

5.0 METHOD AND APPROACH FOR ANALYSIS OF DATA

5.1 What is Content Analysis?

Content analysis is a procedure for studying the content and themes of written or transcribed qualitative data, usually by reducing it into more structured or concise units of information. Most content analysis uses a scheme of categories that are relevant to the research objectives and a number of judges who systematically characterise the content in terms of these categories. The judgments have to be validated, or agreed on between the judges (who make their interpretations independently) to ensure that the judgments of content are not idiosyncratic.

In this research, content analysis procedures were used to address the information provided by the free and open-ended responses to interview questions. Such a procedure has a number of advantages over more conventional 'fixed-response' interview techniques. It allows whatever is salient to the respondent to emerge and therefore prevents the undue constraint of answers by the fixed options usually given for survey questions. In an exploratory study primarily concerned with experiences, such freedom of response was particularly important. Researchers could not know beforehand the possible diversity that might be offered, and therefore would have found it impossible to provide a sufficient range of fixed-response options. This means, however, that it is likely that the coding system will be quite complex and diverse, at least if the responses contain such diversity and the coding scheme reflects this. It also means that the proportion of respondents referring to a single category is less likely to be a large majority, yet the category is still important given that it was spontaneously expressed, rather than prompted in a fixed-format interview.

5.2 Developing the Coding Categories

The coding categories were, in the first instance, derived from the literature on outdoor recreation experiences, in particular from the work of Scherl (1988a, 1988b, 1990) and Hunnam (1990), both in Australia, and Graefe et al. (1988), in the United States of America. The common characteristic of all of these studies is that they used content analysis to summarise data obtained from visitors' reports of outdoor recreation and tourism experiences elicited through open-ended research designs.

Scherl (1988a, 1990) developed a hierarchical taxonomy of the different domains of a wilderness experience based on the literature, some other research results on participants' constructions of their experience (Scherl 1988b), and validated against log book contents. The work of Hunnam (1990) is the first on outdoor recreation and tourism experiences in a reef environment using an open-ended interviewing methodology. He, similarly, based on previous research (particularly the work of Scherl 1988a) and his own understanding of the reef experience situation from a manager's perspective, used a number of experience factors as the framework for content analysis of day visitors' interview transcripts. Graefe et al. (1988) developed a number of user behaviour interview scales based on a literature review completed by Berger and Schreyer (1986). These scales were used to content analyse interviews with people rafting rivers in the United States.

The research carried out during the 1980s has grown out of a perception that we need to understand the nature of outdoor recreation experiences from the perspective of the participants and with as little imposition of meaning as possible from the researcher. It is, however, pertinent to make the observation that there are similarities between these experience domains, factors and behaviour scales and the Recreation Experience Preference (REP) scales developed by Driver and co-workers in the 1970s (e.g. Driver 1977).

Although in the present study an initial, tentative list of categories was devised based on previous research, the final taxonomy and the list of categories used to code the questions with specific themes emerged from the data contained in the interviews. A number of people involved in the project, including some of the researchers, interviewers, the people responsible for data analyses and the coders, went through a process of 'mock' coding, that is they attempted to fit the content

of the interviews into the existing list of categories. (This process was undertaken during the interview period by the first author; for one week after coming back by the first and third authors and one of the interviewers; and for five days in a group context by the coders and all of the project researchers.) When the content could not be placed into any category a new one would be developed. After each person did this all of the new additions were integrated into the existing categories during a group discussion. Moreover, the taxonomy was also structured by the group in a way that made logical sense for coding.

In this way, a comprehensive list of categories was developed (see appendix 3 for details). Because the survey had both general experiential and specifically focused items, the coding categories were developed separately for different questions, and were tailored to the focus of the item (see appendix 4). For several general questions about experience and motivations (questions 1, 2, 22, 23, 24), the categories were formed into a single taxonomy applied to all of these general items. This taxonomy contained a superordinate level, and within each section in the superordinate level there were a number of subordinate categories (e.g. Scherl 1988a, 1990). The superordinate categories were self/experience, type of activities, social environment, physical environment—nature factors, physical environment—natural conditions, physical environment—interpretative, environment—human interactions, managerial/organisational factors, trip overall, and miscellaneous (figure 2). There were varying numbers of sub-categories within each of these. Therefore, as an example, respondents answering **question 1**, *'Tell me about your visit to Lady Musgrave and what sort of experience has today been for you'* might have talked about positive feelings; the things that they did that were important enough to be mentioned in response to such a general question about experience; some aspects of the social environment and the natural environment that were salient to them; and perhaps made some reflections about managerial organisational factors, the weather, or the history of the place. The taxonomy coding scheme, as it was developed, would have captured the importance of these general categories, as well as the specific items within them.

