

Allan H. Church
Janine Wacławski

Foreword by Allen I. Kraut

Designing and Using Organizational Surveys

A Seven-Step Process



JOSSEY-BASS
A Wiley Company
San Francisco

Copyright © 1998 by Allan H. Church and Janine Waclawski.

Published by Jossey-Bass

A Wiley Imprint

989 Market Street, San Francisco, CA 94103-1741 www.josseybass.com

First Jossey-Bass edition published in 2001. This book was originally published in 1998 by Gower Publishing Limited, England, under the title *Designing and Using Organizational Surveys* by Allan H. Church and Janine Waclawski.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400, fax 978-646-8600, or on the Web at www.copyright.com. Requests to the publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, 201-748-6011, fax 201-748-6008, or online at <http://www.wiley.com/go/permissions>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Readers should be aware that Internet Web sites offered as citations and/or sources for further information may have changed or disappeared between the time this was written and when it is read.

Jossey-Bass books and products are available through most bookstores. To contact Jossey-Bass directly call our Customer Care Department within the U.S. at 800-956-7739, outside the U.S. at 317-572-3986, or fax 317-572-4002.

Jossey-Bass also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

Church, Allan H., date-

Designing and using organizational surveys: a seven-step process / Allan H. Church,
Janine Waclawski.—1st ed.

p. cm.—(Jossey-Bass business & management series)

ISBN 0-7879-5677-5

1. Employee attitude surveys. I. Waclawski, Janine, date- II. Title. III. Series.

HF5549.5.A83 C483 2001

658.3'14'0723—dc21

00-011967

Printed in the United States of America

FIRST EDITION

HB Printing

10 9 8 7 6 5 4

The Jossey-Bass
Business & Management Series

Contents

Tables, Figures, and Exhibits	xiii
Foreword	xvii
<i>Allen I. Kraut</i>	
Acknowledgments	xxi
The Authors	xxv
Introduction	1
<i>What Is a Survey?</i>	4
<i>A Brief History of Surveys</i>	8
<i>Contemporary Use of Surveys</i>	10
<i>Surveys in Contemporary Organizational Life</i>	12
<i>The Seven Steps to Effective Organizational Surveys</i>	17
1. Step One: Pooling Resources	27
<i>Setting Clear Strategic Objectives</i>	31
<i>Obtaining Commitment</i>	38
<i>Overcoming Resistance and Apathy</i>	42
<i>Maintaining Confidentiality</i>	43
<i>Deciding What Information to Collect</i>	45
<i>Balancing Priorities</i>	46
<i>Checklist for Step One</i>	49
2. Step Two: Developing a World-Class Survey	51
<i>Using a Survey Design Team</i>	53
<i>Gathering Preliminary Information</i>	55

Identifying Key Issues	56
Discussing Your Findings	58
Drafting the Initial Survey Document	60
Piloting the Survey	84
Checklist for Step Two	87
3. Step Three: Communicating Objectives	89
The CPR Model of Organizational Communication	91
First Contact with Employees	96
Communicating the Survey	100
Sample Survey Introduction	106
Guidelines for Communicating to Employees	107
Recognizing Informal Systems	109
Checklist for Step Three	111
4. Step Four: Administering the Survey	113
Timing of Administration	114
Working with the Project Plan	116
Sample Versus Census	120
Methods of Administration and Data Collection	122
Paper Versus Electronic Methods: A Comparison	137
Response Rates	143
Learning While Doing	146
Checklist for Step Four	147
5. Step Five: Interpreting Results	149
The Role of Statistics	150
The Importance of Timing	154
Data Entry	159
Data Preparation	162
Item-Level Analysis	172
Conceptual-Level Analysis	178
Comparative Analysis	186
Content Analysis of Write-In Comments	193
Checklist for Step Five	199

6. Step Six: Delivering the Findings	201
<i>Understanding the Roll-Out Process</i>	203
<i>Preparing the Survey Report</i>	207
<i>Balancing Expectations and Reality</i>	225
<i>Checklist for Step Six</i>	227
7. Step Seven: Learning into Action	229
<i>Using Surveys to Create Lasting Change</i>	232
<i>Barriers to the Transfer of Ownership</i>	233
<i>A Commitment to Action</i>	239
<i>Four Approaches to Survey Action Planning</i>	241
<i>Five Critical Factors That Determine the Success</i>	
<i>of Survey Action Planning</i>	258
<i>The Action Planning Process</i>	259
<i>Linking Survey Results to Other Measures of Performance</i>	267
<i>Building Systems for Evaluating Success</i>	271
<i>The Evolving Role of the Survey Practitioner</i>	272
<i>Checklist for Step Seven</i>	276
References	279
Index	287

Step Five: Interpreting Results

Look to the essence of a thing, whether it be a point of doctrine, of practice, or of interpretation.

Marcus Aurelius Antoninus

The prospect of statistically analyzing data obtained from thousands of responses collected during a survey can be daunting for many people, at least at first. Most people are simply not used to working with so much raw information. Once the initial shock of the size and complexity of the dataset has been absorbed, however, the budding practitioner will find that the analysis and interpretation phase is really a process of creating and communicating a compelling story regarding the current state of the organization (or whatever the survey topic).

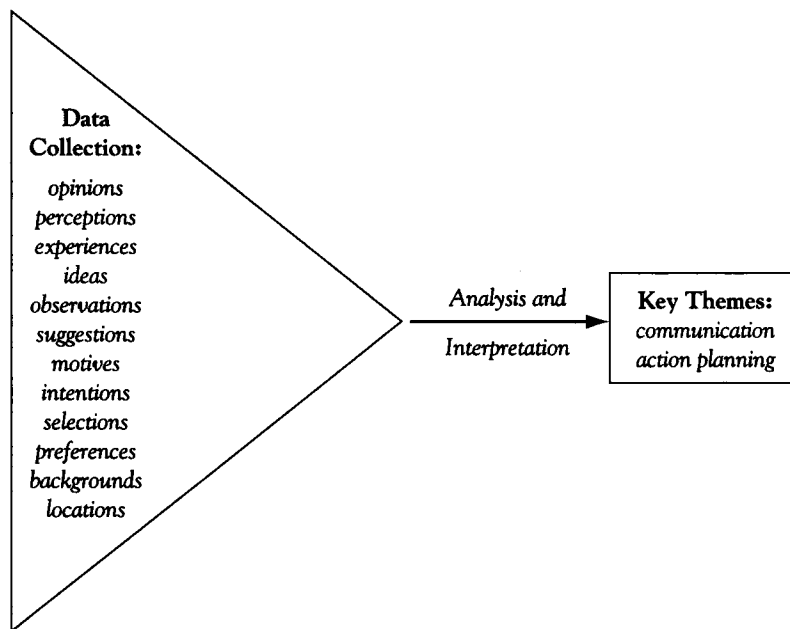
This step is concerned with discovering what is going on and why. In addition to detailing the stages involved in the analysis process, this chapter covers some broader issues related to survey interpretation, such as how to balance client expectations with realities (and limitations) inherent in the data, the use of norms and benchmarking data, and how to use write-in comments to enrich the interpretation—the compelling story that is the sum total of the data analysis process.

Before exploring any dataset, remember the basic goal of your efforts during this phase: information reduction. Whereas the concentration during both the survey development and the administration stages has been on collecting more rather than less information, now it is time to compress the volumes of opinions, attitudes, and perceptions into a single coherent picture for the client or for your own organization. When you begin to interpret survey results, you are engaging in a process of data synthesis; the intended outcome is an accurate and meaningful distillation of a few main themes, trends, and issues, as they are reflected in the larger set of responses (see Figure 5.1).

The Role of Statistics

Survey interpretation often involves the effective use of advanced statistical techniques such as multiple regression, item response

Figure 5.1. The Distillation Process.



analysis, analysis of variance, factor analysis, reliability analysis, and even structural equation modeling. But these sophisticated tools are not *required* to produce an interesting and compelling set of findings. Like other types of tools, statistics are only helpful to people who know how to use them and understand their strengths and weaknesses. Sometimes the uninformed use of statistics can go a long way to getting the survey practitioner into trouble. For example, someone knows just enough to ask the right questions but not enough to fully understand the complexities inherent in the answers. We have been involved in a number of survey analysis efforts (sometimes being called in just for this portion of the effort). One client's recurring plea was, "Tell me which findings are statistically significant." This is a valid query. However, statistical significance is affected by four primary factors:

1. The relative strength of the observed effect (the difference between mean scores for a given pair of questions or between averages across a series of items for two different regions or functions)
2. The amount of variance obtained in the measures (the degree of similarity or disparity among individual responses to the same questions)
3. The total size of the dataset (one hundred employees versus ten thousand employees)
4. The degree of precision or confidence required to state that a result is indeed statistically significant (for example, 90 percent confident, 95 percent confident, 99 percent confident)

Therefore the question does not always result in a completely straightforward answer, which in turn, often confuses the client even more. To demonstrate this point, see Table 5.1, which is part of a letter provided in response to this very issue that we sent to one of our clients during a large staff survey effort.

Table 5.1. Response to Question on Statistical Significance.

In response to your request regarding significant differences among 1993 and 1994 survey data, as well as within a single report (directorate vs. department, item vs. item, and so forth), I have conducted a number of test analyses in order to provide you with the following rules of thumb. Please keep in mind that the information listed below is only a set of general recommendations for determining important and meaningful differences. Thus, in some cases, not all significant differences may be meaningful to consider (particularly if the sample sizes are very large), and not all nonsignificant differences may be unimportant.

Although the degree of variability among the different items does vary considerably, given the number of comparisons and the use of a similar scale (1-5) throughout most of the items, the primary consideration in determining a statistically significant difference among the survey results is the sample size (or number of respondents).

Sample Sizes for Each Comparison Group	Approximate Difference in Means Needed to Reach Statistical Significance
500 and above	.2
300-500	.3
150-300	.4
50-150	.5
20-50	.6

These problems with the use of statistics can be compounded further when individuals with limited skill in this area attempt to use some type of statistical computer package to provide statistical results without fully understanding the important assumptions and parameters involved in their use. These programs always produce any number of pages of output, particularly when using more recent versions of popular (and advanced) statistical analysis programs that are fully menu-driven and have interactive help screens regarding analysis commands.

