enduring method of research on libraries."¹ However, little published evidence indicates that factor analysis has been used to interpret library survey data. Michael Halperin and Maureen Strazdon described the application of a similar technique, conjoint analysis, to measure students' preferences for reference service.² Like factor analysis, conjoint analysis places variables in categories. However, the procedures are based on designation of categories prior to administration of the survey. One of the advantages of factor analysis is that the categories or groups emerge through the application of the statistical program. Relationships that are identified among variables may differ remarkably from an original hypothesis.

Linda L. Phillips is Associate Professor and Head, Cooperative Information Services, John C. Hodges Library, and William Lyons is Professor of Political Science and Director of Survey Research in the Social Science Research Institute at The University of Tennessee, Knoxville.

William McGrath et al. applied factor analysis to a formula for determining library collection allocations to academic departments.³ Twenty-two variables represented questions that should be addressed when developing an allocation formula. Data on each variable were analyzed first by multiple correlation and then by factor analysis. The data were reduced to three groups or factors. Although names had been assigned to the three factors prior to analysis, the results showed that the variables that came together in each of the factor categories did not fit the preselected names chosen by the researchers. Factor analysis drew attention to relationships that had not been considered previously. The technique permitted creation of the allocation formula using three factors, where previously there would have been twenty-two.

BACKGROUND

The University of Tennessee, Knoxville, has received national attention for its leadership in the area of assessment of educational outcomes. One UTK evaluative program, the Academic Program Review, is sponsored by the Office of the Provost and has three major components. First, the department being reviewed prepares a descriptive self-study. Second, a review team composed of internal and external reviewers visits the department. Third, a series of reports are written by the review team, the department head, and the university administration. The reports are used to evaluate the department, to establish departmental goals, and to provide a foundation for subsequent strategic planning efforts.

In spring 1987, the university library participated in its first program review. As a supplement to the usual review components, a survey of faculty opinion about the library was conducted. Information about faculty library use, evaluation of the collections and physical facilities, attitudes toward services and policies, and assessment of library priorities were intended to contribute to the program review and to be used as a basis for planning.

The survey was designed and adminis-

tered by one of the authors, a social scientist with extensive survey research experience. A product of discussions among the faculty senate library committee, the library administration, and campus administration, the instrument was drafted in response to a list of topics solicited from the faculty senate library committee. Following the initial draft, the library administration added topics and gave advice on the organization of the questionnaire. The questionnaire addressed a number of points. In addition to overall evaluation of the library, specific judgments were sought regarding the quality of the collection, facilities, and various services provided by the library staff. The instrument also solicited preferences for a range of actual or desired library policies. Finally, a variety of background measures was obtained for each respondent to facilitate comparison of the results across colleges and divisions and by faculty rank.

Every faculty member and graduate teaching assistant at UTK was sent a comprehensive survey instrument in April 1987. Sponsored by the UTK Office of Institutional Research, the survey was sent in three waves, yielding a response rate of over 75 percent. Of 1,690 questionnaires distributed, 932 were returned by faculty (76 percent) and 343 were from graduate teaching assistants (74 percent). The survey produced considerable data—over 1,200 responses to approximately 140 questions.⁴ The authors chose factor analysis as a relevant technique to identify relationships among nineteen of the questions that asked faculty to select among priorities regarding various library services and collections. The following discussion demonstrates how factor analysis may be applied in the interpretation of responses to policy preference questions. Because there were no significant differences between faculty and graduate teaching assistant responses on this portion, responses of both groups are referred to by the generic term *faculty*.

METHOD

Respondents were asked nineteen questions concerning library policies and priorities. (Policy and priority questions from

the survey are shown in appendix A). The policy questions followed a Likert format, in which faculty were asked whether they agreed or disagreed with various statements. For eleven possible "Improvements" to library collections or services, each person was asked to assign priorities, ranging from "not a priority" to "one of the highest priorities." The investigators had particular interest in determining faculty preferences in reduced terms, that is, to find the underlying dimensions of policy preference.

Factor analysis works well for such a data reduction task. Basically, factor analytic procedures try to "fit" a matrix of correlations of variables into the smallest number of dimensions. The analysis always starts with the same number of dimensions as variables. The number of dimensions is gradually reduced until the amount of variance left unexplained amount the correlation matrix exceeds acceptability. At that point new composite variables may be created (one for each of the remaining dimensions), and existing variables are correlated with these new composite variables. These new composite variables, or factors, derive their identity from the context of the variables with which they are strongly correlated. These correlations are often called factor loadings.