5.3 The Process of Content Analysis

The material provided in visitors' accounts of their experiences and perceptions of management issues during their stay at Lady Musgrave Island and Reef resulted in valuable records, reinforcing Miles and Huberman's (1984) suggestion that qualitative data are a 'source of well-grounded, rich descriptions and explanation of process occurring in local context' (p. 41). In this study, the interviews were content analysed by two judges.

All of the interviews were transcribed and the coding unit (i.e. the unit of analysis) was defined as the string of words and sentences elicited as a response to one question. Therefore, judges read each question separately, and, using the set of categories developed for that question, decided how applicable each of the categories was to the content. If the content in any part of the question 'satisfied' a category, then this category was recorded as applicable (see appendix 5). No attempt was made to measure the proportion of the total response to a question that was devoted to a particular topic, since the quantity of talk was thought to not necessarily be proportional to importance. It may, for example, only take a few words to say, 'I've had a great day, and it just feels so good to be here in such a beautiful natural setting', but it may require much more talk to attempt to explain the strange nudibranch and its behaviour that a respondent may have seen while snorkelling.

The content analysis procedure used here was one that allowed judges to characterise the content of a response into as many coding categories as they thought appropriate to capture that content. The categories were therefore not used in a mutually exclusive way where only one category was chosen into which was 'fitted' as a 'best fit' all of a unit of content. Judges were also asked to assess the graded extent to which a category was appropriate to the content of a response. The graded scale that was used for this purpose varied from a rating of '0' for 'not at all relevant' through, say, 3 for 'relevant' and 4 for 'very relevant' (see Scherl and Smithson 1987 and Smithson 1987, where this work is reported, for more details). Thus for each unit of context judges selected multiple categories.

TAXONOMY

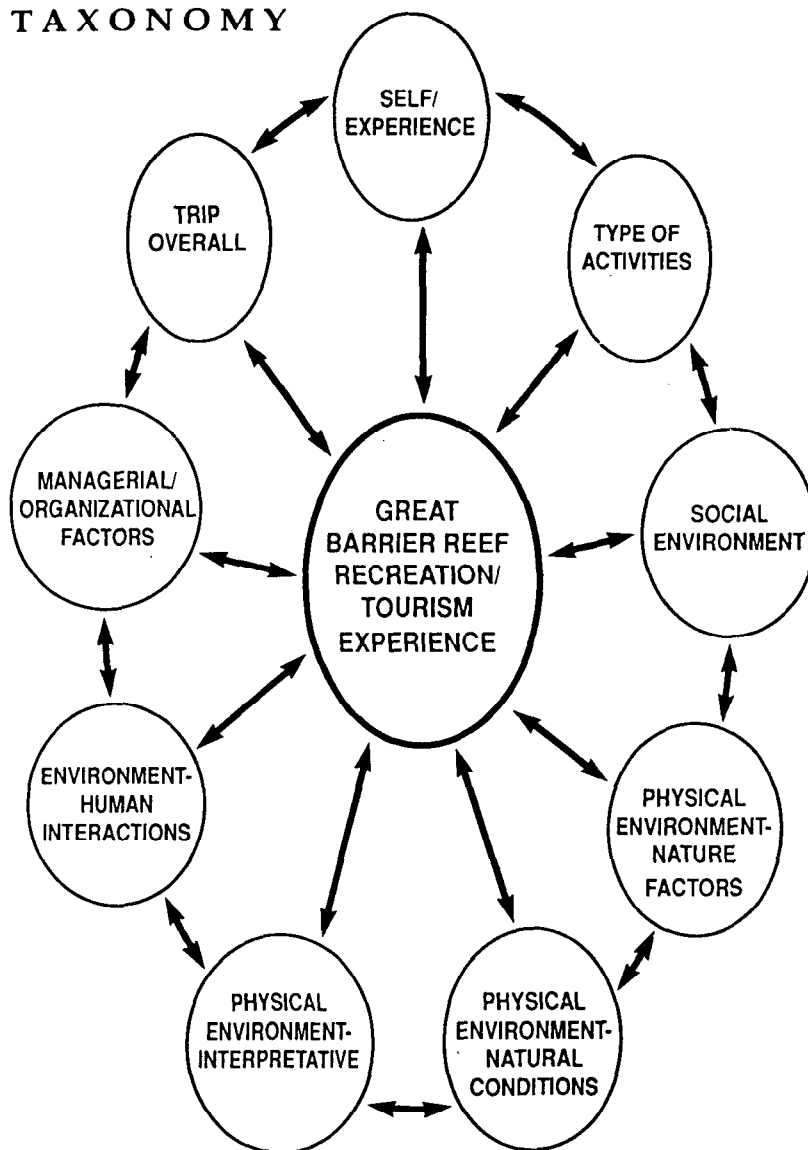


Figure 2. Taxonomy of recreation/tourism experience in the Great Barrier Reef Marine Park