Moreover, the inappropriate application of a specific technique is not the only problem with these easy-to-use programs. While working with another consulting firm on the analysis of a recent survey effort, for example, one of the more junior (and non-statistically trained) consultants who had access to the full dataset of more than twenty thousand cases managed to perform, via drop-down menus, a series of simple recodes and select-if statements to answer a specific question. Unfortunately, he forgot that these data screens had been put in place when he started to reanalyze the data for another request, and there was no program audit trail to follow. The result was an entirely spurious set of findings that required significant senior-consultant-level intervention and correction at a later date.

Thus good advice in cases like this is to use only programs and procedures with which one is very familiar. If more advanced analyses are desired or required, the survey practitioners should find someone with experience in survey analysis to assist in the process. And if experience has shown us anything, we strongly recommend keeping a formal programming audit trail (or using a program code file instead of those deceptively easy drop-down menus) of every step taken in a survey analysis process from transformations to labels to recodes to option used on statistical procedures. This way the results can always be recreated by the survey practitioner or other interested parties months, even years later when returning to the dataset.

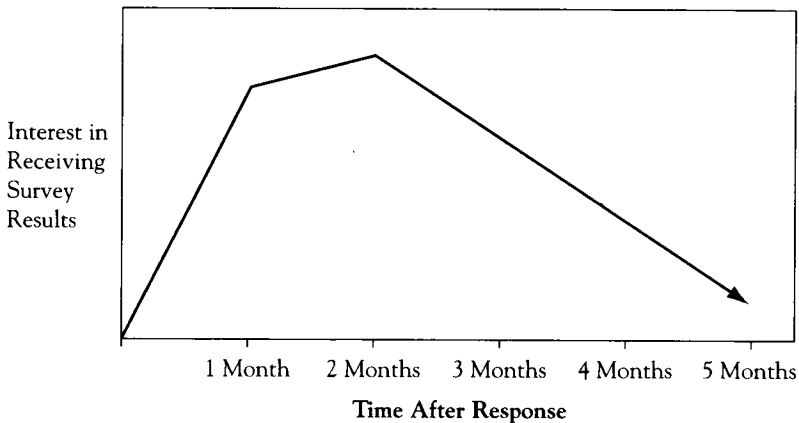
Keep in mind, however, that the knowledge and use of advanced research methods and data manipulation tools alone do not guarantee a useful or actionable interpretation process. As with most other things in life, survey analysis and interpretation requires a combination of both skill and experience in working with this type of information. Many people in organizations who have advanced, research-related degrees but limited survey experience would probably be willing to run complex analyses if asked. In many cases these same individuals will likely discover that it is much more difficult to create a complete and compelling story based on survey results

than it is to simply run their favorite statistical procedures using a computer. This is one reason that some external organizational consultants and some internal practitioners specialize in the conduct and interpretation of organizational survey efforts.

The Importance of Timing

Regardless of the level of expertise and statistical knowledge being brought to bear in this process, it is an unfortunate reality of most survey efforts, at least in our experience and that of others (for example, Kraut, 1996b), that the analysis and interpretation stage is allotted the *shortest* amount of time in the total project plan. It might be desirable to have a two- or three-month period in which to conduct a comprehensive set of analyses and fully examine the data using all possible options, but this is not usually an option. Most survey participants want to see the results as soon as possible, and this results in pressure on the survey analysis team. Managers and employees seem to understand that it can take several months to develop the appropriate questions for use in the survey instrument itself; however, once they have completed their survey and dropped it in the mail, the desire to see the results is almost immediate and tends to peak in just a few weeks (see Figure 5.2). Because people often complete their surveys at different times during an extended administration period, it can easily be a month between receipt of the first and the last survey response forms. By the close of the administration process, those people responding early are already eager to see the results and have been waiting several weeks. Moreover, in some organizations with advanced electronic survey collection and reporting processes, results are expected within a few days or even overnight. Whatever the timeframe and process used, the energy created by participation in a survey effort can indeed be a positive force for organization development and change (for example, Nadler, 1977), but it can also become a barrier when the feeling turns to one of either distrust or apathy regarding the outcome of the assessment process.

Figure 5.2. Relationship Between Response Timing and Level of Interest in the Results.



These effects translate to a need for the survey team to be able to work hard and fast during the analysis and interpretation stage. The time from final data collection to the first presentation of an overall report describing significant findings and trends to the survey sponsor or senior management is typically about two to three weeks from the close of the survey effort (in order to capitalize on as many returns as possible). Although automated on-line intranet or Web-based survey systems that allow almost immediate reporting can certainly shorten the report roll-out timeframe (particularly in the form of straightforward reports to individual units and managers), significant time and effort will still be required for the survey team to conduct a comprehensive set of analyses and recommendations for presentation to the senior management of the organization. Although some executives might simply want to pore over the item means and frequencies, if no compelling story is offered (based on a well-grounded and comprehensive analysis of the results), the impact of the survey effort is likely to be seriously limited. Unfortunately, developing that compelling story does take time, making effective time management a critical skill during this stage. Of course, this timing can be further compounded

by the need for content summaries of write-in comments if these are collected (as described later in this chapter). Whatever the approach taken, there will still probably be a lag in time between the initial delivery of the findings to various constituents (whether senior leadership or individual managers) and the full-scale communication to all organizational employees. The only exceptions to this rule are targeted survey applications (for example, perceptions regarding a relatively small work group, function, or limited topic) or in organizations where the survey process has become fully institutionalized and accepted. In the latter case, however, it may still be problematic to provide immediate results to some members without having a unified and fully developed interpretation of the results for the top of the house. In short, timing is everything. Keeping the timeframe between close of administration and final reporting as short as feasibly and ethically possible is an important factor in achieving success. Moreover, the shorter the time between when management sees the results and when the rest of the organization sees them is critical to building support and belief in the process, as well as its results.

In some instances the complexities of timing can be further compounded, particularly when senior managers are not pleased with the results obtained and have not readied themselves to accept criticism and acknowledge problems. In such cases, there is often a tendency to withhold the communication of the results to employees until after the message has been reworked and polished so that it better suits their particular objectives. Some level of hesitation and apprehension should be expected when delivering the first set of survey findings to the sponsor or senior management. However, if the data are held too long, the results can become meaningless to employees, which damages both the extent to which the findings will be taken seriously and the participation in future survey efforts. Rapid turnaround in the analysis and interpretation stage is required.

Because this is the only stage in the entire survey process that is almost completely in the hands of the practitioner and survey team, it should be done with an eye toward balancing speed and accuracy.

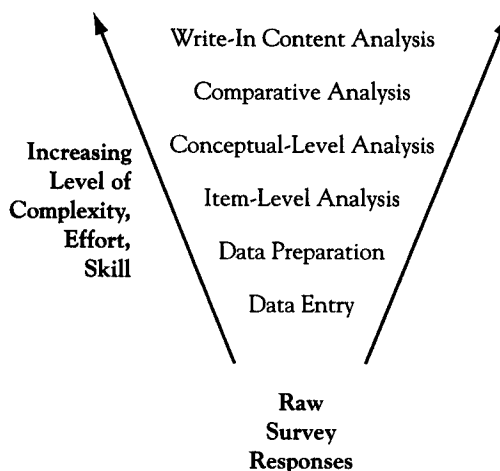
Often this means preparing much of the groundwork for analysis before the closing date for the survey passes by doing the following:

- Making sure the data entry system is fully operational and has been tested for all types of contingencies (assuming that an active entry system like voice response or on-line entry is not being used, in which case the entry system has presumably already been made fully functioning)
- Installing and testing all data analysis software
- Writing and testing initial analysis code using sample data (or preliminary data if possible or available)
- Obtaining results from previous surveys or related comparison data for benchmarking purposes
- Managing the clients' expectations and anxiety regarding issues with the collected data (for example, ensuring confidentiality of individual responses, tallying the number of calls and queries into the process, providing updates on daily and weekly response rates, and preparing clients for unexpected issues or problems that may arise in the data)
- Preparing a comprehensive and thoughtful data analysis plan (for example, becoming familiar with a theoretical or conceptual model regarding potential relationships among variables or large sets of items, specifying additional relationships or trends to explore, targeting initial and final review dates for various elements of the analysis process, assigning roles and responsibilities for various tasks)
- Developing a framework or outline of the final survey report. This is often based on the model used in the

data collection process and will help the analyst and the client prepare mentally and psychologically for the final product

Having prepared the groundwork, the heart of the analysis process is now at hand. The remainder of this step provides an overview of the six main stages involved in the data analysis and interpretation process. Figure 5.3 provides a graphic representation of these stages in the form of an inverted triangle. Note that as one moves through each of the steps, the level of effort, time, and complexity tends to increase. For the most part, this is because each stage requires greater levels of statistical processing and analysis capabilities. The process of analyzing write-in comments, however, is the most time consuming, primarily because of the nature of working with individual responses (often voluminous in nature), as opposed to simple numerics. A detailed discussion of the statistical procedures related to survey analysis is not included here, but pointers as to the utility of certain techniques and procedures, as well as references for more information, are provided during discussion of the various stages.

Figure 5.3. Six Stages of Survey Analysis.



Data Entry

The first stage for any sort of analysis or interpretation process requires that the survey responses be entered into a database. Regardless of the administration method used, the individual responses must be converted from their natural state, whether it be optical scan forms, e-mail responses, pen-and-paper survey questionnaires, fax-back scans, voice response data storage systems, or on-line intranet or Web-based response systems, and converted into some raw stream of numbers and text characters that can be read, manipulated, and analyzed by some type of sophisticated computer software package. Remember, however, that data entry can take time, particularly if the data are in pen and paper form and must be keyed. A variety of professional data entry services are available, but this method can be slower than other options, for example, various electronic data collection methods such as e-mail and on-line approaches (review Step Four); therefore, extra time may be needed to compile information this way. The process of transcribing write-in comments, for example, can be extremely time consuming and complex when multiple languages are involved. Although some applications do work with text (including scanned text from fax-backs and optical scan forms, as well as those entered directly into on-line databases), the quality of the interpretative summaries generated have yet to match good, old-fashioned human analysis.