To reveal the multiple dimensions produced by factor analysis, factor loadings are rotated according to various criteria, the most common of which is a Varimax Rotation. The rotation procedure groups variables that tend to correlate highly with some factors and considers them together. This allows factors to be defined in terms of the variables with which they are highly correlated. In most rotation procedures, the rotated factors are treated as though they were statistically independent. Simply put, this independence allows the analyst to deal with factors that do not statistically overlap with other factors.

As factor analysis was initially developed in the context of psychology, the following hypothetical situation shows an example of its application. Suppose one is analyzing a number of variables dealing with creativity and finds a distinct, two-

factor solution. What conclusion could be drawn? The first factor is characterized by high correlation among variables measuring artistic ability, and the second factor is a composite of measures tapping abilities in abstract reasoning. We could infer that creativity is two dimensional and that any discussion implying only one dimension would overlook the true nature of the larger concept.

FINDINGS

Respondents' answers to nineteen questions were subjected to factor analysis. The criterion used by the package employed for the analysis, SPSSPC (Kaiser's Criterion) selected a six-factor solution (eigenvalue = 1). However, we decided to relinquish a small amount of explanatory power for the advantage of dealing with fewer dimensions and further reduced the number of factors to four.

"There is always some aspect of the variance among all respondents that is specific to a variable and some remaining variance that represents the underlying dimension of thinking."

The four factors explained 46 percent of the variance among the policy variables. This means that slightly less than half of the variance among the variables is "common variance," subject to explanation by the four-factor solution. Conversely, slightly more than half of the variance among the variables is "unique variance" not accounted for by the solution. This division of variance into common and unique is important. The logic of the division is as follows. Even though the responses to a question about, say, the desirability of increasing library hours taps what users may think specifically about hours (unique), it also may well reflect some broader, underlying dimension of respondent thinking, such as overall feeling about a more accessible library (common). There is always some aspect of the variance among all respondents that is specific to a variable and some remaining

variance that represents the underlying dimension of thinking. Thus, by reducing the numbers of dimensions from nineteen to four, we simplify our consideration of that set of questions. This simplification is at the expense of working with a solution that captures only about half of the variation with which we started. When we deal with the common proportion of the variance, we lose the unique variance that is captured by the individual variables. However, the underlying assumption inherent in the decision to work with the composite factor rather than the unique variable is that the underlying dimensionality of thinking is of paramount interest rather than the specifics of the individual question.

The four-factor rotated solution is shown in table 1. The name provided to each factor reflects the underlying composite that summarizes those variables

with high loadings. Variables associated with each factor are grouped together in table 1 to facilitate interpretation and discussion.

DISCUSSION

The variables with high loadings on the first factor are:

1. Agreement with the desirability of the library's maintaining access to national databases

2. Higher priority status to subsidized computer literature searches

3. Higher priority status to capacity to send and receive library messages on the personal computer

4. Higher priority status to computerized reference service

These four variables appear to tap a common dimension: computerized access to information beyond the confines of the local collection. The fact that they load on

TABLE 1
LIBRARY POLICIES AND PRIORITIES: FOUR-FACTOR SOLUTION

Factors	Computerized Access	Circulation	Specialized Collection	General Collection
Computerized Access				
Increase computerized reference services	.82636	.01943	.05073	.03044
Subsidize computerized literature searches	.73488	.14812	.06689	-.01884
Increase PC message capabilities	.71655	.02657	.03337	.04614
More national database access *	.55707	.18675	.06249	.07876
Circulation				
Open 24 hours*	.06470	.78066	.04433	-.01405
Increase hours	-.01109	.76204	.18924	.00957
Open every day*	.04974	.69938	-.05723	.11997
Check out nonprint*	.08587	.48072	-.05263	-.08605
Check out periodicals*	.03203	.46269	-.07152	.08381
Specialized Collection				
Increase A/V acquisitions	-.00100	-.00002	.68036	-.07796
Increase staff	.04098	.02193	.66165	-.00594
Increase rare books	.02514	.00645	.53434	.28309
Increase reference holdings	.19651	-.07335	.48778	.33361
Increase user education	.40880	-.01958	.47638	-.20624
General Collection				
Increase book acquisitions	-.06574	.01055	.22104	.74490
Increase journal acquisitions	.13433	.01551	.07171	.73245
Spend on collection*	.02404	.09266	-.33818	.52575
% Explained variance	16.1%	11%	9%	7.7%
Cumulative % of variance explained = 43.8%				

*Policy question: High value indicates agreement. Other questions refer to priorities, and high values indicate assignment of high priority.

one factor suggests that there is commonality among these variables and that there is statistical justification to the discussion of this commonality. In such a discussion, it is helpful to use a name for the dimension that would reflect what is shared among these variables, Computerized Access. Finally, the fact that the variables can be grouped statistically suggests that survey respondents tend to think of them together and apart from other variables. The factor analysis, however, does not reveal anything about the degree to which respondents hold one variable or the other in higher or lower priority. It only indicates that these items vary together, i.e., that they may be placed together in the Computerized Access category.

