Chapter 8

The Identification and Analysis of Outliers in Employee Survey Data as a Means for Improving the Reporting of Survey Results

Anthony Dillon
Centre for Positive Psychology and Education, School of Education, University of Western Sydney

Jeff Patrick
Australian Institute of Health Service Management, Faculty of Business, University of Tasmania

Krosnick (1999) observes that the literature is replete with new insights that demand dramatic revisions in the conventional wisdom of survey methodology. Such dramatic revisions are necessary to draw out the deeper relationships inherent in employee data sets. With this in mind, this chapter aims to explore novel methods of analysing employee survey data. It is intended that these methods will assist management in deriving a clearer and more complete picture of what the information contained in the survey is telling them about their organisation. The proposed novel methods of analysing and reporting employee survey data are not meant to replace the existing more conventional methods, but rather are meant to supplement these methods.

Analysing Survey Data

When conducting employee surveys, managers are usually more interested in the major themes or trends that emerge when the data are analysed collectively, than in individual responses (Church & Waclawski, 1998). They typically want to know what the average or most common response is for each item or subscale. For some survey datasets an elementary statistical analysis (that uses means) is adequate for reporting results. Managers can see at a glance if there is a problem within the organisation, or where they have succeeded. Such an approach to reporting results is understandable in a climate of increased accountability where managers expect their organisations to be bottom-line oriented (Fralicx et al., 2002). While such analyses are useful for identifying the main findings in a dataset, there is still the less obvious, yet useful, information waiting to be found. According to Walters (1990), it is the ability to go beyond the obvious factors to identify further issues that distinguishes the most successful surveys from others. If this useful information is not identified due to a substandard analysis, then the utilisation of the findings will also be compromised.

Limitations of Means and Percentage Scores

The data collected from organisational surveys, in part, reflects the nature of the organisation's structure, which is not always simple. Large organisations are often multilevel in nature, and therefore data collected from a survey conducted in such organisations will be multilevel in nature. That is, the data are hierarchically structured or nested (Chan, 2006). For example, within an organisation, employees may belong to various units defined by location or function, where the units combine to form different departments. Within a unit, the staff may have different ranks and job classifications. According to Chan, the single-level approach to multilevel data, which assumes a "one-size-fits-all" solution, is inadequate. Consequently, attempting to apply averages and percentages to overlapping organisational units can be misleading. For this situation, while means and percentage scores can be calculated across the organisation as a whole (single-level approach) to provide some useful information at the aggregated level, these statistics are limited in their capacity to provide or uncover the full depth of information that is available within the dataset – information that is likely to be valuable to managers.

An example at this stage may be useful. Assume that employees are surveyed with regards to their levels of satisfaction with their employer. Levels of attitude could be measured using either a single item, or from a subscale. One could expect that levels of satisfaction could range from being very dissatisfied to very satisfied. Neither of these responses would necessarily be considered extreme or unusual. Suppose also that only 3% of respondents expressed that they were either dissatisfied or very dissatisfied. Further, suppose that this percentage score is down from 12% in the previous year. Use of the elementary percentage score has indicated a relative degree of success for the organisation as a whole. However, if a disproportionately higher number of those employees who expressed that they were either dissatisfied or very dissatisfied came from one work unit within the organisation, then the apparent success would clearly not be universal. What may mask this relative concentration of discontent still further, is that the individual work unit in question may have had an acceptable average level of satisfaction among its staff, simply because there was a sufficiently large number of staff within that unit who indicated that they were very satisfied. Findings such as these would not be detected through the routine use of elementary statistical procedures. While this single-level approach may be adequate for describing individual cases or specific subsets of cases (such as individual work units) when they are representative of the organisation as a whole, it is inadequate where diversity exists throughout the organisation (Rousseau, 1985).

An obvious solution to this problem therefore, might be to conduct analyses that recognise the multi-level nature of organisations (Hofmann, 2002). Analyses could be conducted where the focus is on smaller subsets of data that enable employee responses to be evaluated for the various groupings that exist in the organisation, such as work units, age groups, geographical location, etc. As a simple example, an analysis could be conducted to determine if some work units have a higher number of employees who are dissatisfied with their work life than would be otherwise expected from chance alone. Such an approach is conceptually straightforward and only requires that the variables of interest be identified (e.g., work unit and employee satisfaction), and assessing the level of satisfaction for each work unit by using either means or percentage scores.

The procedure just described, whilst thorough, is often tedious, particularly for datasets where some of the variables may contain many categories or levels. Further, where results for individual work units reflect the general findings of the organisation, the value of such analysis can also be drawn into question. This situation may lead to a more ad hoc approach to examining the results of individual work units, driven by the inquisitiveness of individual managers. Clearly, this later approach would be less than thorough.

A better approach is to adopt a strategy that specifically seeks out those cases that are not typical or representative of the data as a whole – variables that are easily masked or overlooked when adopting the single-level approach. Such cases are referred to as outliers. It is expected that a strategy that focuses on identifying outliers will prove useful in augmenting the statistical techniques most commonly used (summary statistics in the form of means and percentage scores) when analysing survey data.

Outliers

The topic of outliers in datasets is covered at length in the literature (e.g. Barnett & Lewis, 1994; Hawkins, 1980; Lovie, 1986). Although there is no precise definition of an outlier (de Vaus, 2002a), Rousseeuw and van Zomeren (1990) describe outliers as those observations that do not follow the pattern of the majority of data. Despite their importance, Osborne and Overbay (2004) report that a casual observation of the literature suggests that most studies rarely report checking for outliers of any sort. This is an important observation, as many papers and studies have shown that datasets free of outliers are the exception (Filzmoser, 2005). This is particularly relevant to staff survey data, where failure to detect outliers during the analysis phase of the survey process may mean that management is potentially overlooking valuable information (as demonstrated by earlier hypothetical example).

Different Types of Outliers

Depending on the purpose of the analysis and the nature of the data, it is convenient to categorise outliers into one of two types. The first of these is the univariate outlier, which is a case with an extreme value on one variable (Tabachnick & Fidell, 2001). The second type, the multivariate outlier, represents a combination of scores on two or more variables that is different from the majority of the data (Tabachnick & Fidell). This chapter will examine univariate outliers in employee survey datasets.

