DOCUMENT RESUME

ED 259 687 HE 018 615

AUTHOR	Seppanen, Loretta J.
TITLE	Using Relational Data Base Management Systems Capabilities to Increase the Usefulness of Open-Ended Survey Responses. AIR 1985 Annual Forum Paper.
PUB DATE	May 85
NOTE	21p.; Paper presented at the Annual Forum of the Association for Institutional Research (25th, Portland, OR, April 28-May 1, 1985).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Attitude Measures; Computer Oriented Programs; *Computer Software; *Databases; Higher Education; Management Information Systems; *Microcomputers; Research Methodology; Research Problems; *Statistical Analysis; *Surveys
IDENTIFIERS	*AIR Forum; *Open Ended Questions

ABSTRACT

The use of Relational Database Management Systems (RDBMS), a type of microcomputer application software, to analyze open-ended survey questions is discussed. Using open-ended questions allows researchers to ask respondents to express themselves freely about their attitudes and beliefs. This approach also can elicit a precise answer even though the list of possible answers is too large to list in the instrument or too long for most respondents to read (such as student major or home town). The characteristics of an RDBMS that allow for analysis of open-ended questions are: flexibility to create fields after the database has been designed, ability to join databases, and ability to sort on any field in the database. RDBMS can handle unstructured data and can use the relational operators "join" and "project" when using CONDOR, or the relational operator "copy" when using dBASE. The RDBMS can be used to clarify who is responding to open-ended questions in surveys, thus making the comments more useful, even in cases of underenumeration. The underenumeration problem can also be approached through effective design techniques. Another feature of RDBMS is that it allows the creation of a data entry screen. In addition to identifying technical considerations related to the use of RDBMS, an example of the automated Q-sort is provided. (SW)

ED259687

USING RELATIONAL DATA BASE MANAGEMENT SYSTEMS CAPABILITIES TO INCREASE THE USEFULNESS OF OPEN-ENDED SURVEY RESPONSES

April 24, 1985

Dr. Loretta J. Seppanen

Systems Information Manager

Washington State Board for Community College Education

319 7th Ave., FF-11

Olympia, Washington 98504-3111

(206) 753-3685

JC 850 315

THE ASSOCIATION FOR INSTITUTIONAL RESEARCH

This paper was presented at the Twenty-Fifth Annual Forum of the Association for Institutional Research held at the Portland Hilton in Portland, Oregon, April 28-May 1, 1985. This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of Forum papers.

Elizabeth F. Fox, Chair
Forum Publications Editorial
Advisory Committee

Abstract

A Relational Data Base Management System (RDBMS) and good instrument design can be used together to overcome the problems previously associated with analysis of open-ended questions in surveys.

Using open-ended questions allows the researcher to

* ask respondents to express themselves freely related to their attitudes and beliefs, especially to clarify a closed-ended evaluation or to explore a previously unresearched topic, and
* elicit a precise answer even though the list of possible answers is too large to list in the instrument or too long for most respondents to read (such as "student major" or " home town").

The characteristics of an RDBMS which allow for analysis of open-ended questions are the

* flexibility to create fields after the database has been designed,
* ability to join databases, and
* ability to sort on any field in the database.

This paper describes how CONDOR, an RDBMS, is used to allow efficient analysis of open-ended survey questions.

Why We Use Closed-Ended Survey Questions

Higher education researchers use surveys to determine the

* satisfaction of current and former students;
* satisfaction of employers and transfer institutions;
* attitudes of students, staff, and the service community toward policy issues; and
* the needs of current and potential students and employers.

A review of the surveys used in most of our colleges would reveal a strong bias for structured closed-ended questions. Even factual information is not requested in an open ended-fashion in many surveys. Instead, respondents are asked to provide their occupation code, for example, from a lengthy list of such codes.

What has inclined us as researchers to use such structured closed-ended questions? I believe there have been two forces at work on our thinking and behavior:

* The **desire to make research** on human **behavior more scientific**, and therefore to assure that data is more quantifiable, and
* The **technological changes** which have allowed researchers to manipulate large volumes of quantified data on human behavior.

The former force leads us to mistrust qualitative evaluation of student, faculty, and service community opinion because it increases our awareness of the role of selective perception and evaluation of data. Taken to extreme, this orientation leads us to basing change only on the results of well-designed surveys with appropriate controls to prevent undersampling and oversampling and statistical error, rather than upon educated hunches about what should happen in a given environment.

The second force is technological change. Along with the desire for more scientifically based decision making regarding human behavior, came new developments in computers which allowed researchers to automate statistical analysis.