In comparison, data entry using optical scan forms depends largely on the speed and quality of the scanning machine itself. With the latest and greatest Opscan model from vendors such as National Computer Systems (NCS), data can be scanned easily, quickly, and with a high degree of accuracy. Although various electronic systems and other automated (for example, voice response) data capture methods have gained considerable popularity in recent years and may well one day eclipse paper approaches, the optical scan method still remains the dominant mode for the majority of survey efforts. As discussed in Chapter Four, this is due in large part to a combination of

factors, including familiarity, ease, independence, and universality of response approach (it can be completed almost anywhere), perceptions of increased confidentiality, and accuracy. In any case, in order to maximize the time available for analysis, there is always a need for speed of entry as well.

There are undoubtedly a few practitioners lurking in the field who still prefer using hand calculations for the entire analysis process, but the advantages in both speed and applications of personal computers make the process of working with survey data relatively simple, at least in theory. The extent to which this conversion process proves difficult varies with both the method of collection and the sophistication of the technical staff. For example, although the technical people professed the speed and utility of an on-line, Web-based response system in one survey effort in which we were involved, when it came time to convert the data to numerical form for use in a database, it took them two weeks to download the information in a usable form. However, there is usually a way to extract the information in a manner that will be compatible with a variety of software packages.

A detailed discussion of the mechanics of these operations is beyond the scope of this chapter. Suffice it to say that most systems can read and write one of two common options: (1) American Standard Code for Information Interchange (also known as ASCII, pronounced "askey") text files, where each response is represented by a continuous stream of numbers followed by a fixed return, or (2) comma-delimited files (also known as CSV files), which are text files wherein each variable or question is separated by a comma in a long string of text. More sophisticated options include writing specific formats that are directly compatible with various popular database programs. Ensure that the individual managing the database system has a thorough understanding of the nature of the datastream being produced for analysis. The most important information for this process is as follows:

- *The exact order or sequence in which responses to questions are listed*, including any demographic or background variables (for example, years in the organization, age, location code, gender, department, or responses to survey questions should be included. In many cases the order of variables reflects the order in which the questions appear on the survey, but certain applications may require that items be listed in a variety of different sequences, such as when different sets of custom items are used for various functions or departments or when the order has been varied for formatting or research reasons.

- *The number of records or lines of data per individual*. Each line of data is followed by a hard return. In some cases and with some computer systems and applications, however, the number of columns needed to keep an entire set of responses on one line may exceed the number available. In these situations, multiple lines may be needed for each person's response.

- *The type of record format used to write the cases*. There are two options here. The first is called a fixed format in which the data contain exactly the same number of spaces for each variable or question, regardless of whether a response was obtained. With this type of datastream, it is easy to check that everything is in order with the raw scores because they always end on the same last column (and have the same number of lines per response).

The second option is called a variable format; the responses literally vary in length, or the number of spaces used is based on the pattern of responses obtained. Thus, if an individual answers only two questions on a survey of one hundred items, only two spaces would be used in his or her datastream. This type of formatting system is particularly useful when responses to questions might vary in length, for example, in a name, occupational title, or description. Note that in these types of files, some other indicator is required to

separate responses to individual questions. Typical options are to include a comma (as in the CSV file described earlier), a tab, or even a letter such as X or Z.

This information may seem simplistic; it is necessary nonetheless. If there are errors or miscommunications in the setup information, it is almost certain that the analyses conducted will be entirely spurious and meaningless.

Three types of software categories and a host of products within each category are currently available to the practitioner with respect to the type of program that can be used for data analysis and interpretation. Table 5.2 provides an overview of their strengths and weaknesses, along with some sample applications for each type.

The differences among these products were once considerable. However, in recent years the number and level of basic functions has begun to make them almost interchangeable for relatively simple and straightforward types of analyses. Features improve with each new version. Almost all of these programs, for example, have the capability to produce excellent, high-quality reports and graphics in their most recent versions.

The choice of a program for use in this stage is dependent on a combination of (1) the range of functions and types of analyses available in the program, (2) the nature and complexity of the specific questions and investigations of interest, and (3) the familiarity, experience, and skill level of the individual or survey team tasked with this process. The decision to use one or more of these programs should have been made before the data were entered.

Data Preparation

Some people think that once raw survey data have been entered into a database system they are automatically ready for analysis; in actuality, an important step must be taken to prepare the data for subsequent exploration. That stage is called data preparation (cleaning) because it represents the process of identifying and removing,

Table 5.2. Computer Software Options for Analysis.

Software Category	Positives	Negatives	Sample Product Options
Statistical	Provides the latest and most advanced statistical techniques Highly complex and flexible programming language Capable of working with extremely large files and numerous variables	Difficult and complex to use with limited experience Will produce results, even if it is not appropriate to do so More than is needed for basic analyses	Statistical Package for the Social Sciences (SPSS) Statistical Analysis System (SAS) BMDP, LISREL
Database	Provides relational and multiple linkages among variables Highly complex and flexible programming language Works well with a variety of file formats and structures	Difficult and complex to use with limited experience Often limited to only basic level of analysis for specific set of conditions More than is needed for basic analyses	dBASE, Microsoft Access Paradox Foxpro
Spreadsheet	Provides basic and easily understood functions for simple analyses Moderately complex and flexible in options Moderately complex and flexible in options Likely to have the widest range and flexibility regarding graphics and charts	Very limited analysis Some restraints regarding number of variables and cases per worksheet	Lotus 1-2-3 Microsoft Excel, Quattro Pro

or at least correcting, the various types of problematic responses that often occur in any data collection process, including those collected electronically. No survey is immune. Many people who are only tangentially associated with survey efforts (including some of our clients) are aware that this process occurs. Nonetheless the raw data should be as clean and free of errors as possible before moving forward to the analysis and interpretation process. Researchers and practitioners often find themselves in difficult situations midway through an analysis, or occasionally even later in the communication and roll-out process, only to find that blanks (also known as nonresponses or missing cases) were being treated as real values in a set of computational averages. For example, many programs use a "9" to represent blank responses. When blanks are included, the mean scores for various items are artificially increased or skewed. Since these types of problems can affect the outcome of various analyses and therefore the nature of the survey story itself, the data must be examined and prepared thoroughly before proceeding with the next stage in the process. The survey items that are identified during the analysis phase as being the highest or most important to the organization should be due to actual perceptions or ratings made by respondents rather than because some chose *not* to answer the question and therefore skewed the results.

Several types of common problems or issues must be identified in the data preparation process. The main ones are listed next:

- Missing, incomplete, or partially completed responses
 - Blank optical scan form or questionnaire
 - First five questions answered; the rest left blank
 - Duplicate responses from same individual
- Problematic or intentional response patterns
 - All middle scores (for example, all 3s)
 - All extreme scores (for example, all 1s or 5s)
 - Design scores (for example, scores that make a graphic or range in clear sequences)

Responses to entire opscan form when the appropriate number is far fewer

Redundant or repeating scores

(for example, 1, 2, 3, 4, 5, 1, 2, 3, 4, 5, 1, 2, 3, 4, 5)

- Incorrect use of scales

Indicating a noninteger response on a scale with integer options (for example, 2.5)

Selecting more than one response when only one is requested

Negatively worded items

Inconsistencies in responses to similar items

- Damaged forms, computer malfunctions

Optical scan forms that are torn, faxed, or otherwise made unreadable to machines

Crashed disks, database systems that drop cases, glitches in writing cases

Each of these problems is discussed in greater detail later in this section, along with some suggested solutions. Although the identification of and ultimate solution to some concerns are relatively easy and straightforward, others are more subtle and may require considerable attention as well as some important decision making on the part of the survey practitioner or team. However, the problems that occur during this process are those due to individual respondents (for example, not following directions) and not the instrument itself (for example, badly constructed or unreliable items), or the method of administration (for example, software bugs or server accessibility problems). In the former the solution usually lies with the individual datastream itself; in the latter, the solution is more systemic; it concerns modification of all responses for a given item or section.

Missing, Incomplete, or Partially Completed Responses

The first and most easily identifiable type of data problem that must be corrected is that of missing, incomplete, or only partially completed

responses to a survey instrument. It is an unfortunate fact that employees often return survey forms that have not been completed; sometimes only a few items carry responses. This can also occur with on-line intranet, e-mail, and even voice-response based-systems, depending on how they are configured and whether or not individuals quit, exit, or otherwise become detached in mid-response. Because blank returns can be accidentally counted toward the total response rates, or worse, included in various computational and analysis work as described in the earlier example, it is best to identify and remove them from the dataset. These types of forms or datastreams are easily seen in either hard copy form or on a database system. Once identified, if they are found to be completely blank they should be discarded or removed from the database. These forms can and should, of course, be counted as a blank return in the overall response rate but not as a usable form. Partially completed responses should be retained and used for analysis purposes unless the number of completed items is less than 10 percent of the total set of questions provided (for example, less than ten completed out of a total of one hundred items on a survey). Less than 10 percent complete suggests a significant problem with that respondent and raises questions about the validity of those data. Of course, this means ensuring (through programming) that various electronic data-entry systems allow for incomplete responses. Some on-line surveys we have seen, for example, will not let users submit their responses to the server or central database until the entire form is complete. Although useful from an item-completion perspective, respondents can see this approach as frustrating and overly rigid, particularly if no means for saving answers part-way and returning to the form later have been provided.

Duplicate Responses from the Same Individual

Another potential problem in survey work is duplicate responses, particularly when multiple forms of response options are provided such as a pen-and-paper survey, voice response option, or a Web-based reply. Some administration systems have built-in safeguards to

prevent this problem from occurring (for example, providing individually coded questionnaires, key codes, or response forms per each employee and not allowing unidentifiable responses to be used). But these are less common because some respondents may think the confidentiality of responses would be compromised. Unless one of these types of systems is used, however, the dataset should be examined for duplicate responses, as these can artificially alter the mean score and response rate obtained. It is quite possible (and has happened in one dataset we have seen), for example, that one disgruntled employee might try to have his responses entered into the system ten or more times. This is particularly relevant if his manager is being rated on a section of the survey. A few duplications of this nature probably do not overly affect the survey findings, but they can be unfairly damaging to the specific manager or department receiving feedback.