The univariate outlier is easier to conceptualise as much of people's everyday experience is easily explained or represented by one dimension. Typically, univariate outliers are those cases whose numerical score lies in the tails of a probability or frequency distribution. Furthermore, not only can univariate outliers be defined by their distance from the majority of data within a given dataset, but also their distance is characterised by direction from the majority of data. This feature makes univariate outliers easy to interpret. For example, univariate outliers can be either above or below the majority of the data. For a given scale, if a score is considered desirable or positive the further it is above the central (or mean) value, then a score far from the central value in the opposite direction is considered undesirable or negative.

Theoretical Considerations for Dealing with Survey Data

While the focus of this study is not concerned with questionnaire construction, it is important to provide a general understanding of the theory of scaling, particularly as it applies to employee surveys. Such an understanding will be helpful in appreciating some of the limitations or concerns when undertaking statistical analyses with employee survey data.

Many of the items in employee surveys use a Likert scale. While this scale is straightforward and simple to apply (McIver & Carmines, 1981), it does have limitations (van Alphen, Halfens, Hasman, & Imbos, 1994). One such limitation is that the Likert scale is essentially an ordinal scale. Many authors (e.g., Cliff, 1996; Mertens, 2005; Townsend & Ashby, 1984) suggest that some descriptive and inferential statistical analyses differ for ordinal and interval data. While these same researchers would therefore recommend the use of non-parametric methods when dealing with ordinal data, this recommendation is not supported by all researchers (e.g., Gregoire & Driver, 1987; Kim, 1975). Kim suggests that if the use of parametric procedures is desired when dealing with ordinal data, ordinal strategies are no better than parametric strategies at meeting the basic requirements of multivariate analysis when pursuing the goals of scientific research, namely successive estimation and refinement.

Wright (2003) suggests that adequate justification should be provided for the choice of a statistical procedure for a given measurement scale. It is acknowledged that the variables used in the dataset of this study are measured on ordinal scales, and therefore can never be truly distributed normally. A simple reason for this, is that the normal distribution has infinite tails (Thomas, 1982), and as Likert scales have absolute minimum and maximum values, the range of allowed scores is strictly bounded. However, if it can be demonstrated that the distribution of scores for the subscales to be examined are approximately normal or at least have a reasonable degree of symmetry, then the use of some parametric statistics such as the mean and the standard deviation will be justified. Otherwise, non-parametric measures such as percentile values and the median should be used as required.

Outliers in Survey Datasets

According to Lance, Stewart and Carretta (1996), the detection and treatment of outliers has received scant attention in the human resource management literature. The obvious absence of literature that brings together research on staff surveys with the concept of statistical outliers argues well for the need of this present study.

While the identification of individual outliers potentially provides useful information to management, the nature of their distribution throughout the dataset can provide further useful information for managers (de Vaus, 2002b). The purpose of this study is to ascertain whether the outliers contained in an exemplar dataset are randomly (and hence naturally) distributed throughout the organisation, or whether they cluster together, and by their distributional pattern point to otherwise hidden issues within particular sections of the organisation.

Consider the example discussed earlier. It was described how 3% of respondents expressed dissatisfaction with their workplace. In any sufficiently large workplace, a portion of workers who are dissatisfied with their workplace is to be expected. However, if these dissatisfied workers have something in common with each other in addition to their dissatisfaction, for example belonging to the same age group, the same department, or having the same job role, then this finding can tell much about the organisation. That is, the level of dissatisfaction is most likely to be a reflection of some aspect of the organisation, than the individuals themselves. These sorts of findings are not easily uncovered using the routine statistical analyses typically used when reporting on employee survey datasets, and therefore make this study such a worthwhile and important pursuit.

Method

To test the utility of analysing outliers when reporting employee survey data, an archival survey dataset was used. The nature of the particular dataset and the particular findings are less important than demonstrating and evaluating the analytic techniques involved. The dataset used here, though coming from an actual employee survey, is only a vehicle with which to test an alternative statistical procedure for analysing employee survey data. The archival dataset serves as an exemplar dataset and contains data collected from the Queensland Public Agency Staff Survey (QPASS; Hart, Griffin, Wearing, & Cooper, 1996) for 2002. The survey was designed to collect employees' opinions regarding their work environment and job satisfaction. The questionnaire comprised 10 demographic items, 6 items relating to respondents' opinions about the survey, and 76 items specific to organisational climate and job satisfaction. Participation in the survey, though strongly encouraged, was not compulsory. The questionnaire was administered in paper-and-pencil form. A page of instructions was provided to participants that included the usual assurances of confidentiality and the availability of feedback on aggregate results.

Demographics

In total, the dataset contains 1,959 records for a large Queensland Public Service Department that comprised four districts (D1, D2, D3, D4). Table 1 provides a summary of participants by district and age group. Each of the variables used in this study (including the subscale scores derived from individual items) had less than 5% of their data missing. Furthermore, an analysis of missing values reveals the distribution of missing values to be random.

Table 1 Summary of QPASS Respondents by Age and District, 2002

	District			
Age group	D1	D2	D3	D4
< 30 years	22	214	58	44
31 to 40 years	43	279	101	104
41 to 50 years	59	327	131	120
> 50 years	29	208	100	68

Note. Rows and columns will not add to the total of 1,959, due to missing data.

Survey Variables

Two sets of variables were used in the analyses. The first set relates to employee demographics/characteristics and includes: age, gender, time in current position, time in organisation, and district. Descriptions of these variables (hereafter referred to as employee variables) are provided in Table 2. The second set of variables includes the survey subscales described in the QPASS manual (Hart et al., 1996) and is shown in Table 3. This table also includes sample items of each subscale. Subscales were analysed in this study instead of individual survey items, for reasons that will be explained shortly.

Table 2 Definitions of the Employee Variables Used for Analyses

Variable	*Description*	*Values*
Age	*Employees' age at the time of survey.*	*< 30 years, 31 to 40 years, 41 to 50 years, > 50 years*
Gender	*Employees' gender.*	*male, female*
District	*Districts within the organisation.*	*D1, D2, D3, D4*
Time in current Position	*How long the employee has worked in their current position.*	*< 1 year, 1 to 2 years, 3 to 5 years, > 5 years*
Time in Organisation	*How long the employee has worked for the organisation.*	*< 1 year, 1 to 2 years, 3 to 5 years, > 5 years*

Note. Originally this variable included age group categories of < 21 years and > 60 years. These categories were recoded as < 30 years and > 50 years respectively to enable the chi-square test to be applied.