The second factor was named Circulation because items relating to that factor—longer hours, more days open, and more liberal check-out procedures for periodicals and audiovisual materials—involved access to and use of library collections. It is interesting that these items come together on a single factor, for it would not be readily apparent that respondents think of the hours and days of library operation along the same dimension as increased check-out policies.

The third and fourth factors tap dimensions primarily dealing with the collection. The third was labeled Specialized Collection because it is defined by such types of materials in the collection as reference sources, rare books, and audiovisual programs. Two items concerning staff and staff functions also appear on this factor—increased education of the user community and increased staff. How would one interpret this seeming anomaly? One explanation is that all of the variables in group three serve a specialized clientele. As faculty tend not to seek assistance in their use of information resources, the availability of user education and staff assistance appears to be considered a specialized feature of the library.

The final dimension underlying the collective thinking of the faculty on library policies had to do with the general collection of books and journals. Obviously, respondents think of these two items together and apart from the other collection variables. Of further note is that when the

respondents think of spending on the collection in general, they may be thinking of the aggregate size of the collection rather than its composition. In fact, the negative loading ($-.338$) for this variable on the specialized collection factor suggests a weak negative relationship between the priority for spending on the entire collection and the priority assigned for development of the specialized collection.

The above discussion suggests five possible applications. First, when library staff and/or administrators communicate with faculty on library policies, they can consider the policy dimensions used by the faculty. Some policy matters might be addressed collectively, in packages consistent with those identified in this analysis. Second, the fourfold policy categorization may be useful in assessing internal library organization. The divisions in faculty con-

“The underlying assumption inherent in the decision to work with the composite factor rather than the unique variable is that the underlying dimensionality of thinking is of paramount interest rather than the specifics of the individual question.”

ceptualization among computerized access, circulation, special collections/services, and general collection do not closely mirror present traditional organizational structures. The linking of user education and staff resources with specialized collections/services provokes thought and discussion about the relationship of the library organization to faculty perceptions.

Third, the emergence of the computerized access category is a gratifying sign that faculty, who are traditionally interested primarily in local library holdings, may be thinking in more global terms for access to information. In fact, if the instrument had included questions related to interlibrary borrowing policy or services that had loaded highly on Factor 1, the category might have been named External Access. A fourth application of factor analysis is to improve future research instru-

ments and methods through further exploration and specification of the underlying dimensions of response. Variables related to externally-focused services should be included in future surveys, either to test the validity of the external access interpretation or to improve the precision of the survey instrument. Finally, the factor analytic solution may be used as a point of comparison for other studies, possibly about student perceptions, at this or other institutions. As technological changes permit the delivery of innovative and unforeseen services, a replication of this study would be advantageous.

CONCLUSIONS

Factor analysis was a useful technique to investigate the dimensionality underlying

the attitudes of the faculty at the University of Tennessee, Knoxville, with regard to library policies. In general, faculty think of library policies in four dimensions. These dimensions—computerized access, circulation, specialized collection/services, and general collection—summarize faculty attitudes on nineteen policy questions.

The utility of factor analysis as a data-reduction strategy is well documented throughout the social sciences. As statistical techniques become more prevalent in the assessment of library survey data, the use of this along with other methods, should facilitate the understanding of attitude structures of various library clientele.

REFERENCES

1. F. William Summers, "History and Development of the Survey Model for Planning," *Drexel Library Quarterly* 21:41 (Fall 1985).
2. Michael Halperin and Maureen Strazdon, "Measuring Students' Preferences for Reference Service: A Conjoint Analysis," *Library Quarterly* 50:208-24 (Apr. 1980).
3. William E. McGrath, Ralph C. Huntsinger, and Gary R. Barber, "An Allocation Formula Derived from a Factor Analysis of Academic Departments," *College & Research Libraries* 30:51-62 (Jan. 1969).
4. William Lyons and Linda Phillips, *Library Survey Summary* (Bethesda, Md.: ERIC Document Reproduction Service, 1989; ED 297 742, IR 052 414).