The researchers' love affair with numeric analysis of human behavior heated up just about fifteen years ago when SPSS (Statistical Package for the Social Sciences) was first up and running on a few university mainframes around the country. Now that most researchers do even their frequency analyses on SPSS or SAS (Statistical Analysis System), we require that our data be codable in a form these systems can read - simple numerics.

This love affair has inclined us toward pre-coded data, which requires that the researcher determine in advance which categories the respondent will choose. Even post-coding done by technical staff is based on codes established by the researcher before all responses have been studied.

Implicit in this numeric orientation is an assumption that we know how to categorize data even before we look at it. In part, this is a carryover from research designs in which correlations are intended. But, in college surveys, most analysis consists of purely descriptive statistics: frequencies and cross-tabulations. Pre-coded data is not essential for descriptive analysis.

In fact, much of our survey work in colleges would be enhanced if we, as researchers, would let the data speak for itself - find the tone in what respondents say - not just the frequency with which they say it.

The open-ended survey question is one way of allowing for the kind of openness in collecting data that we need. But, how can we systematically analyze the responses to such questions? If the questions are not properly structured or cannot be cross-tabulated with other closed-ended questions, we will have problems with underenumeration, that is, too few respondents answering the question to be safe in concluding that their responses are representative of the total sample (Dillman, 1977). For these reasons, it is not suprising that some advice givers simply suggest not using open-ended questions at all (Pride, 1983).

At least one analytic process has been developed which allows for minimal bias in the review of open-ended questions. I remember the technique, called Q-sort, from my graduate school Research Methods class. You may recall how it goes. The researcher writes all the open-ended statements of respondents on individual 3 by 5 cards. Then she reviews all the cards to let the data suggest a method of organization and categorization. When the data speaks to her (so to speak) she sends the cat to the nearest kennel, stocks the refrigerator, and starts the days of question sorting (Q-sort) by placing each statement in its proper pile somewhere on the living room floor. Large piles are saved for future weekends (when the cat is out of the house) for another iteration of the same process. Finally, after weeks of work and an estranged pet, all comments are properly sorted, bound with rubber bands, and ready for use in report writing.

This method may work for graduate students in a research class, but it is not very practical in a busy institutional research office. If the whole sorting process could be automated, just as SPSS or SAS automated all the number crunching just a few years ago, researchers would have an avenue for analysis of open-ended questions.

Automating the Q-sort is what this paper is about.

Criteria for Use and Analysis of Open-ended Questions

Researchers have learned that open-ended questions are extremely valuable in specific settings as summarized by Dillman, 1984:

* exploratory research where the objective is to find the most salient aspect of a topic for use in closed-ended questions in later studies
* when respondents need to vent frustrations or state strong opinions
* in partially closed-ended questions where the explanation of the option "other" is desired
* when it would be unnecessarily time-consuming for the respondent to read a long list of possible responses for a closed-ended question (i.e., make of car), and
* clarifying closed-ended responses.

An automated Q-sort should be able to categorize each comment after reviewing all the responses. Just as in the manual process, the automated approach must be repeatable again and again.

To determine the impact of underenumeration, there is a need for the automated process to do much more than was possible in the manual setting. Specifically, responses to open-ended questions must be cross-tabulated with closed-ended questions to better determine the inclinations of those who make comments.

The automated Q-sort needs to be a system flexible enough to allow categories to be developed after the data has been reviewed. No standard file management system nor hierarchical data base management system can achieve these objectives.

Using a Relational Data Base Management System

The type of microcomputer application software which will meet the need for analysis of open-ended questions is a Relational Data Base Management System (RDBMS).

Kruglinksi (1983) characterised RDBMS as a data base product with the following features:

* Allows operations on an entire database with a single command
* Does not require that all information needs be planned in advance. In fact, relationships are specified at the time of inquiry rather than in advance

* Contains the relational operators "project" and "join". "Project" is an operation which creates a new relation by selecting a subset of the existing relation, and "join" simply combines two separate relations.

Examples of microcomputer RDBMS are dBASE II and dBASE III, R:base, and CONDOR. The author's experience is entirely with CONDOR; however, the principles described here apply to any true RDBMS whether on a micro or larger computer.

The characteristics of the RDBMS which allow it to automate the Q-sort process are:

* use of the relational operators "join" and "project" (or "copy" in dBASE), and
* the ability to handle unstructured data.