Table 3 Description of the QPASS Subscales Used in the Analyses

Subscale	Definition	Sample Items
Quality of Work Life	Conditions of life at work are excellent, giving everything important that might be wanted.	"In most ways my work life is close to my ideal."
Individual Morale	Feeling positive, enthusiastic, proud, cheerful, and energized at work.	"Feeling positive at work."
Psychological Distress	Feeling tense, afraid, unhappy, anxious, negative, uneasy, and depressed at work.	"Feeling tense at work."
Organisational Climate		
Workplace Morale	Staff show enthusiasm, pride in their work, team spirit and energy.	"There is a good team spirit in this work area."
Workplace Distress	Staff feel frustrated, stressed, tense, and anxious and depressed about their work.	"Staff in this work area experience a lot of stress."
Supportive Leadership	Managers are approachable, dependable, supportive, and know the problems faced by staff, and communicate well with them.	"There is support from the supervisors in this work area."
Participative Decision-Making	Staff are asked to participate in decisions, and given opportunities to express their views.	"I am happy with the decision-making processes used in this work place."
Role Clarity	Expectations, work objectives, responsibilities, and authority are clearly defined.	"I am always clear about what others expect of me."
Professional Interaction	Acceptance and support from others, with involvement, sharing, good communication and help when needed.	"Staff frequently discuss and share ideas with each other about how best to carry out their work."
Appraisal and Recognition	Quality and regular recognition and feedback on work performance.	"I am encouraged in my work by praise, thanks or other recognition."
Professional Growth	Interest, encouragement, opportunity for training, career development and professional growth.	"I am encouraged to pursue further professional development."
Goal Congruence	Personal goals are in agreement with workplace goals which are clearly stated and easily understood.	"The staff are committed to the workplace's goals and values."
Excessive Work Demands	Staff are overloaded with constant pressure to keep working, leaving no time to relax.	"There is too much expected of staff in this workplace."

Note. From Manual for the QPASS Survey, by P. M. Hart, M. Griffin, A. J. Wearing, and C. L. Cooper, 1996, Brisbane: Organisational Climate and Performance Project, Office of the Public Service.

Alternative Procedures for Analysing Outliers

The statistical procedure used in this study comprises two stages. The first stage involves the identification of outliers in the survey dataset, while the second stage examines the distribution of outliers in relation to selected employee variables. Specifically, the analysis stage involves checking if the outliers identified for the subscales are randomly distributed throughout the organisation (for a specified employee variable such as district, age group, gender etc.), or if there is some pattern to their distribution. The results yielded from these procedures will be contrasted with the results yielded from the standard techniques typically used in the analysis of employee survey datasets.

Stage 1 - Identifying Outliers

Identifying univariate outliers is a relatively straightforward process. However, before they can be identified, careful consideration needs to be given to how they will be defined for a particular dataset. The definition used to define an outlier in this study is dependent upon the nature of the distribution of responses. Typically, for variables distributed normally, outliers are defined in terms of a specified distance from the mean. For employee survey data it is useful to define outliers as those cases that lie 1.64 standard deviations or more either side of the mean. The values of ±1.64 standard deviations from the mean correspond to the upper and lower 5% of scores respectively. This definition is somewhat arbitrary, and is considerably more relaxed than the definition usually used for univariate outliers. Univariate outliers are usually considered to be those cases that lie in the range of ±3.29 standard deviations or greater from the mean (Tabachnick & Fidell, 2001). The reason for using less stringent criteria in this study is that even if a relatively large dataset is used, too many cases that are substantially different from the majority of other cases (i.e., those cases that lie between 1.64 and 3.29 standard deviations of the mean) will be missed if the usual criterion for defining outliers is used. This logic underlines the arguments of Wainer (1976), which suggests that less extreme outliers are still of interest even though they may not meet the strict definitions of an outlier typically provided in statistical research (e.g., Tabachnick & Fidell).

If the distribution of data is approximately normal, then the values of ± 1.64 standard deviations from the mean should coincide closely with the 5th and 95th percentile points of the distribution of scores. However, if the distribution is not normal and lacks symmetry, then the 5th and 95th percentile points will not necessarily lie 1.64 standard deviations either side of the mean. When this happens, outliers can still be identified by selecting those scores that lie in the tails of the distribution beyond the 5th and 95th percentile points. However, in order to ensure that the cases lying beyond these percentile points are sufficiently distant from most other scores, it is necessary to examine the level of kurtosis of each distribution. Large positive kurtosis values are not desirable, as this indicates a relative unanimity of data near the mean, with the tails of the distribution shortened. With shortened tails, the distribution of scores will not extend to the maximum and minimum permissible values of the subscales. If this happens, then even those scores located beyond the 5th and 95th percentile points will not be sufficiently different from the majority of data to meet the definition of an outlier for this study. In such cases, there will be no outliers to identify and analyse.

While formal inferential tests are available to determine the significance of kurtosis values, Tabachnick and Fidell (2001) recommend examining the shape of the frequency distributions, as significant values of kurtosis are likely to be detected for large samples, even if the kurtosis value is close to zero. Furthermore, at the very least, for each subscale there should be cases that are very close to, or equal to, the maximum and minimum scores allowable for the subscale. If those conditions are not met, then none of the scores for the subscale will meet the criteria required to be classed as an outlier.

This process of selecting extreme cases that are distant from the majority of cases is made easier if the variability of scores is higher. That is, univariate outliers are easier to identify if there is more room on a scale for scores to vary. If the allowed responses for individual items can only fall between 1 and 5 (or even between 1 and 7), as is often the case for Likert scales, then it is difficult for meaningful outliers to emerge, and hence be identified. However, the use of subscales permits a greater variation in scores among respondents (e.g., some subscale scores can range between 5 points and upwards of 25 points, depending on how many individual items make up the subscale). Therefore, it is recommended that only subscales, rather than individual items, be analysed when using the alternative procedure described in this study.
An examination of frequency distributions for each subscale for the exemplar dataset of this study showed an absence of any positive kurtosis. Furthermore, for each subscale there were scores ranging over the full range of all permissible values. This meant that for each subscale there were scores that were sufficiently distant from the majority of data to meet the specifications for being an outlier.