In practice, duplicate or multiple responses are relatively rare. Clients often ask that this issue be examined, but usually only a small number appear to be duplicates. We have only seen significant cases of multiple responses in three or four organizations out of many surveyed. The likelihood of receiving exactly the same set of responses is low in surveys with a large number of items. In fact, exactly the same pattern of responses (whether they are duplicates or not) is uncommon in large surveys. Finding such duplications in any dataset depends on (1) the number of questions asked and (2) the number of responses obtained. More items on a questionnaire means a reduced probability of finding exact matches that are not duplications, but larger databases tend to offset this, with greater probabilities of duplications. Based on probability theory, with sixty items or more and at least ten demographics, exact matches that are *not* duplications are highly unlikely, regardless of the size of the dataset.

Identifying duplicate responses can be a difficult task, particularly in a large dataset with many items and many responses. The most efficient way to test for this problem is to use a sophisticated computer program to (1) sort the entire dataset in order of every variable, including the demographic responses, and then (2) using a

"lag function" test for an exact matching pattern of responses between each datastream and the next. When such a program is not available, the only other method is to sort the data using a basic word processing program and examine for duplicates by hand, which is not recommended. When an exact duplication is confirmed, however, and item number and sample size are large enough for this occurrence to be unlikely given probability theory, the multiple responses should be deleted. Keep in mind, however, that one does not want to ever discard data with the same pattern of responses if they are in fact from different individuals, so careful examination of this issue is required.

Problematic or Intentional Response Patterns

Another problem in survey work that is more common is related to intentional response problems. Examples include such patterns as providing all middle or extreme scores on a survey (for example, all 3s, 1s, or 5s on a 5-point scale) or drawing patterns on a response form instead of responding. The latter is only possible with an optical scan form; the former occurs in all types of survey administration methods and is probably the most pervasive response problem short of return rates. Hostility, fear, apathy, and anxiety are typically the causes of such problems. Because these data represent totally meaningless results at best or totally biased ones (positively or negatively) at worst, they need to be identified and removed before analyses and conclusions can be drawn with confidence.

In theory, the process of identification is simply a case of checking individual ratings for patterns that do not reflect opinions but some other type of response set. In practice, though, it is often difficult to spot these problems. It is simple to notice responses with all the same scale option but often much harder to find those with graphic patterns or ascending and descending patterns without looking through the individual questionnaires, optical scan forms, or data strings; this can be time consuming. Therefore, many of these problems do get past the practitioner. With any luck, their impact in larger datasets is minimal.

Once identified, the datastream should be removed. The problem, of course, is that the problematic pattern is often not as clear-cut as all responses using the same scale point. Instead, where the pattern may be questionable (for example, mostly 3s or 1s or 5s, but not all), the decision is less clear and is likely to depend on (1) total number of responses received and (2) similar types of ratings from others.

Incorrect Use of Scales

A related issue but one that is less likely to reflect an intentional attempt at influencing the results is that of using the scales provided incorrectly. Regardless of how many points are provided on a scale (3, 5, 7, even 9) some respondents will decide to select a point not included as an option, such as 4.5 instead of a 4 or a 5. This problem is worse with even than with odd scales (see Step Two for more detail on the difference between these options); however, the problem is persistent, regardless of the type of item. The other manifestation of the incorrect use of scales is when people choose more than one response option such as both 4 and 5 or both yes and no in response to a question. These responses are likely to occur more often in pen-and-paper, optical scan, and even simple e-mail replies than in voice response, disk-based, or on-line systems; those response choices are typically monitored and screened by a computer.

If such a problem is present, the best option in both instances is to use an average (or the midpoint chosen as in the first example) of the total set of responses so that the data can be retained for analysis. Many data systems do not allow noninteger responses; the record formatting may result in a missing entry for these responses, thereby making this option unavailable. In such situations, the practitioner can try to merge these data points into the database at a later date or, more commonly, ensure that they are entered into the system as blanks. It is not always desirable to discard data that appear to be legitimate; however, a scale was provided to accurately capture the perceptions of the respondent. If everyone failed to respond appropriately, the item itself would be suspect and likely discarded completely. But with only

a few midpoint responses, it is better to assume that the respondent could not appropriately make a discrimination on the existing scale, and their responses for those questions should be dropped (but not for that entire respondent). If the responses for a given item are truly categorical in nature (for example, "What is your primary work location?" or yes-or-no questions such as, "Do you have a personal computer at your desk?") the data must be discarded if they do not fit the response options provided.

Negatively Worded Items

Another type of data problem concerns negatively worded questions. Some survey professionals consider these to be useful for testing respondents' awareness and keeping them on their toes; other practitioners, such as ourselves, consider these items more problematic than beneficial (see Step Two). The issue here is that people often (1) fail to read the items carefully enough and therefore respond incorrectly or (2) do not understand the negative wording. Presumably, the latter concern has been corrected during a thorough pilot testing process, but sometimes bad items do slip through. In any case, always check negatively worded items to ensure that individuals responded appropriately and accurately.

Often the best way to check data is by comparing a mean score for the negatively worded item with a positively worded one with a similar content or theme. If the item did not produce ratings that parallel those of other items from the same individual or other types of abnormalities are inherent in the response set, it probably should be dropped from analysis. If a negatively worded item is to be used for analysis purposes, responses should be recoded so that they match the direction of the other ratings in the survey. For example, if 1 was positive and 5 negative for a particular item, but most others on the survey had 5 as positive and 1 as negative, then the first item should be reverse-scored. This way, the means will be more consistent. The only caution here is that this recoding process should be carefully noted in reports and communications using this item.

Inconsistencies in Responses to Similar Items

The final problem inherent in individual responses is response inconsistency. This problem is both difficult to identify and even more difficult to correct but should be mentioned nonetheless. Often a result of inattention or negativity toward the survey effort, two items with the same general content can receive divergent ratings. Conversely, a negatively worded item might not reflect the opposite of a positively worded one with a similar content focus, as described earlier (see Table 5.3 for an example). These problems are often difficult to identify on an individual level without complex programming using a statistical or database computer program (reliability is a method for testing the general tendency across the entire dataset). However, they often remain in the database despite the interests of the practitioner.

Damaged Forms and Computer Malfunctions

The problems of damaged forms and computer malfunctions are common occurrences that often need to be dealt with during data preparation. With optical scan systems, for example, forms can become easily damaged and otherwise unreadable when torn, faxed, or marked with pen instead of pencil (for those systems that read only pencil). Similarly, the reliability of optical character recognition

Table 5.3. Consistency in Responses

	Positive and Negative Questions	Responses
Consistent responses	a. To what extent is your manager effective at communicating with others?	5
	b. To what extent does your manager have difficulty communicating with others?	1
Inconsistent responses	a. To what extent is your manager effective at communicating with others?	5
	b. To what extent does your manager have difficulty communicating with others?	5

(OCR) used in many fax-back approaches is such that a significant portion of forms will come through with blank or incorrectly scanned information and text, which will require attention. These responses have to be either "re-bubbled" or entered into the data system by hand. In either case, there is a significant loss in both speed and accuracy for those responses. Similarly, frequent backups of active data collection methods such as voice response units or on-line Web systems are highly recommended because no hard copy of the scores is being obtained. A simple glitch in the computer's procedure could cause a permanent loss to the database.

Item-Level Analysis

Now that the data have been successfully cleaned and prepared, the real part of the analysis and interpretation process can begin. As with most complex processes, several levels of analysis are involved in examining survey data (see Figure 5.3). The first, most intuitive, and least statistically complex approach to data exploration is the item-level analysis. As the name suggests, at its most basic this is simply the process of examining the frequencies and descriptive statistics (for example, means, standard deviations, ranges, and percentage of nonresponses or "don't knows") for each item on the survey across all responses obtained. Table 5.4 provides an overview of the information that can be obtained from each of these simple calculations. Because this type of information can easily be generated using almost any type of computer program, including spreadsheets, statistical packages, or databases, it is common among practitioners without advanced degrees or experience in research procedures for entire analyses to be conducted using only these core methods. Even if more advanced analyses are intended, the data should first be examined from this perspective.

There are two objectives or intended outcomes for this type of procedure. First, because the average responses to each of the survey questions are the primary source for all subsequent analyses and

Table 5.4. Descriptive Statistics and the Information Provided.

Descriptive Statistic	Information
Frequency	The relative number (and percentage) of people responding to each scale option for a given item. Most useful for demographic and categorical questions.
Mean	The average response obtained for a given item across all respondents. Most useful for Likert-scale items (for example, extent present, satisfaction, agree-disagree). Often used to identify the top 5 or 10 highest-rated items (most positive), as well as the lowest-rated (5 or 10) items (most negative).
Standard deviation	The degree to which ratings of a given item ranged across all respondents. Very useful in determining if item "worked" or not across different sets of respondents. Also useful in determining the degree to which responses followed a normal distribution or were significantly skewed toward one end or another of the scale. Good indicator of whether or not the item will be helpful in more advanced analyses to determine complex relationships as well.
Range	The degree to which respondents rated the item using the full scale (all points provided) versus choosing middle-range scores. Lack of full use of scale in large samples may suggest a problem with a particular item.
Percent nonresponse, not applicable, or don't know	The relative number of people who did not or could not provide a response to a given item. Can be very useful as a data point itself (for example, percentage "don't know" about how the formal performance appraisal process works), as well as important in identifying items that may be problematic or not well understood by respondents. Often items with high percentages on nonresponse are indicated as such on a formal report, and then dropped from more complex analyses.

interpretation, items that did not work effectively or at all should be identified. Similar to the data preparation stage described earlier, the idea here is to identify any problematic items so that they can be corrected, put aside for the time being, or, in extreme situations, removed from the database. These are items that, for one reason or another, received significantly fewer responses in the form of "don't knows" or blanks than most others on the survey. However, these types of questions may still be useful and should be reported somewhere in the final presentation. When 15 percent or more of the responses were blank, however, these should be interpreted with some caution or flagged for more detailed examination. Similarly, if responses to negatively worded items appear to be in opposition to the general trend on positive questions with similar content, there may be a question as to their validity as well. Advanced statistics such as reliability analysis (whether the items move together with each other in patterns) and item response theory analysis (whether the items reach appropriate thresholds that suggest good response characteristics) are available, but the quickest and easiest tests are made using simpler measures.