The flexibility of choosing a number of categories to represent visitors' responses is compatible with the definition of an outdoor recreation experience in this study, i.e. that an outdoor recreation experience is considered to be multi-dimensional with a number of these dimensions interacting simultaneously. The methodological flexibility of co-occurrence of categories for coding written material was introduced by Scherl (1988a, 1990) in an attempt to capture this multi-faceted and interacting aspect of outdoor recreation and tourism experiences.

While the use of this technique clearly captures much of the richness and co-occurrence of features of experience, **it also means that readers should be cautious about adding up the reported frequency counts for groupings of categories.** Since the categories were not mutually exclusive, readers cannot simply refer to tables of results and add up, say, the 126 people who mentioned that they went snorkelling and the 72 who said they walked around the island to say 198 people went snorkelling or walked around the island. Some clearly did both, while others may have done one or the other. Where applicable a valid combination of category frequencies was obtained by only counting once the people who mentioned both categories. In places in the text where these additions have been computed, they are reported in *italics*.

5.4 Preliminary Data Reduction

Given that the category lists developed through the above procedure were extensive, the first step of the data reduction procedure was to sort and retain the categories that had proven reasonably salient after all the interviews had been coded. Firstly, after a preliminary combination of the two judges' sets of ratings, categories that were not at all relevant to the content were eliminated, although records were kept of those because this non-use may have been of interest in itself (i.e. those are the categories that do not appear in the results tables but are in the coding scheme, the taxonomy and categories are presented in appendices 3 and 4). Then, since there were many categories that had very low levels of applicability (only applicable to two or three per cent of respondents), an attempt was made to ascertain whether there were natural cut-off points which could be used as an elimination level for categories that were not used to any appreciable extent by the judges. A 'scree' test similar to that used in Factor Analysis was conducted in an attempt to locate such cut-off points. Plots of decreasing frequency of use of all the *non-zero* categories were carried out. Figure 3 shows two examples, i.e. question 1 and question 10.

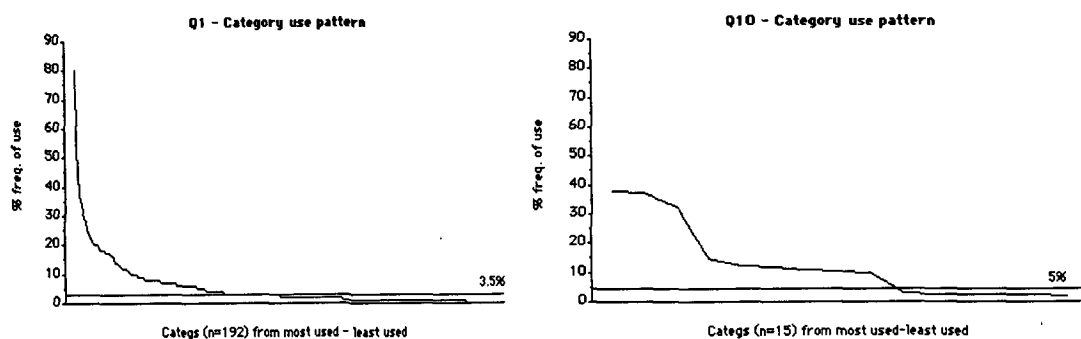


Figure 3. Examples of 'scree' plots of category relevance to content

These graphs, and those for the other questions contained in appendix 6, contain lines indicating the cut-off points. These might be interpreted as the points above which category relevance to content seems to increase appreciably, or sharply. Below them, as you can see, use is at a consistently low level. Results presented in the tables are only the categories that were salient at a level above the cut-off points, which are represented by the lines across the graphs.

5.5 Inter-judge Agreement

Using the category relevance cut-off points, the original judges' files were cleansed of all zero and near zero categories prior to the assessment of inter-rater agreement. This assessment used a variance partition based on Smithson's (1987) 'T2' fuzziness coefficient. This technique, which is similar in many ways to conventional analysis of variance, partitions the total variation in the ratings of applicability of categories. Taking as input the two judges' data sets for all the interviews, the program partitions variation into that due to inter-judge disagreement or difference (*between* judges in ANOVA terms), and that due to the spread or variability of all values across the graded scale, which indicates fuzziness or a lot of gradation in the applicability of categories (*within* variation in ANOVA terms). A coefficient is computed from this partition which represents the proportion of variance due to inter-judge disagreement. This coefficient varies from zero to one, with zero representing extremely low disagreement (high agreement) and one representing maximum possible disagreement. Thus, in the ensuing tables, this measure of *disagreement* is reported for each question. On all of the questions it proved to be extremely low, reflecting independent consensus among the judges about the quantified applicability of the categories used, to the content they characterised. The files of the two judges were then recombined in a reduced version using the geometric mean of the two ratings, and producing the final data set for further analysis.