Once determining that the scores for a subscale spanned the full range allowable, it was then necessary to determine if outliers should be defined in terms of their distance from the mean, or by percentile scores. To determine which method to use, the frequency distribution of each subscale was observed for symmetry. All of the distributions showed only moderate skewing, and so those cases that lie 1.64 standard deviations or more either side of the mean were considered outliers.

Stage 2 - Analysing Outliers

It is expected that within any large dataset there will be a small proportion of cases that are atypical in nature, and will thus be recognised as outliers. Once outliers have been identified for each subscale, the next step is to determine if they are randomly distributed throughout the organisation, or whether they follow some pattern. The number of outliers observed within a given category (e.g., a particular age group or district) can be compared with what would be expected given the number of cases in that category using a chi-square test. For example, if the proportions of survey participants across four different age groups are observed to be 40%, 30%, 20%, and 10% respectively, than it would be expected that the number of outliers for any survey subscale be proportionality split to reflect these ratios, assuming the number of outliers was uniformly distributed.

What is unclear at this level of analysis of the chi-square test, is exactly where the discrepancies lie. A significant chi-square test only reveals that at least one of the observed counts is not a chance deviation from its expected count; it does not reveal which one (Siegel & Castellan, 1988). A closer examination of the magnitude of the residuals (differences between observed and expected values) for each cell of the chi-square table however, can indicate which cells are contributing to the chi-square test being significant. One way to determine which residuals are most likely contributing to a significant result is to use the adjusted standardised residual (ASR). This statistic is the residual divided by an estimate of the standard error. ASRs with an absolute value greater than 2 are statistically significant (Ishigooka et al., 1998).

Results

To determine the utility of analysing univariate outliers for interpreting employee survey data, it is helpful to provide an example of the sorts of results typically reported in an employee survey, and then show the results that would be produced using the novel statistical procedures discussed in this chapter for comparison. This section will commence with a presentation of the sorts of results typically generated when analysing employee survey data. This will be followed by a presentation of the sorts of results that can be produced from an analysis of outliers. It is important to note, that the examples provided here are presented in a format suitable for management. They are not necessarily presented in a format that would be provided to survey participants or published in an in-house document for general readership.

Standard Procedure for Reporting Survey Results

One method of presenting results for the QPASS proposed by Hart et al. (1996) is to calculate the mean score for each subscale so they can be compared with each other. Such an approach is common and provides information about the relative importance of each construct being measured. Given that the size of subscales varies widely (depending on how many questions were used to construct it), it is first necessary to convert the subscale scores to a common metric. Hart et al. recommend converting respondents' scores for each subscale to a percentage score. Table 4 shows the mean percentage scores for each subscale for the entire organisation, as well as for the four individual districts.

Table 4 Mean Percentage Scores for Each QPASS Subscale by District, 2002

	District				
Subscale	D1	D2	D3	D4	Total
Quality of Work Life	50.1	51.8	54.2	45.6	51.1
Individual Morale	53.3	55.7	59.6	48.7	55.1
Psychological Distress	30.9	30.8	32.1	31.4	31.2
Workplace Morale	51.3	53.1	49.8	43.7	50.7
Workplace Distress	57.0	55.2	53.4	61.0	56.0
Supportive Leadership	55.2	57.6	54.9	48.0	55.2
Participative Decision-making	47.5	52.3	51.5	41.4	49.9
Role Clarity	58.1	61.4	59.5	54.7	59.6
Professional Interaction	60.4	62.8	61.0	59.0	61.6
Appraisal and Recognition	46.5	51.9	47.2	40.9	48.6
Professional Growth	47.0	51.6	48.9	40.8	48.8
Goal Congruence	55.5	59.0	55.9	49.9	56.5
Excessive Work Demands	59.0	59.7	53.3	56.4	57.7

A casual perusal of these tables allows the reader to easily see any noteworthy findings such as trends, unexpected or unusual results, etc. For example, the table shows that issues relating to professional interaction (as measured by the Professional Interaction subscale) are generally scored highly compared to the other subscales. Data from these tables can also be presented pictorially in the form of pie charts or bar charts to facilitate interpretation.

In summary, survey results that focus on reporting averages are easily produced, readily accessible and easy to understand. They are useful when the most important information from an employee survey dataset is quickly required. Furthermore, they are appropriate for a general readership. However, while the use of averages provides a good insight into what is relevant to most staff, they are limited where diversity exists among staff and the different subgroups (e.g., age groups, location, etc.) within the organisation. Analysing outliers is one way of investigating the implications of such diversity.

Alternative Procedure for Reporting Survey Results

Once the outliers were identified for each subscale using the procedures previously discussed, a series of chi-square tests was conducted to determine the nature of the distribution of the outliers throughout the dataset (i.e., were the outliers distributed randomly, or was there some pattern to their distribution). As may be expected, not every chi-square test yields a significant result. Table 5 provides a summary of the results, with the significant results appropriately identified.

Table 5 Summary of Chi-square Results for the Analyses of Univariate Outliers

Subscale	TIO[a]	TICP[b]	District	Gender	Age[c]
Quality of Work Life	12.5	17.4*	3.1	0.0	4.0
Individual Morale	32.3*	24.9*	22.3*	4.6	11.4
Psychological Distress	9.6	13.9*	14.3*	1.1	20.4*
Workplace Morale	30.8*	24.3*	19.9*	3.3	8.8
Workplace Distress	22.7*	14.4*	12.0	2.0	6.8
Supportive Leadership	25.2*	17.3*	12.4	1.0	6.9
Participative Decision-Making	14.7*	13.9*	25.5*	4.0	5.7
Role Clarity	4.8	5.4	11.7	2.3	13.8*
Professional Interaction	21.1*	13.4*	12.3	3.6	4.3
Appraisal and Recognition	24.2*	13.7*	23.5*	7.2*	12.4
Professional Growth	47.9*	36.7*	35.6*	3.6	19.8*
Goal Congruence	19.7*	13.5*	13.4*	7.3*	13.4*
Excessive Work Demands	18.3*	28.2*	8.1	2.8	3.9

*Note. N ranged between 1,838 and 1,920, due to missing data. Chi-square analyses were conducted for all variables with 6 degrees of freedom, except for gender, which had 2 degrees of freedom. [a]TIO = Time in Organisation. [b]TICP = Time in Current Position. *p < .05.*

A requirement of the chi-square test is that no more than 20% of the expected counts are less than 5 (Moore, 2000). For this dataset the age group variable yielded expected counts less than five for more than 20% of the expected counts for all subscales. To overcome this problem, those respondents whose age was less than 21 years were grouped into an age category comprising all respondents under the age of 30 years. Similarly, those respondents over the age of 60 years were grouped into an age category comprising all respondents over the age of 50 years.