Frequencies are often the most useful here. In Table 5.5, for example, it is clear that although the full range of the scale was used

Table 5.5. Identifying Response Patterns.

	1	2	3	4	5	DK
Q1. To what extent are promotions and assignments based on fair and objective assessment of people's skills?	3.2%	7.5%	37.5%	23.3%	8.0%	20.5%
Q2. To what extent do you feel adequately compensated for the work that you do?	25.1%	19.1%	32.0%	17.1%	6.6%	0.1%

on question 1, 20.5 percent of respondents did not feel they could answer the question regarding the fairness of promotions. Compare this with the responses obtained for question 2 regarding the perceived adequacy of the compensation plan. In this case, the nonresponse information on question 1 is actually usable and useful data since this suggests other possibilities.

Of course, in the results for this one item (question 1) alone, two entirely different interpretations can be made, which in turn have very different recommendations for action associated with them. It could be, for example, that many respondents are simply unaware of the organizational processes and criteria used for promotion because they are not well communicated either by the HR function or by their manager. It is also possible that this finding reflects a set of negative attitudes or experiences toward the formal promotion system.

Or to make matters more complex, perhaps the 20.5 percent are all originating from only one or two departments or functions and are not a general trend across all respondents. This is a good example of the complexities inherent in the survey interpretation process, showing it to be an art as well as a science. There are never definitive answers to be obtained from survey responses; these require face-to-face conversations with employees via interviews or focus groups (see Step Two). However, one of the best ways for a practitioner to determine which of these interpretations is more likely to be accurate is through examination of other, related items with similar content, the use of advanced statistics to determine complex relationships among groups of items, or through an analysis of the write-in comments, if these were collected.

The second objective in item-level analysis is to determine the highest- and lowest-rated items and to begin to look for emerging patterns or themes in the data. This involves a simple sorting and examining of all the relevant items on the survey from high to low. Some practitioners prefer to base their analysis work on categorical schemes such as "percent favorable" (the percent responding to the top two scale points on a 5-point scale such as 4 and 5), but the most

complete method that uses all the data obtained is to use the average rating for each item. Once all the items have been sorted from highest to lowest, it is a relatively simple task to examine the top ten or twenty and the corresponding lowest ten or twenty items for (1) specific item ratings and (2) any general patterns that might be evident. Table 5.6 provides a sample listing of this type of analysis.

Keep in mind that this process is a relative one, that is, it assumes that the highest- and lowest-rated scores obtained are indeed positive and negative findings, respectively. Although this is a workable assumption most of the time, it is not always true. For example, if the five highest-rated items on a survey using a 5-point scale are all less than 3.10, the survey team would be hard pressed to claim that these responses are *positive aspects of the organization* because the mean scores are so close to the midpoint of the scale. Rather, the interpretation would probably have to be that employees in the organization are generally unhappy or dissatisfied (or perhaps simply tough raters) and that these areas received the highest ratings relative to other questions. Conversely, if the lowest ratings received, using the same 5-point scale, are all above 3.10, this sug-

Table 5.6. Sample Item Sorting by Mean Score Obtained.

Item text	Mean	Ranking
1. To what extent does your manager stand up for what he or she believes?	4.21	1
2. To what extent does your manager take public responsibility for mistakes?	3.23	2
3. To what extent does your manager contribute actively and openly to the work of the team?	3.02	3
4. To what extent does your manager operate truthfully when delivering good or bad news	2.07	4
5. To what extent does your manager deal with problems objectively and fairly?	1.99	5
6. To what extent does your manager take appropriate action about poor performance from staff?	1.65	6

gests that employees are somewhat pleased overall and that the problem areas are probably not important issues for most people. Once again, although such patterns are uncommon, most large surveys using a 5-point scale do yield mean results with some items above a 4 and some below a 2, making interpretation difficult.

There is one caveat to this approach. If survey questions are to be examined in this manner, it is imperative that two conditions be met: (1) all negatively worded items must be recoded to match the scales of the positively worded ones, and (2) all items ranked together must use the same scales or, at the very least, scales with similar points to them.

Only compare items in this manner with like characteristics. There is no point, for example, in ranking means for items on a 7-point scale along with those using a 3-point scale. That does not make sense because the 3-point items will always yield lower means. Similarly, if a series of items has been rated using a unipolar extent scale, such as "the extent to which this behavior is practiced," it is probably not appropriate to rank these along with items rated using a bipolar agree-disagree scale, even if the number of points used in both scales is the same. In cases like these with significantly different scale lengths, points, or related issues (for example, ranked items or categorical responses), these questions should be examined independently of each other.

Besides the sorting process itself, a few rules of thumb are helpful when evaluating item-level mean scores in a survey. First, don't expect the average of the entire survey to be the midpoint of the scale provided. Just because you use a 5-point scale does not mean that the average should be about a 3, or should be a 5 with a 9-point scale. Most practitioners concentrate on the highest and lowest items, but if a midpoint cut-off value is needed, the best method for determining one is to compute an overall average rating across all survey items, assuming similar scales. This value is then used to calibrate the midpoint and assess the degree to which various items are generally positive or negative.

For the data presented in Table 5.6, for example, the total mean is 2.70. This indicates that any items falling below this can be considered anywhere from mildly to significantly negatively rated. Conversely, those above this rating are generally positive to very positive as far as the survey results themselves are concerned.

This calibration process will provide the most accurate method for assessing positive and negative values. A rule of thumb based on survey experience is that all items that receive mean ratings above a 4 on a 5-point scale or above a 6 on a 7-point scale are generally positive and strong findings. Similarly, most items with mean scores below a 2.0 or even a 2.5 at the low end, regardless of the scale, tend to represent significant problem areas. Whether or not any or all of these items become the focal points for the overall interpretation depends on other factors, including the extent to which more complex conceptual models, frameworks, or analysis are used to explore relationships. Once again, the idea is to craft a compelling story that captures the essence of the survey data and yet will motivate the client to action.

Conceptual-Level Analysis

Once the survey practitioner or survey team has a thorough understanding of the descriptive statistics and item-level means, the next stage in the data interpretation process is a conceptual-level analysis. Although this stage does not have to be overly complex nor does it require a working knowledge of advanced statistics, this is where the most advanced procedures and theoretical models are applied to provide a more complete and rich understanding of the relationships inherent in the survey data collected. This is what survey practitioners live for and where their expertise is best used. The goal is to use some type of overarching framework, model, theory, values, principles, or even a set of observed relationships among variables, that best fits or describes what is going on in the results. The task is to generate four or five main areas or themes out of the entire survey

dataset that can be used to explain or characterize the results and drive action planning. This is not always an easy or a simple task, but if done correctly it can make the survey results literally come alive for the client and the organization as a whole.

Two basic approaches to conducting a conceptual level analysis are (1) to use an existing model and (2) build one empirically from the data itself. These approaches are described next.

Using an Existing Model

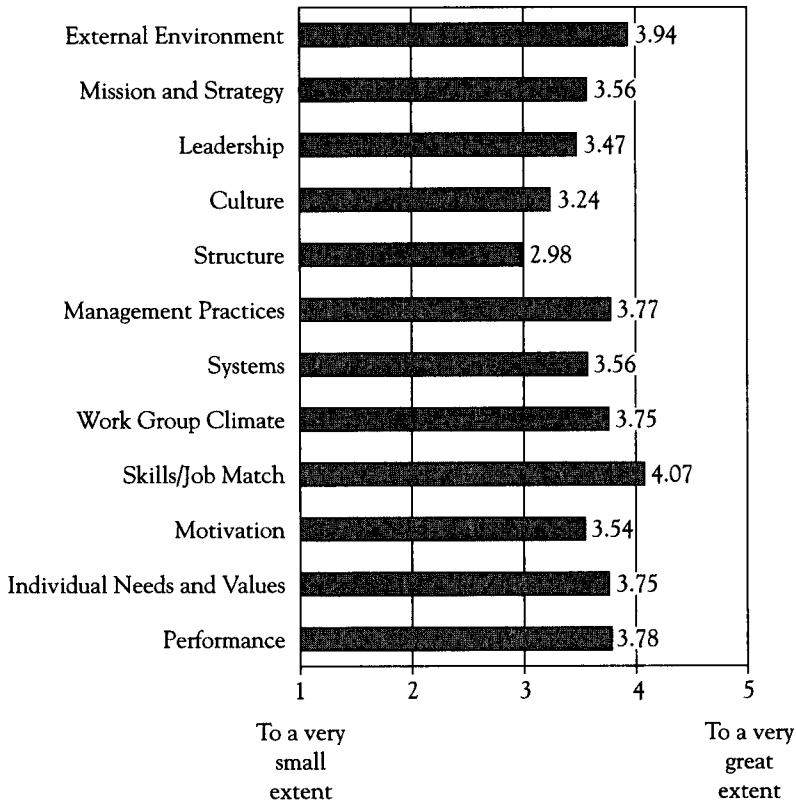
Probably the most common application of a conceptual analysis is to apply an existing framework to the interpretation and perhaps even the analysis plan for the data. Often this same model may have been used to guide the item development and construction process with respect to content areas (see Step Two). In any case the idea is to provide an overarching means of connecting different ideas and aspects of organizational life together in a comprehensive whole. It is for this reason, then, that in many instances the guiding framework is provided by a well-developed, theoretically sound, and widely used organizational model such as the Burke-Litwin model of organizational change and performance (Burke and Litwin, 1992), the Nadler-Tushman congruence model (Nadler and Tushman, 1992), or the Linkage Research model (Wiley, 1996), which has been used successfully at Sears (Rucci, Kim, and Quinn, 1998). Readers interested in learning more about either of these two models or other approaches to diagnostic work in organization development and change efforts should see Howard (1994) for a good summary. However, it matters less which perspective is chosen, as long as it provides enough latitude in its categories, descriptions, or proposed relationships among variables to be relevant to the issues at hand.