APPENDIX A. LIBRARY POLICY QUESTIONS FROM UTK LIBRARY SURVEY

Libraries at UTK: A Survey of Faculty Opinion

Please circle the number which corresponds to the response which best reflects how you feel. Space for additional comments is provided at the end of the survey.

17. In this section, we would like to know how you feel about various library policies.

[In the actual instrument the following scale was placed to the right of each statement.]

STRONGLY DISAGREE	DISAGREE	NEUTRAL	AGREE	STRONGLY AGREE
1	2	3	4	5

- a. The library should spend any available money on its collection, even if service declined.
- b. When acquisition funds are limited, the library should purchase books "on approval."
- c. It is very important that the library increase the accessibility of national databases.
- d. The main library should be open twenty-four hours a day.
- e. The main library should be open every day, even when classes are not in session.
- f. Periodicals should be allowed to be checked out.
- g. Nonprint materials should be allowed to be checked out.
- h. The library should provide a pickup and delivery service to academic units.

18. Below is a list of improvements that have been suggested for the library. Given the fact that funds are limited, please indicate the priority status that you feel should be assigned to each.

[In the actual instrument the following scale was placed to the right of each statement.]

NOT A PRIORITY 1	A LOW PRIORITY 2	A HIGH PRIORITY 3	ONE OF THE HIGHEST PRIORITIES 4
------------------------	------------------------	-------------------------	---------------------------------------

- a. Increased hours
- b. Increased book acquisitions
- c. Increased journal subscriptions
- d. Increased staff
- e. Increased audiovisual acquisitions
- f. Increased manuscript and rare book acquisitions
- g. Increased reference holdings
- h. Increased education of the user community
- i. Subsidized computer literature searches
- j. Ability to send messages to and from library via personal computer
- k. Increased access to computerized reference services

Introducing the New Standard in Research Indexes

Composite Index for CRC Handbooks 3rd Edition

Now, consult a single index containing 600,000 entries to locate data from over 140,000 pages of scientific information. Introducing the *Composite Index for CRC Handbooks, 3rd Edition*, a merged, unified three-volume index representing over 300 CRC Handbook volumes. Never before has such a comprehensive and useful reference resource been offered to the scientific community. CRC Handbooks include topics from the fields of biomedical sciences, biology, chemistry, engineering and computer science, physics and mathematics. By including subjects, compounds, and names of organisms, the Composite Index helps you locate information within these topics quickly, saving you valuable research time.

CD-ROM

The entire *Composite Index* has been transferred to a CD-ROM disk that is included with the printed index volumes at no additional charge. The CD-ROM's search program offers superior search capabilities that make finding information in the *Composite Index* faster, easier, and more efficient than you ever thought possible. Electronic bookmarks, on-line program help, customization options, and output capability are only a few of the added features included with the CD-ROM. Menu-driven and designed to run on IBM PC/XT/AT, PS/2 or 100% IBM-compatible computers, the CD-ROM version of the *Composite Index* is simple to use and maintain.

One-Day FAX Page Delivery Service

Need an article from a CRC Handbook fast? CRC Press's One-Day FAX Service puts Handbook information in your hands within 24 hours. This unique service provides libraries with access to every CRC Handbook—over 140,000 pages of scientific data—without purchasing all of the Handbooks. The cost per FAX request is a mere \$5.00 for the first two pages and \$1.50 for every page thereafter (telecommunication charges included). Your first year's annual user's license fee is included in the initial purchase price. Obtaining research information has never been so easy!

Annual Supplements

To keep the *Composite Index* in pace with the rate of scientific developments, a yearly supplement is available on a subscription basis. This supplement will consist of a 200-300 page hardbound volume, a CD-ROM disk containing the entire updated *Composite Index*, as well as a renewal of the user's license for the One-Day FAX Service.

Whether you are a scientist or student, librarian or engineer, the *Composite Index for CRC Handbooks, 3rd Edition* provides fast, easy access to information from CRC Handbooks. Libraries should consider the *Composite Index for CRC Handbooks, 3rd Edition* an indispensable reference acquisition.

Pre-Publication Special

SAVE \$200!

Order now through December 15, 1990 and receive \$200.00 off with our special pre-publication price of \$995.00 — Catalog No. 284ZV.



CRC Press, Inc., 2000 Corporate Blvd., N.W., Boca Raton, FL 33431
TO ORDER: Call toll free 1-800-272-7737 Monday through Friday
(Continental U.S. only). In Florida, call collect 407-994-0563.