An Example of the Automated Q-Sort

Just as in the manual Q-sort, the first step when using RDBMS is to review the data in total to determine what categories present themselves. In the automated situation, the comments of all respondents to any single open-ended question will be available on a screen or in a nicely printed list. The "magic" required to avoid the step of creating all those 3 by 5 cards is described in the **Data Entry Process** section of this paper.

Let's use as our example a factual type of open-ended question from a student follow-up study: What is your current occupation? Here is part of what the researcher would see on the database screen:

Accountant

Accounting technician

Administrative Assistant

Admin Asst

Artist

Bookkeeper in spouses's business

Commercial fisherman

Floor nurse, RN

Sales clerk

School aid

Secretary

.eacher

The researcher reviews the list of comments and creates categories. Once categories are determined, the researcher puts a code by each sentence representing its code in the category system. This is a process which can be easily reiterated, so the first time through, the coding may be as simple as positive vs. negative comments.

In this occupations example, the researcher might want to code the level of education generally required for the occupation. At this stage it is helpful to make categories consistent and to tally the number for occupations with more than one response:

BA Accountant

AA Accounting technician

AA Administrative Assistant 2

BA Artist

AA Bookkeeper in spouses's business

HS Commercial fisherman

AA Floor nurse, RN

HS Sales clerk

HS School aid

AA Secretary

BA Teacher

Now - and this illustrates the flexibility of an RDBMS - a new field is created in the database for this new code. Because this field now exists, the researcher can do a wide variety of useful things:

* List only the occupations which require an associate degree level

    Accounting technician

    Administrative Assistant 2

    Bookkeeper in spouses's business

    Floor nurse, RN

    Secretary

* Q-sort this new list again into subject area categories which match the degrees offered at the college. This reiteration of the process requires the creation of yet another field in the database

  Certificate level accounting occupations:

  Accounting technician

  Bookkeeper in spouses's business

  Clerical occupations:

  Administrative Assistant 2

  Secretary

  Allied Health occupations:

  Floor nurse, RN

* Cross tab the occupations with any other question in the survey, for example, to find out if more of one sex tended to be in certain occupations, or if those with "personal interest intent" are in different occupations from those with "transfer" or "job related" intents

  Personal Interest Student's Occupation:

  Accountant

  Artist

  Bookkeeper in spouses's business

  Sales clerk

  Teacher

BEST COPY

All of these functions can be done in the microcomputer, or the new code can be made numeric and transferred to a mainframe statistical package for analysis. The microcomputer method is used when there is a preference for retaining the exact words of the respondent, not just the category of response.

The strategies illustrated above apply equally well to sentence-long comments from respondents as they do to factual statements.

Instrument Design for Open-ended Questions

The RDBMS can be used to clarify who is responding to open-ended questions in surveys, thus making the comments more useful, even in cases of underenumeration. The problem of underenumeration can also be approached through effective design techniques.

I find the highest response rate is on partially closed-ended questions, those which ask the respondent to specify the meaning of "other" or to list something factual. These questions get at least a 95% response rate, which is about the same as closed-ended questions. For example, 99% of respondents who are employed provide a meaningful response to the occupation question.

My experience indicates that, except for populations that are hesitant about surveys in general, 75% to 99% will respond to a general open-ended question placed near the end of a well designed study which asks for specific recommendations for the program/college. These narrative statements help to capture the general attitude of respondents more clearly than the closed-ended responses.

To overcome problems of misunderstanding the open-ended question, I structure opportunities for comments throughout the evaluation sections of a survey. For example, if the respondent is evaluating the helpfulness of four student services, they are offered space to comment on each service immediately after evaluating the service as shown in the sample below:

JOB PLACEMENT SERVICES Comments:	VERY MUCH	SOMEWHAT	VERY LITTLE	DID NOT USE
ADMISSIONS AND RECORDS OFFICE (getting transcripts) Comments:	VERY MUCH	SOMEWHAT	VERY LITTLE	DID NOT USE
COURSE CATALOG, SCHEDULE & NEWSPAPER ADS FOR REGISTRATION Comments:	VERY MUCH	SOMEWHAT	VERY LITTLE	DID NOT USE
WOMEN'S CENTER (program for men & women) Comments	VERY MUCH	SOMEWHAT	VERY LITTLE	DID NOT USE

This method also results in underenumeration problems as most respondents do not care to comment on everything, but each comment is focused. Additionally, I have found it helpful to present the open-ended responses sorted by how the respondent commented on the closed-ended question.