For this reason many survey efforts are designed and interpreted, not around a theoretical model but with regard to some set of important values, principles, or concepts that are specific to the organization itself. Once again, as we have mentioned several times before in earlier chapters, it is important to link the survey objectives, content,

and interpretation to relevant, important, and visible elements of the organization such as the mission statement, strategic initiatives, leadership principles, core competencies, or components of a new vision statement. This is one of the means by which survey results are given significance and therefore make an impact.

Such was the case at SmithKline Beecham (Burke and Jackson, 1991), for example, where an organizational survey effort was used during the initial stages of the merger to (1) assess main issues that needed to be dealt with during the coming years in the management of the merger process and (2) simultaneously communicate the new, desired culture and leadership practices deemed important for the future success of the organization. The Burke-Litwin model was used as a driving force behind the design and interpretation of the survey tool (Burke and Jackson, 1991), but in the end the results were reframed and communicated to employees using the new five core values and nine leadership principles. The use of more than one model or framework is not always easy to reconcile when examining the raw data and creating a story. However, it is often done when working with different professional groups and practitioners because most external survey consultants already have their own models well in hand.

When working with any type of preexisting model or framework, the standard and most basic approach is to use the primary variables, factors, or content areas provided in the model (for example, senior leadership, information systems, culture, motivation, reward systems) to classify and subsequently create averages or summary scores for each area. Questions reflecting various aspects of communication such as clarity, speed, and consistency of messages received might be placed in a category reflecting corporate communication systems. This way the information contained in the multitude of individual items can be condensed into a series of eight to twelve main categories that are all theoretically and conceptually similar in content. Figure 5.4 provides a sample of this type of conceptual analysis based on a comparison of summary scores.

Figure 5.4. Sample Conceptual-Level Analysis Using Summary Scores.

This type of conceptual approach (generating averages) is well within the means of a basic computer program. However, the added benefit of more advanced packages and statistics here is that each item in the category selected can be empirically tested using procedures such as reliability analysis, confirmatory factor analysis, multiple regressions, and even structural equation modeling. Regardless of whether the item was categorized at the design and construction stage before the survey was administered or after the data have all been collected and examined at the individual item level, these types of statistical procedures can help place individual items within main categories. Of course, if the conceptual analysis does not go beyond the categorization scheme, and the content averages such as those

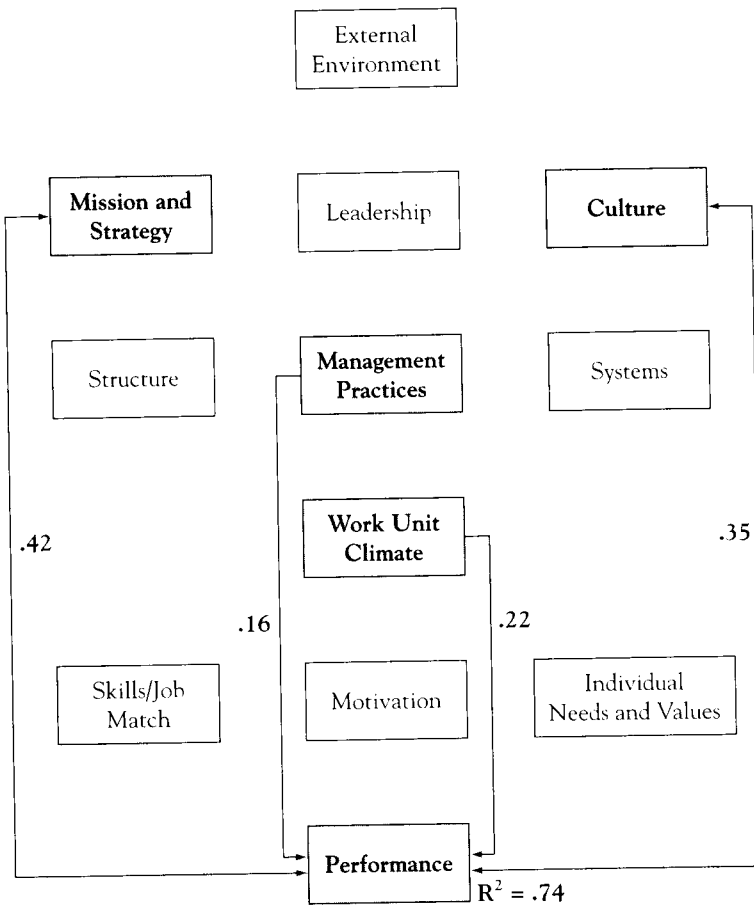
listed in Figure 5.4 are not used for personnel-related decision-making purposes, it matters less whether or not the item truly belongs in a given category. When content areas such as these are important or truly meaningful, as in our use of the Burke-Litwin model for guiding diagnosis and organizational change initiatives, it is important that the items be tested and moved or deleted if necessary so that model best fits the data at hand before proceeding further.

For those interested in going even further with this type of approach, the next logical application of a conceptual model is to test for relationships among the larger variables or categories described using inferential statistics. Although this can only be done using advanced statistical modeling techniques such as multiple regression or structural equation modeling (there is no simple substitute other than, perhaps, simple correlations, which are often difficult to interpret and can be misleading), the results of this type of advanced analysis are often worth the extra effort and can yield important insights into the underlying dynamics of a dataset. Without going into too much detail here, this process involves an analysis of the relationships among the variables described in the conceptual model or categorizing framework being used. By examining which types of variables are covaried with (or related to) others, a link can be drawn between changes or movements in one area and corresponding changes or movements in another. For those interested in more information regarding multivariate statistics in general, see Pedhauzer (1982) and Tabachnick and Fidell (1989). More specific information regarding the use of these methods in survey analysis can be found in Babbie (1973), Berk (1983), Schumacker and Lomax (1996), and Stolzenberg and Land (1982). For some applied examples of this approach to using surveys, see Burke, Coruzzi, and Church, 1996; Wacławski, 1996a; or Wiley, 1996.

Figure 5.5 provides an example of what the outcome of this type of more inferential approach might look like when using the same framework as in Figure 5.4.

Based on an examination of these results (and with some knowledge of the organization being surveyed), it is apparent that perfor-

Figure 5.5. Sample Conceptual-Level Analysis Using Advanced Modeling Techniques.



Source: Burke, Coruzzi, and Church, p. 55. In A. I. Kraut (ed.), *Organizational Surveys: Tools for Assessment and Change*. Copyright © 1996 by Jossey-Bass, Inc. Reprinted by permission of Jossey-Bass, Inc., a subsidiary of John Wiley & Sons, Inc.

mance in this financial services organization was being driven primarily by a combination of employees' understanding and awareness of the newly articulated mission and strategy, the rigidity inherent in the existing culture, the degree to which the day-to-day climate was generally positive or negative regarding the change effort, and the behaviors exhibited by middle management in support of the change

effort. Based on these findings and the strength of the relationships observed, some of the initiatives proposed included (1) enhancing and clarifying communications regarding the mission and strategy and (2) multisource feedback aimed at middle managers in order to change their immediate behaviors, positively enhance the climate experienced by employees and, over time, have an impact on the culture of the organization.

However, a word or two of caution is needed regarding the use of these types of modeling procedures. The results produced from this type of survey analysis do present a strong picture of the problems and potential solutions in the organization, but there is potential to cloud or obscure the more straightforward findings that are typically obtained. For example, significant problems may appear in a survey when people consistently provide low ratings and are clearly unhappy in their jobs, but this does not appear in this type of analysis because responses do not covary or move with other items. If everyone in the organization is in complete agreement that the management information systems are antiquated and lacking in speed and facility, that finding is likely to be missed if a modeling analysis is the primary means of the interpretation and subsequent action planning process; this variable will not be likely to covary with any others. This is one of the main reasons the standard item-level analysis approach is always recommended as the first step in any analysis plan, that is, so the survey team has a solid understanding first of the basic highs and lows from the individual ratings themselves before moving on to more complex exploration.

Creating a New Model

Besides the use of existing models or frameworks, the other main option for a conceptual-level survey analysis, albeit less often employed, is to generate a new model based on the relationships inherent in the dataset at hand. The survey responses themselves are examined using a number of different sophisticated statistical techniques (for example, factor analyses and structural equation modeling) in order

to create a conceptual framework or set of interrelated categories that best describe the data obtained. Although some practitioners, particularly the more traditional researchers, would argue that this approach is fundamentally flawed due to its tautological nature (arguing instead for a theoretically derived framework), in some situations the approach can be very useful. Occasionally, for example, the data, as analyzed according to an existing framework, may not make enough of a compelling story to motivate people to action. In other situations, it may be that certain individuals would rather use the data (rather than a formal model) as the motivating factor. In fact, for organizations with senior leaders who refuse to accept the notion that any preexisting model or theory can represent the issues in their organization, creating a new framework is often one of the few safe alternatives to chaos and ultimate rejection of the survey results.

There are other situations when creating a new model may be useful in conjunction with an analysis plan driven by an existing conceptual model or framework. Let us say you have conducted a survey with a series of items representing each of the twelve boxes or variables corresponding to the Burke-Litwin model. Now, in looking at the results, you realize that although you have a clearly defined conceptual set of items in the management practices category (ratings of specific behaviors of middle managers by their subordinates), over twenty-five different behaviors make up this content area, making a general interpretation of management practices difficult. You could conduct a basic item-level analysis by simply sorting the items in this section from most to least practiced. However, a more empirical approach would be to use a "principle components" factor analysis procedure to isolate four or five groupings of items that covary together statistically. In this application, the content of the items drives what the categories or factors are called. These factors could then be used to create management practices summary scores that are likely to make further interpretation and feedback more useful to others who work with these results. For most people, it is much easier to grasp only a few concepts, such as

communication style, rewarding and recognizing others, personifying leadership, managing the task, and empowering others than it is to work through a list of twenty-five items and associated responses. And remember, these data are for only one section out of possibly twelve on the survey, so it is important that it be examined and presented in a way that people can understand and ultimately use for positive change.