For consistency, clarity and ease of demonstration, the employee variable of district from Table 2 will be chosen and analysed with respect to a selected subset of subscales from Table 3. The subscales chosen are Individual Morale, Workplace Morale, Participative Decision-Making, Appraisal and Recognition, and Professional Growth. This combination of variables gives 5 separate chi-square analyses. The results of these analyses are shown in Tables 6 to 10. For each analysis, a brief account of the main findings are provided. To facilitate interpretation of each analysis, those cells contributing to the significant result of the chi-square tests are appropriately identified. The utility of these analyses is discussed in the next section of this chapter. A more meaningful discussion of these results will be provided in the discussion section of this chapter.

At this stage, it is worth noting that the effect size of each chi-square analysis was small, Cramér's V < .1 (Cohen, 1988). A post hoc calculation of power was performed with 6 degrees of freedom, an alpha level of .05, sample sizes in the vicinity of 1,800 to 1,900 (depending on how many cases were missing), and effect sizes ranging between .06 and .1. For these values, power ranged between 47% and 92% (Faul & Erdfelder, 1992). The implications and meaning of such small effect sizes for a study of this nature are not known. However, as will be discussed later in this chapter, there is still value in discussing results that yield small effect sizes. Therefore, a sample of entries from Table 5 that yielded significant results will be discussed in the paragraphs that follow.

District and Individual Morale

The chi-square results for this analysis are shown in Table 6. While 22.0 positive outliers were expected for D3, there were actually 35. This observation shows that D3 had more people than expected who felt very favourably about issues relating to individual morale. For D4, while 14.7 negative outliers were expected, 22 were observed. This shows that there were more people than expected who felt very negative about issues relating to individual morale. Consistent with this finding, is that while there were 19.0 positive outliers expected for D4, only 6 were observed. This shows that D4 had fewer people than expected who felt very positive about issues relating to individual morale compared with the other districts in the organisation.

Table 6 Chi-Square Analysis of District and Outlier Status for the Individual Morale Subscale

District	Negative outliers	Non-outliers	Positive outliers
D1	6 (7.1)	145 (141.8)	7 (9.1)
D2	43 (45.0)	899 (899.1)	60 (57.9)
D3	13 (17.1)	333 (341.9)	35 (22.0) [a]
D4	22 (14.7) [a]	300 (294.3)	6 (19.0) [a]

Note. Observed counts are shown in each cell with the *expected counts in parentheses. [a]Cells with an adjusted standardised residual greater than 2.0.*

District and Workplace Morale

The chi-square results for this analysis are shown in Table 7. For D1 and D3, all expected counts appear to be close to their observed counts. However, for D2, while there was an expected count of 50.8 positive outliers, there were actually 66. This suggests that there were more people than expected in this district who felt very positive about issues relating to workplace morale than expected when compared with other districts. This is the reverse of D4. While there were 16.5 positive outliers expected, only 3 were observed. This finding shows that there were fewer people who felt very positive about issues relating to workplace morale than would be expected compared with the other districts.

Table 7 Chi-Square Analysis of District and Outlier Status for the Workplace Morale Subscale

District	Negative outliers	Non-outliers	Positive outliers
D1	5 (7.5)	146 (142.7)	7 (7.9)
D2	44 (48.1)	909 (920.0)	66 (50.8) [a]
D3	19 (18.8)	360 (359.3)	19 (19.8)
D4	22 (15.6)	305 (298.0)	3 (16.5) [a]

Note. Observed counts are shown in each cell with the expected counts in parentheses. [a]Cells with an adjusted standardised residual greater than 2.0.

District and Participative Decision-Making

The chi-square results for this analysis are shown in Table 8. For D2, there was an expected count of 73.4 positive outliers, with 89 observed. This shows that there were more people in this district than expected who felt very positive in regards to issues relating to participative decision-making. For D4, there was an expected count of 23.6 positive outliers, with 6 observed. This shows that for D4, there were fewer people than expected who felt very positive in regards to issues relating to participative decision-making. Consistent with this result is that while there was an expected count 16.4 negative outliers, 27 were observed, showing that D4 had more people than expected who felt very negative in regards to issues relating to participative decision-making.

Table 8 Chi-Square Analysis of District and Outlier Status for the Participative Decision-Making Subscale

District	Negative outliers	Non-outliers	Positive outliers
D1	7 (7.9)	140 (138.8)	11 (11.3)
D2	42 (50.9)	893 (899.7)	89 (73.4) [a]
D3	19 (19.9)	350 (351.5)	31 (28.7)
D4	27 (16.4) [a]	297 (290.0)	6 (23.6) [a]

Note. Observed counts are shown in each cell with the expected counts in parentheses. [a]Cells with an adjusted standardised residual greater than 2.0.

District and Appraisal and Recognition

The chi-square results for this analysis are shown in Table 9. For D2, there was an expected count of 54.2 positive outliers, with 73 observed, showing that there were more people than expected who felt very positive in regards to issues relating to appraisal and recognition. With regards to the negative outliers, there was also an expected count of 54.2, with 41 observed, showing fewer people than expected who felt very negative in regards to issues relating to appraisal and recognition. For D4, while there was an expected count of 17.4 positive outliers, only 6 were observed, showing that there were fewer people than expected who felt very positive in regards to issues relating to appraisal and recognition. While there was also an expected count of 17.4 negative outliers for this subscale, 25 were observed. This shows that there were more people than expected in this district who felt very negative in regards to issues relating to appraisal and recognition.