5.6 Fuzziness of Categories

It is important when *graded* scales of applicability of categories are used, as in this case, that assessment is made of the extent of this gradation within each category. For example, if the judges, when using a graded scale, mostly judged across all cases a category dichotomously as either 'highly applicable' or 'not at all applicable', then the use of gradation across the range of the scale is small, the category is said to have been used in a non-fuzzy way, and the graded scale of category salience adds little information above simple frequencies. If, however, judges were often using middle level ratings on the scale like 'slightly applicable', 'moderately applicable' etc. (a more 'fuzzy' use of the category) then this fuzziness can be measured and incorporated into the analysis. In the case where many middle level ratings are used, then simple frequencies of respondents for whom a category was salient might not reflect the 'true' extent of its salience, since a count of the numbers of people to whom it was applicable (to any degree) does not capture more subtle trends in the gradedness or fuzziness of the judgments. For example, the content of 10 interviews judged as slightly applicable gives the same frequency count as 10 responses judged highly applicable, yet the salience of the category in the latter group is much higher. In such situations, salience of categories is more accurately represented by 'fuzzy per cents' which take account of the full 'gradedness' of the applicability of the category.

One simple way to calculate such a 'fuzzy' category salience proportion is to add up the total graded ratings of content in categories, and for each category divide this by the maximum possible rating total for a category, which would be the number of respondents multiplied by the maximum range of the graded scale. If, however, the categories were not used in a 'fuzzy' way, then the frequencies of respondents mentioning the category does closely parallel the graded 'fuzzy' salience of the category, and 'fuzzy per cent' proportions provide little extra information.

In this exercise therefore, the fuzziness of each category was assessed using the 'T2' fuzziness coefficient (Smithson 1987). This coefficient measures, as explained above, the extent to which judges' ratings were or were not dichotomous or clearly in or out of the category, as opposed to partly in and partly out. It varies from zero to one with zero being highly 'un-fuzzy.'

Overall, the fuzziness indices were extremely low for almost all of the categories. The average overall was 0.02. This meant that the coders were generally using the categories in a clear-cut way. Either the category was judged as highly relevant to a fragment of content, or it was not at all relevant.

In terms of analysis and presentation of results, this meant that there would be little value in adding fuzzy per cents to the tables. Therefore the tables contain only frequency of use of the category and the percentage of total respondents that this frequency represents.

5.7 Data as a Resource for Further Analysis

Overview of Analysis

Material obtained during the open-ended interviews was content analysed using two judges who independently rated the content for its fit into a comprehensive set of coding categories developed by the authors. Judges assess the graded salience of each category to the units of content, with the aim of reasonably capturing and summarising this content in terms of the categorical scheme. Judges' ratings were checked for independent agreement, then combined to produce a single data set of category salience ratings. Categories that were not at all relevant were eliminated using a systematic technique, and tables were produced showing frequencies of respondents whose content fell into each category. Some specific management issues were addressed through the construction of indices of perceived crowding and perceived high density, and the computation of measures of association and difference between user groups on these indices. Because of the enormity of the data set, further specific management questions were addressed through examination of category frequencies. This was the extent of data analyses that this research team could conduct within the time constraints and resource availability.

Resource for Further Analysis

In many respects the data collated from this study represents an important resource in itself. Time and financial constraints (on such a project) limited analyses actually conducted and reported. Apart from those presented here, a number of other analysis possibilities were considered, and represent future possibilities for exploration of this data set. It would, for example, be informative to engage further data reduction techniques such as Factor Analysis and Cluster Analysis on limited subsets of the most salient categories in various subsets of this data. The authors, for example, are particularly interested in whether patterns of association exist between sets of experiential categories and sets of activity categories. It would also be informative to assess whether identifiable factors or clusters would correspond to the user group differences that are already apparent in the data. This would also inform inquiries into whether there are other possible experiential groupings of users other than the *a priori* groupings used in the comparisons contained in this study. Other possibilities include computing measures of association separately across subsets of categories to address specific theoretical or management questions. These, and no doubt others, are things that still remain to be done on this large and complex data set.