Comparative Analysis

An entirely different technique that is applicable and commonly used in large-scale survey work is comparative analysis. Simply put, comparative analysis is directed primarily at understanding a set of results from a survey effort in comparison to some other similar type of information collected either at the same time, or a different time, or even in a different organization. The information and subsequently the interpretations and recommendations obtained from these types of investigations are considerably divergent in focus from the types described in the analysis options earlier. Rather than specifying general relationships or patterns in the data as in the conceptual analysis, they are more likely to point to significant differences in perceptions among specific groups (for example, U.S. employees versus U.K. employees in a multinational corporation), across extended timeframes (for example, time 1 versus time 2) or across different organizations and industries (for example, the external benchmarking of best practices). These three examples represent quite well the three general types of comparative analysis that can be performed on survey data. Each is described in more detail next.

Comparisons by Type

The first and most intuitive form of comparative analysis is one that examines differences in survey ratings (using either items, conceptual variables and summary scores, or even a total average survey response) by various groups, functions, areas, experiences, demographic

information, or other categorizing variables that differentiate among types of people. For want of a better term, we have labeled these *comparisons by type*. The primary statistics used in these cases tend to be analyses of variance (ANOVA) because they focus on group differences. The most common comparative analyses of this type are done by background characteristics such as tenure, education, gender, or ethnicity to look for significant trends in responses. Other useful options might include questions that assess the respondent's participation or experience with a particular event such as a performance appraisal, various training courses, or even a prior survey effort. Table 5.7 provides some sample variables of this type and the corresponding types of interesting questions that might be answered with such an analysis.

A word of caution is needed here. Although the results obtained from these types of analysis can provide important information for

Table 5.7. Comparative Analyses Using Demographic Variables.

Variable	Question or Information of Interest
Location	Are there major differences among respondents from different regions or countries?
Age	Do younger employees feel differently about the company or their jobs than older ones?
Tenure	Do employees who have been with the company longer have more positive or negative attitudes?
Ethnicity	Do different groups experience the same level of treatment from managers?
Training	Do employees who have taken various training classes rate their knowledge of their jobs higher or have higher morale?
Gender	Do males and females differ in their perceptions of the fairness of reward and recognition policies?
Level	Do senior managers have different perceptions about the strengths and weaknesses of the organization than do middle managers or shop-floor service providers?

interpretation and action planning purposes, the survey practitioner should not go too far with these comparative tests. In most cases there are likely to be far more variables available for comparative investigations than are practical, given the need for a timely and concise analysis and interpretation of the results. Remember that the purpose of analyzing data in the first place is to condense the information collected into main trends and relationships, not to expand the amount of information provided. However, some practitioners have a tendency to err on the side of information overload by providing numerous iterations of variables cut by demographics and other categories to the point that the core messages in the data are lost to the client or survey sponsor. Although this type of detail is extremely important in the action planning stages (more on how this is generated and used in Steps Six and Seven), it can be cumbersome and ineffectual to provide too much information of this type too early in the process. In most cases, the only comparative results that should be retained and explored in detail at the analysis level are those that truly *reflect main issues* that need to be recognized and considered at the survey sponsor (typically corporate) level. In a merger of two organizations, for example, differences between employees from the two former organizations might be of extreme importance and relevance, particularly if there are significant and powerful differences in perceptions. It probably makes less sense to include this type of information, however, in an interpretation when few differences are evident or the issue is less relevant. Results such as these can be, and often are, added as an appendix to a formal report later in the process.

Comparisons Over Time

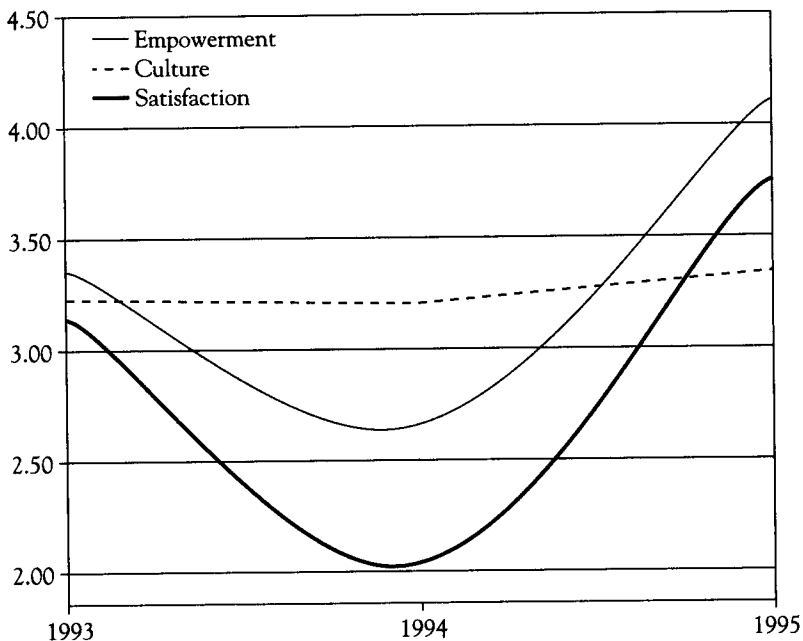
The second type of comparative analysis is achieved using additional data external to the current survey effort, that is, *comparisons over time*. Often the burning question here is, Have perceptions improved since last year? Most organizations with an institutionalized survey process, for example IBM or others in the Mayflower Group (John-

son, 1996) or the IT Survey Group, generate time-based comparative analyses as part of the standard analysis procedure to look for trends in employees' attitudes and perceptions. With these types of survey systems, senior management, HR, or the OD group can easily see whether the changes made between assessments yielded a positive or negative effect on employees' ratings. This is one reason organizations undergoing significant change efforts in their mission, vision, values, or structure implement an organizational survey effort to measure the impact and success of the change process over time (for example, Burke, Coruzzi, and Church, 1996; Wacławski, 1996b).

Figure 5.6 provides a good example of how one of these analyses might be used to explore changes in main content-related areas during the introduction and internalization of new empowerment-related management practices.

For this example let us assume that these new practices were intended to have a significant impact on the extent to which employ-

Figure 5.6. Sample Comparative Analysis Using Ratings Over Time.



ees felt empowered in their jobs, as well as the extent to which the organizational culture was supportive of innovation in the workplace. In examining the data, it is clear from this analysis, even though only three conceptual variables are being used, that (1) perceptions of empowerment and employee satisfaction seem to be highly related, and (2) although these were negative when the new practices were first introduced (1994), employees appear to be feeling much more positive about both aspects in 1995, at which time the new concepts have taken hold. In comparison, it is interesting that the culture of the organization does not seem to have been affected one way or the other during this change process. Ratings in this area were surprisingly stable over the course of two years. Because significant changes in organizational culture often take more than a year or two, however, this lack of effect is not particularly surprising.

If it is not already apparent, the negative aspect to this approach is the availability of comparative data from prior assessments. This type of analysis is only available when more than one survey has been conducted over time and all or some of the same assessment items can be used. This is one reason a thorough item-construction process should be followed during the first survey administration; at least a majority of the survey questions will yield meaningful and important results now and in the future if used again. Between changing employees, roles, and corporate agenda, it is often difficult to retain the same survey design or process during subsequent administrations unless the system has become fully institutionalized. The energy and excitement that was behind the first survey effort has often been spent on resulting initiatives or other, competing efforts by the time the second or third survey would be valuable. We have seen many organizations make effective use of a survey during a change effort, for example, only to decide a year later that a follow-up survey is either (1) too costly to find out that "not enough" has changed, or (2) too burdensome to respond to on the part of employees or management, in view of all the other initiatives in progress. Of course, this highlights the need for the survey practitioner to be vigilant in try-

ing to instill the importance of repeated assessments for tracking change up-front, at the initial contracting stages. Doing so after the fact can be challenging.

External Comparisons

The third and final type of comparative analysis for survey results concerns the use of *external comparisons* or information, also known as benchmarking data, to provide a reference point for the ratings obtained. As we discussed in the Introduction, benchmarking is a process of comparing the results obtained from an internal survey with a predetermined external measure or benchmark to determine the relative standing of the organization's functions and operations. These external benchmarks include the organization's top competitors or perhaps the best competitors in the same industry. Sometimes companies in entirely different industries but with similar types of processes or issues facing them are used for comparison purposes. Regardless of the nature of the benchmarking database, the purpose of this type of analysis is to determine how well the current organization is doing in various areas relative to other firms that have collected the same or very similar information. At its most specific form, benchmarking has been used for documenting specific processes and procedures for enhancing service and product quality (for example, Camp, 1995; Spendolini, 1992). In other situations, such as survey assessment efforts, the focus tends to be more on a comparison of issues like satisfaction with compensation, benefits options offered by competitors, the level of training and education required for various jobs, and the extent to which managers engage in specific types of behaviors (for example, Johnson, 1996). In fact, benchmarking has become so popular that some survey practitioners and survey-based consulting firms have promoted the use of their benchmarking dataset rather than allow for a custom-designed survey as their primary competitive advantage. For some, a high degree of external comparability more than makes up for a lack of specificity and adaptability in the choice of content or questions.

Some degree of external benchmarking is probably useful most of the time at a general level, particularly given the increasing tendency for organizations to span national boundaries and compete in different markets; however, some practitioners, ourselves included, are concerned with an apparent overemphasis on this type of analysis. By itself, benchmarking organizational assessment survey data is likely to tell the analysis team very little and provides for a rather weak set of interpretative statements. It may add some flavor and background to the results obtained, but it cannot and probably should not be used as the core set of analyses for any large-scale organizational survey effort. What is going on in one's own organization should be far more important and meaningful than how the data compare favorably or unfavorably with other companies at a more general level. We have seen far too many clients, for example, debate the merits of an external comparison or discredit some low survey rating obtained in their own company, such as a mean rating on a 5-point scale of 1.96 for employee morale, because it was equally low or lower in other organizations included in the benchmark. From our perspective, if an organization has taken the time and money to conduct a survey properly, it should use the data gathered from that process and the relative strengths and weaknesses as described earlier in this step as the basis for its analyses, interpretations, recommendations, and action plans.