Table 9 Chi-Square Analysis of District and Outlier Status for the Appraisal and Recognition Subscale

District	Negative outliers	Non-outliers	Positive outliers
D1	10 (8.3)	140 (139.5)	6 (8.3)
D2	41 (54.2) [a]	909 (914.5)	73 (54.2) [a]
D3	25 (21.0)	356 (354.9)	16 (21.0)
D4	25 (17.4) [a]	298 (294.1)	6 (17.4) [a]

Note. Observed counts are shown in each cell with the expected counts in parentheses. [a]Cells with an adjusted standardised residual greater than 2.0.

District and Professional Growth

The chi-square results for this analysis are shown in Table 10. For D2 there was an expected count of 71.0 positive outliers, with 97 observed. This shows that there were more people than expected who felt very positive in regards to issues relating to professional growth. Consistent with this finding was that, while there was an expected count of 67.8 negative outliers for D2, 55 were observed. This shows that there were fewer people than expected in this district who felt very negative in regards to issues relating to professional growth. For D3, while there was an expected count of 26.7 negative outliers for this subscale, 36 were observed, showing that there were more people than expected who felt very negative in regards to issues relating to professional growth. For D4 there was an expected count of 23.0 positive outliers, with only 3 observed, showing that there were fewer people than expected in this district who felt very positive in regards to issues relating to professional growth.

TABLE 10 Chi-Square Analysis of District and Outlier Status for the Professional Growth Subscale

District	Negative outliers	Non-outliers	Positive outliers
D1	11 (10.6)	139 (136.4)	8 (11.1)
D2	55 (67.8) [a]	862 (875.2)	97 (71.0) [a]
D3	36 (26.7) [a]	338 (344.4)	25 (27.9)
D4	25 (22.0)	301 (284.0)	3 (23.0) [a]

Note. Observed counts are shown in each cell with the expected counts in parentheses. [a]Cells with an adjusted standardised residual greater than 2.0.

Discussion

This section will contrast the two sets of results with the specific aim of evaluating the utility of the results generated from the second approach. Both sets of results will be assessed in terms of the strengths and limitations. However, given that the first approach is widely practised and well understood, more discussion will be given to the latter approach.

Standard Approach to Reporting Survey Results

The first set of results provides readily interpretable information. For example, an examination of Table 4 reveals that D4 generally scores lower than the other districts for most of the subscales, except for Psychological Distress, Workplace Distress, and Excessive Work Demands. Higher scores on these subscales indicate problems.

If the overall results are what matters most to managers (and it often is in large organisations), then results that focus on means and totals are fine. Identifying areas of strengths and weaknesses using this approach readily provides valuable information for management to respond to in their quest to improve performance, employee satisfaction, and the management of economic resources. However, this study has argued that there is much to be gained by identifying those employees whose opinions do not reflect the opinions held by the majority. The discussion that follows will provide specific examples of the sort of information that can be identified using the alternative procedure that is not easily obtainable from a routine analysis of employee survey data.

Alternative Approach to Reporting Survey Results

As an example of the sort of information that can be identified by analysing outliers, consider the employee variable of gender. As can be seen from Table 4, the only two subscales that yielded significant results for gender were Appraisal and Recognition, and Goal Congruence. That is, with the exception of these two subscales, there was no evidence to suggest that the distribution of extreme scores (outliers) was any different for males than for females, other than what might be expected by chance.

An examination of Tables 6 to 10 (which are based on analyses of outliers) provides information that is generally consistent with that obtained from Table 5 (analyses using the standard approach). For example, those districts with high percentage scores (relative to the mean as shown in Table 4) for a particular subscale, generally had a higher observed count for positive outliers than what would be expected if outliers were randomly distributed. Further, consistency for this finding was often demonstrated when an observed count for negative outliers was lower than the expected count. These findings are not surprising. However, using outliers provides a window of opportunity to observe results that are not easily seen using the standard approach to analysing survey data. For example, while Table 4 shows that the mean subscale scores for D4 are moderately lower than the mean subscale scores of other districts, it is not known if this finding is because D4 generally has an homogenous set of moderately lower-than average scores, or because it has several very low scores (negative outliers) which tend to drag the mean down.

Using the Individual Morale subscale as an example, Table 6 shows D4 to have more negative outliers than would be expected if the distribution of outliers for this subscale was independent of district. Furthermore, this same table shows that even though D4 has fewer survey respondents than D3, it still has more negative outliers than D3 (22 for D4 as opposed to 13 for D3). These results suggest that for the Individual Morale subscale for D4, the low mean score of 48.7% (see Table 4) is partly attributable to the disproportionately higher number of negative outliers.

As another example, consider the mean scores for D1 and D3 for the subscale of Professional Growth. While Table 4 shows the percentage score for each of these two districts (47.0% and 48.9% respectively) to be very close to that of the mean percentage score of all districts (48.8%), Table 10 provides additional information not obtainable from Table 4. For D1, the expected numbers of positive and negative outliers are close to their respective observed scores. However, for D3, the number of observed negative outliers is significantly greater than its corresponding expected number (36 observed and 26.7 expected). So despite having a favourable mean score (by comparison with the overall mean), this total score tends to mask the disproportionately higher number of people with very negative attitudes and opinions regarding issues relating to professional growth (e.g., opportunity for training, and career development).

These examples clearly demonstrate the type of information that can be missed when focussing on a single score to describe employee survey data. Focusing on outliers not only reveals more information about subgroups of employees, but it can also provide an insight into the organisation. For example, if people with extreme scores on a subscale are not randomly distributed throughout the organisation, but tend to bunch in particular categories of a variable (e.g., district), then this finding is likely a reflection of the organisation, rather than the individual.

Potential Limitations

The above-mentioned examples only provide a snapshot of what is available when analysing outliers in an employee survey dataset. They strongly suggest therefore, that the identification and analysis of outliers is a worthwhile pursuit. However, in the interests of providing a fair evaluation, weaknesses or concerns of this approach need to be addressed. The following topics will be addressed in the paragraphs that follow: (a) effect size, (b) sample size, (c) confidentiality, (d) skills required to conduct the analyses, (e) taking action, and (f) interpretation of significant results.