The Data Entry Process

So far, the comments of respondents have somehow magically arrived on the researcher's computer screen for review and for use in the automated Q-sort. Actually, the process of data entry is quite simple and efficient.

The RDBMS allows for creation of a data entry screen. CONDOR allows for design of an especially attractive data entry screen. Here is an example of a screen display for entry:

```
****************CON'T AK NATIVE STUDENT SURVEY F84 - SCREEN 4********************

 ID 11    REC.4

 13.1 1  13.1A 111111111111111111111111111111111111111111111111111111111111111
 13.2 1  13.2A 111111111111111111111111111111111111111111111111111111111111111
 13.3 1  13.3A 111111111111111111111111111111111111111111111111111111111111111
 13.4 1  13.4A 111111111111111111111111111111111111111111111111111111111111111

 14 1  15 1  16 1  16OTH 11111111111111111111111111111111111111111111111111111
 16.A 1  17 11  18 1  19.1 1  19.2 1  19.3 1  19.4 1  19.5 1  19.6 1  19.7 1
 20 1  21 11111111111111111111111111111111111111111111111111111111111 22 1
 23 11111111111111111111111111111111111111111111111111111111111111111
```

The "1" represents spaces on the sceen (inverse video) which can be filled with characters by the data entry worker. Each set of "1" is a data field which may have specifications as to whether the entry can be alpha, numeric, or both. If numeric, the range of numbers which can be entered can be pre-set.

The long "1" units are open-ended comment spaces which follow immediately after the related closed-ended question. The data entry person simply types in the respondent's statement. In the rare case of a wordy respondent, the data entry person abbreviates the response to fit in the allowed space.

Once the data entry is completed, the "project" command is used to separate the numeric fields from the alpha fields. Numerics are transferred to the mainframe computer for standard analysis by a statistical package. Two databases now exist:

* the first in the mainframe with numeric information only, and
* the second in the microcomputer RDBMS with all the information.

One additional advantage of the RDBMS is that data is entered only once. The comments of respondents, once entered into the database, can be used in the final report without ever being typed again. Typos can be corrected in either of two places, 1) in the database on the data entry sceens or 2) when list of statements is transferred to a wordprocessing file for final report writing. This feature greatly reduces the amount of clerical work involved in preparing reports for surveys.

<u>Technical Considerations Related to Use of RDBMS</u>

Each RDBMS has limitations on:

* Number of characters per field,
* Fields or bytes per record,
* Screens per record, and
* Records per database file.

These characteristics place limits on the researcher which may require some creative planning.

In this review, I will comment only on the limits of CONDOR 20-3.

* CONDOR allows 127 characters per field, which implies that comments of respondents could be as long as 127 characters, a fairly wordy sentence. Don't believe it for one minute.

  While it is possible to enter and list a 127 character field, CONDOR is really structured on an 80 column card concept. Consequently, the system places carriage returns at the end of every 80 column line. Once the researcher starts "joining" and "projecting" a few times, these unwanted carriage returns will create havoc with the comment field. I choose, instead, to use comment fields of about 65 characters. When longer comments are likely, I allow more than one field per comment.

* CONDOR all 127 fields and 1024 bytes per record.

Since it would be very hard to get 127 fields on one screen, the field limit is not a problem. Also, since CONDOR stores numbers very economically, the byte length is also not a problem.

* CONDOR is limited to 1 screen per record. The one screen per record limit is a problem. Most lengthy surveys require two or more screens for data entry. Fortunately, the "join" command allows all these records from different screens to be combined as needed during data manipulation.

Each RDBMS will have unique limits and strengths in use for survey data entry and open-ended analysis.

While every survey done in a college setting is unique, much of the data manipulation to be done in the RDBMS is repetitious. This is where macro type commands can be used to save considerable time.

CONDOR allows the researcher to program the PROJECT, JOIN, SORT, SELECT, PRINT commands needed for a particular analysis and save that work for future sessions which require similar functions.

Since the microcomputer RDBMS works reasonably slowly, this programmable feature allows the researcher to set up a few requests for information, set the system to work, head for lunch, and come back an hour later with pages of useful analysis ready for final review.

References

Dillman, D. A. Mail and Telephone Surveys: The Total Design Method. New York: John Wiley & Sons, Inc., 1978.

Kruglinski, D. Data Base Management Systems: A Guide to Microcomputer Software. Berkeley: Osborne/McGraw-Hill, 1983.

Pride, C. Let's take a survey. Attitude and Opinion Research: Why You Need it/How to do it. Washington, D.C.: Council for Advancement and Support of Education, 1983.