One final point is needed regarding the use of benchmarking. As we stated in the Introduction, people often think of the word *benchmarking* in terms of external indicators; however, internal survey ratings can be used as benchmarks too. Practitioners often refer to the results of an initial survey effort, for example, as being the benchmark or baseline for the future. Data from high-performing work teams, departments, functions, divisions, business units, or entire countries can be used as internal benchmarks as well. Of course, in the nomenclature used in this step, these would probably be classified as *comparisons by type* or *comparisons by time*, but the idea across

all three applications is the same: to compare one set of results to another to look for strengths and weaknesses and enhance the overall survey analysis picture.

Content Analysis of Write-In Comments

The last stage in the survey analysis process concerns an entirely different type of data altogether—the written comments provided to various short-answer or open-ended questions such as, What is exciting about your job? or What makes you come to work each day? As discussed in Step Two, large-scale surveys often include one or more of these qualitative questions, in conjunction with a series of quantitative ratings, in order to provide greater depth and clarity during the analysis process. Unlike the myriad of available statistical options when working with raw numerical responses, the analysis of write-in comments is relatively simple and direct.

The process most frequently employed here is called *content analysis*, which basically means taking the individual, idiosyncratic comments and opinions and converting them into a number of main categories or themes for interpretation purposes. If this procedure sounds familiar, it is because this is exactly the same process that is used by many OD, I-O, and HRD professionals to analyze responses from focus groups, interviews, and other types of qualitative data collection. We have already introduced the use of this technique in Step Two with respect to generating survey item content and providing feedback based on information gathering during meetings with organizational members.

Content analysis is a relatively straightforward process and requires no special skill other than the ability to conceptualize and sort responses based on a number of categories or themes. Here are some sample write-in comments taken from several different large-scale survey efforts in response to the question, What are the blocks, hindrances, barriers that prevent you from doing your work effectively?

Change is happening so fast and there is too much of it. So many managers are inexperienced in their new roles when I need information.

Cronyism and sexism dictate advancement far more than do ability and performance. This is still a white, male-dominated institution.

It's quite appalling that this organization still condones a high-handed and humiliating managerial style—cronyism, unfortunately, is not always kind toward people of color.

The bureaucraticness. I'm filling out this survey under lights that blew out over three weeks ago and though reported have still not been fixed.

Lack of any forward planning; no business plan seems to last more than six months. No investment is planned until equipment is worn out or sudden events require it, for example, new safety legislation.

Customer service should be more proactive and less reactive; this company seems to manage with a fire drill mentality.

Departments that do not like or trust one another mainly due to upbringing, education, or class.

Comments and criticism are discouraged in this company and are often held against people who make them. We have a long way to go, and a few "feel good" meetings and employee surveys are not going to mean as much as positive and honest action.

Management is blatant about favoritism, not allowing for others to learn or advance to the same level of growth or promotion.

A very sad lack of management that we can respect. We are appalled at the way we are treated as staff and at the seeming lack of understanding of our feelings of helplessness as we see the erosion of our standards and ways of treating each other.

Staff attitude of "If you don't want the job, there are many others who do."

Morale went out of the window months ago, and yet we are issued a mission statement that probably cost quite a bit of money and would be just as appropriately written on the side of an ice-cream truck.

Frequent management U-turns about our role and responsibilities. Lack of understanding by managers of what we actually have to do to carry out our jobs effectively.

Due to cuts, closures, and reorganizations I no longer have an experienced or accessible senior manager I can turn to.

During the merger senior management lied to employees and were not concerned about our jobs. They did not help those of us who lost our jobs and had to find other employment. They were only concerned about themselves and were not the least bit sensitive toward the rest of us.

The most frustrating thing is our lack of reliable equipment. Our PCs are always giving us problems, the printers are always breaking, we are spending unnecessary hours making up work we lost because of bad equipment. Please give us good equipment so we can do our jobs more efficiently and be happier coming to work!

Changes are made overnight without advising us and without any training provided, yet we are expected to bumble along without a glitch. When our departments were merged no effort was made to bring people together and make us all feel comfortable.

Marriage leave taken away; no Christmas bonus; benefits are terrible, job security a major problem. Always worried about losing my job—can't sleep at night!

I have an enormous additional workload with reduced support. Experience, expertise, and goodwill count for nothing anymore.

The only negative aspect of the content analysis process is that it often requires a significant amount of time and effort to transcribe or input the comments into a computer form for manipulation (cutting and pasting, tabulating, sorting, and so forth), particularly with optical scan or pen-and-paper methods. Some data collection systems can record the information directly in electronic form (for example, on-line Web sites, e-mail response methods, or voice response units), but these methods often have other difficulties associated with their use, including reduced levels of formal written comments, text length and format limits, respondent unease regarding confidentiality, and occasional problems in transferring the data obtained through these methods to other computer programs. Moreover, although sophisticated text-based content coding software applications exist in the marketplace, the output from these programs is still inferior to the output of a "live" analysis approach. In any case, regardless of the method of administration used, the analysis of quantitative data is laborious and time consuming. The richness of the information obtained, however, often more than makes up for the extra effort involved, which is one reason these types of questions are so popular in survey efforts.

Before we get to describing the specific process by which to conduct a content analysis, we want to underscore two decisions regarding the use of write-ins, since the outcomes of these decisions have important implications for how the data are analyzed and subsequently interpreted.

First and foremost is the question of whether to use a sample or a census of the comments obtained. Should the content analysis be based on all of the individual write-ins obtained, or should it be completed using only a random or perhaps stratified random sample? One's initial reaction might be to use all of the data available; however, given the tight time constraints involved in the survey analysis process and the voluminous amount of information gathered with

such methods (sometimes several pages of text per individual, particularly if there are multiple questions or large numbers of respondents to the survey), more often than not the decision is to use a representative sample of the comments. This is not a problem because (1) an appropriate sample set of comments is likely to be rich enough to provide meaningful and insightful information, and (2) the entire set of responses can always be coded later, after the results have been communicated at the first or second level. Thus, the general rule of thumb here is a 20 percent sample of the total responses. In other words for a survey with two thousand returns, you would use four hundred write-ins for the content analysis. The only concern besides size when choosing a sample is the extent to which various business units, functions, or countries are adequately represented. In some situations, for example, it might be desirable to ensure that responses from different locations are appropriately represented, particularly if there may be differences evident in other areas.

The second decision in the content analysis process is whether or not the results obtained will be fully integrated with the other, more quantitative analyses produced. If integration of these different sources of information is key to the final interpretation and presentation, which is highly recommended but not always possible, the content coding categories of themes used should match as closely as possible the framework or variables used for the conceptual and comparative analyses results. The write-ins should be classified and selected in such a way as to provide additional support and elucidation regarding the key issues and themes identified in the numerical responses. However, in situations where the write-in questions were asked regarding specific topics, there is not adequate time to integrate the findings into a comprehensive whole, or certain individuals are wary of using write-ins because of their qualitative nature, then the content coding categories should reflect the actual themes inherent in the comments themselves, whatever they may be. For example, in the first situation the comments might be coded according to the twelve boxes in the Burke-Litwin model (for example, leadership, culture, mission,

motivation, and systems), whereas in the second case the themes might be more specific such as job security, upgrade technology, increase staff levels, reduce paperwork, or get better managers. The choice here is usually based on a combination of time, resources, types of questions asked, as well as the expectations of the client regarding results.

With these decisions made, the following stages describe the content analysis process itself as it is usually performed:

- Identify the total comment pool available for analysis (sample or census).
- Examine the responses for general tone, potential content themes (assuming main classification categories have not already been identified from a conceptual model or prior analyses), and good representative sample quotes for use in reporting (depending on level of quantitative and qualitative data integration needed).
- Classify all write-in responses from the pool into content categories or themes. It is not considered good practice to classify the same comment into different categories concurrently because it gets listed and counted twice this way. However, it is acceptable to break up a given individual's comments into separate parts or points and classify these in different themes accordingly.
- Refine and adjust the content themes (modification of categories or addition of subcategories—which can be difficult at the software application level) as needed.
- Tabulate the number of responses per category or theme into frequencies once coding is complete so that the relative importance of each issue can be determined.

Table 5.8. Sample Content Analysis Results.*What is exciting about your job? What makes you come to work each day?*

	Number of Responses	Percentage of Total
Successfully helping customers	110	31
Challenging work	76	21
Teamwork among colleagues	60	17
Variety and diversity of work	43	12
Opportunity to contribute	25	7
Nothing but the pay check	24	7
Developing new skills	16	5
Total	354	

Table 5.8 provides an example of what the result of this entire process might look like, based on a sample of 354 responses in a larger survey effort.

Checklist for Step Five

1. Determine the role of statistics.
 - Use those in which you are skilled, and use an expert for those in which you are not.
 - Carefully consider using tests of significance.
2. Remember that timing is very important.
 - Prepare the following in advance (analysis and interpretation are usually allotted the *shortest* amount of time of all seven steps):
 1. Make sure that the data entry system is fully operational and has been tested.
 2. Install and test all data analysis software.
 3. Write and test initial analysis code using sample data.

4. Obtain results from previous surveys for benchmarking purposes.
 5. Manage the clients' expectations and anxiety regarding issues with the data collected.
 6. Prepare a comprehensive and thoughtful data analysis plan.
 7. Develop a framework or outline of the final survey report.
3. Remember the six stages for analysis and interpretation:
- *Data entry.* Formally enter results into some sort of database.
 - *Data preparation.* Check for the following: missing, incomplete, or partially completed responses; duplicate responses from same individual; problematic or intentional response patterns; incorrect use of scales; negatively worded items; inconsistencies in responses to similar items; damaged forms; computer or software malfunctions.
 - *Item-level analysis.*
 - *Conceptual-level analysis.*
 - *Comparative analysis.*
 - *Content analysis of write-in comments.*