Effect Size

As mentioned previously, it is worth noting that while some of the chi-square analyses yielded significant results, their effect size given by Cramér’s V was small on every occasion. While a small effect size may imply triviality, Prentice and Miller (1992) suggest that small effects may have large implications in a practical context. Furthermore, Cohen (1988) advises using the conventional definitions of what constitutes a small, medium, or large effect size as a general frame of reference, and not to take them too literally. With no similar studies to compare this finding with, the full meaning and implications of a small effect size are not known. However, this does not necessarily mean that the claims made in this study are meaningless or unfounded. When conducting research in a new field where the most effective variables to study are not yet known, any new knowledge gained can be considered worthy, even if only small effect sizes are achieved. Supportive of this notion, Prentice and Miller (1992) suggest that whatever their size, effects are important to the extent that they have had a major impact on thinking in a given field.

The sample results given by this new approach, argue well that conventional methods of analysing employee survey data could benefit from a major revision. At the very least, the analyses conducted in this study can demonstrate the utility of this approach when the nature of the small effect size is better understood. Strategies for understanding the small effect size obtained in this study are addressed later when discussing future research.

Sample Size

The sample size used for this study, or more specifically the sample size of some categories of particular variables (e.g., < 21 years age group) meant that some adjustments had to be made in order to use the chi-square test. While these adjustments (i.e., combining smaller groups to make a larger one) resolve the problem on a statistical level, they can do so at a cost - a loss of information. Given that those survey respondents classified as outliers, are themselves, by definition only a small group to begin with (approximately 10% of the sample), total sample size is an important issue. The total sample size and the size of each category of all variables need to be considered when determining if an analysis that uses the chi-square test is appropriate.

Confidentiality

Employee surveys are typically guarded by strict policies regarding confidentiality. Assurance of confidentiality is assumed to promote honesty among survey participants. While anonymity is still maintained with this alternative process, the narrowing down of responses into positive and negative outliers may unsettle some employees. If an employee knows himself or herself to be someone with extreme opinions, particularly unfavourable opinions, they may fear being identified if they respond to survey items honestly. Furthermore, even if confidentiality is assured, it is possible that staff will perceive this new level of analysis as one step closer to a complete loss of privacy in the workplace. Therefore, advancement in the application of statistical procedures and reporting of employee survey results needs to be applied in unison with consideration for the well-being of employees.

Skills Required to Conduct the Analyses

One reason why the use of averages is so common when reporting employee survey results, is that calculating means is very easy. The procedure discussed in this thesis relies on statistical techniques that go beyond basic statistical procedures. However, it is reasonably expected that someone with appropriate experience in statistics would be quite able to conduct the analyses. Organisations would then have to ensure they employed a suitably qualified person. Given the potential gains to be achieved using this alternative procedure, an investment in employing or skilling an appropriate person is justified. The procedure presented in this study is in its embryonic stages. However, once the procedure of analysing outliers has been finely tuned and its efficacy established, development of a software application to perform the analyses would not be too difficult.

Taking Action

Identifying strengths and weaknesses within an organisation based on survey results is fine. However, once identified, how should management respond? Indeed, could management be the problem if weaknesses are identified? The response taken varies between organisations and within organisations. Those who strive for continual improvement embrace such findings, while others may disregard them and continue with business as usual. While the results from this study have demonstrated the potential of the alternative procedures, it is recommended that adopting these procedures should be accompanied with a plan for what action to take once survey results have been reported. If action is not taken once results have been generated, then there is little point in conducting the analyses, or even doing the survey.

Interpretation of Significant Results

In essence, a statistically significant result simply means that an observed effect or difference is assumed to be real, and not just due to chance. In this study, identifying statistically significant results was done to hopefully identify areas of success and areas where there was room for improvement in an organisation. However, the implications of a significant result for the analyses undertaken in this study are not always conveniently categorised in terms of weaknesses or success. Some of the differences in subscale scores, though real (i.e., statistically significant) may be normal, and are neither a strength nor a weakness. For example, if employees are compared on the basis of how long they have been in their current position, it could be expected that employees who have been in their current position for less than a year, to be more enthusiastic and have more favourable attitudes regarding their workplace than employees who have been in the same position for more than a year. This is a problem in questionnaire design in general, and not that of the statistical techniques used to analyse survey data.

Evaluation and Further Research

New ideas should always be thoroughly researched and tested before promoting them as worthwhile. A new product, treatment or procedure normally has to undergo many revisions before it is deemed ready for use. The alternative procedure for analysing employee survey data discussed in this thesis is no exception. However, initial results generated using the alternative procedure are promising. While issues relating to limitations of the procedure were raised and addressed, there was no compelling reason to suggest that further investigation should not be conducted. On the contrary, the findings discussed in this chapter strongly suggest that further research and testing should take place.

Further research could focus on using the same procedure described in this study (or variations thereof) to a range of different employee survey datasets to explore and increase its external validity. The more opportunities available for testing this procedure, the more likely it is that the strengths and weaknesses of the alternative approach, and the conditions under which its application is most optimal, will be better understood. One way of generating datasets different from the one used in this study is to use a different employee survey questionnaire. Alternatively, a different set of subscales derived from the QPASS could be analysed. In addition, administering a survey to an organisation with a different organisational structure and different set of organisational aims from the government department used in this study would also yield a dataset with different attributes to the one used in this study. Finally, aspects of the statistical analyses used in this study can be varied (e.g., using different criteria for defining outliers, or using more stringent alpha levels in the chi-square analyses) to further investigate the utility of this alternative procedure.

With regards to the small effect size, additional research of a qualitative nature (e.g., interviews with staff and employees to ascertain their opinions on workplace issues) may provide an insight into what a small effect size means for a study of this kind. For example, for each subscale and employee variable combination that yielded a significant chi-square result, interviews with staff and management may reveal what meaning these findings have in the organisation. Doing so is very likely to provide evidence either for or against the value of considering the heterogeneity of survey responses (as measured by outliers) as opposed to relying on a single summary statistic (e.g., the mean) when analysing employee survey data.

Conclusion

While survey results that focus on reporting averages tend to be a commonly used form of reporting for survey results, they are not without their limitations and weaknesses. An alternative approach that focussed on studying outliers in employee survey datasets was proposed. This alternative approach was tested using an exemplar dataset and then contrasted with the standard approaches as a means of evaluation. A comparison of the results showed that there is merit in examining outliers, as the alternative procedure produced useful and interesting findings that could not be easily replicated using the standard approach to analysing employee survey data. Though there were limitations associated with the alternative approach (both statistical and practical), these were addressed. The initial results are encouraging, and therefore suggest that further research is a worthwhile endeavour.

References

Barnett, V., & Lewis, T. (1994). Outliers in statistical data. (3rd ed.). Chichester: John Wiley & Sons.

Chan, D. (2006). Multilevel research. In F. T. L. Leong & J. T. Austin (Eds.), The psychology research handbook: A guide for graduate students and research assistants (2nd ed., pp. 401-418). Thousand Oaks: Sage.

Church, A. H., & Waclawski, J. (1998). Designing and using organisational surveys. Cambridge: Gower.

Cliff, N. (1996). Answering ordinal questions with ordinal data using ordinal statistics. Multivariate Behavioral Research, 31(3), 331-350.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, New Jersey: Lawrence Erlbaum.

de Vaus, D. A. (2002a). Analyzing social data: 50 key problems in data analysis. London: Sage.

de Vaus, D. A. (2002b). Surveys in social research (5th ed.). Crows Nest: Allen & Unwin.

Faul, F., & Erdfelder, E. (1992). GPOWER: A priori, post hoc, and compromise power analyses for MS-DOS [Computer Program]. Retrieved October 18, 2006, from http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/

Filzmoser, P. (2005). Identification of multivariate outliers: A performance study. Australian Journal of Statistics, 34(2), 127-138.

Fralicx, R., McCauley, D. P., & Gilbert, P. J. (2002). Solving business problems with the human capital survey. Strategic Communication Management, 6(2), 34-37.

Gregoire, T., & Driver, B. (1987). Analysis of ordinal data to detect population differences. Psychological Bulletin, 101(1), 159-165.

Hart, P. M., Griffin, M. A., Wearing, A. J., & Cooper, C. L. (1996). Manual for the QPASS survey. Brisbane: Organisational climate and performance project, Office of the Public Service.

Hawkins, D. M. (1980). Identification of outliers. London: Chapman and Hall.

Hofmann, D. A. (2002). Issues in multilevel research: Theory development, measurement, and analysis. In S. G. Rogelberg (Ed.), Handbook of research methods in industrial and organisational psychology (pp. 247-274). Malden, MA: Blackwell.

Ishigooka, J., Iwao, M., Suzuki, M., Fukuyama, Y., Murasaki, M., & Miura, S. (1998). Demographic features of patients seeking cosmetic surgery. Psychiatry and Clinical Neurosciences, 52, 283-287.

Kim, J. (1975). Multivariate analysis of ordinal variables. The American Journal of Sociology, 81(2), 261-298.

Lance, C. E., Stewart, A. M., & Carretta, T. R. (1996). On the treatment of outliers in cognitive and psychomotor test data. Military Psychology, 8(1), 43-58.

Lovie, P. (1986). Identifying outliers. In A. D. Lovie (Ed.), New developments in statistics for psychology and the social sciences (pp. 44-69). London: The British Psychological Society.

McIver, J. P., & Carmines, E. G. (1981). Unidimensional scaling. Beverly Hills: Sage.

Mertens, D. (2005). Research and evaluation in education and psychology (2nd ed.). Thousand Oaks, California: Sage.

Moore, D. S. (2000). The basic practice of statistics (2nd ed.). New York: W. H. Freeman.

Nadler, D. A. (1996). Setting expectations and reporting results: Conversations with top management. In A. I. Kraut (Ed.), Organisational surveys (pp. 177-203). San Francisco: Jossey-Bass.

Orr, J. M., Sackett, P. R., & Dubois, C. L. (1991). Outlier detection and treatment in I/O psychology: A survey of researcher beliefs and an empirical illustration. Personnel Psychology, 44, 473-486.

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). Practical Assessment, Research & Evaluation, 9(6). Retrieved October 1, 2006 from http://PAREonline. net/getvn.asp?v=9&n=6

Patrick, J., Albion, M. J., McKeon, C., Fogarty, G. J., & Machin, M. A. (2006). A re-evaluation of the construct and psychometric properties of the Queensland Public Agency Staff Survey. Unpublished manuscript, University of Southern Queensland, Toowoomba.

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. Psychological Bulletin, 112(1), 160-164.

Rousseau, D. M. (1985). Issues of level in organisational research: Multi-level and cross-level perspectives. Research in Organisational Behavior, 7, 1-37.

Rousseeuw, P. J., & van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. Journal of the American Statistical Association, 85(411), 633-639.

Schwager, S. J., & Maragolin, B. H. (1982). Detection of multivariate normal outliers. The Annals of Statistics, 10(3), 943-954.

Siegel, S., & Castellan, N. J. (1988). Nonparametric statistics for the behavioral sciences (2nd ed.). New York: McGraw-Hill.

Tabachnick, B. G., & Fidell, L. S. (2001). Using multivariate statistics (4th ed.). Boston: Allyn and Bacon.

Thomas, H. (1982). IQ, interval scales, and normal distributions. Psychological Bulletin, 91(1), 198-202.

Townsend, J. T., & Ashby, F. G. (1984). Measurement scales and statistics: The misconception misconceived. Psychological Bulletin, 96(2), 394-401.

van Alphen, A., Halfens, R., Hasman, A., & Imbos, T. (1994). Likert or Rasch? Nothing is more applicable than good theory. Journal of Advanced Nursing, 20, 196-201.

Wainer, H. (1976). Robust statistics: A survey and some prescriptions. Journal of Educational Statistics, 1(4), 285-312.

Walters, M. (1990). What about the workers: Getting the best from employee surveys. London: Institute of Personal Management.

Wright, D. B. (2003). Making friends with your data: Improving how statistics are conducted and reported. British Journal of Educational Psychology, 73, 123-136.