The Critical Incident Technique:
An Innovative Qualitative Method of Research

Lorette K. Woolsey
The University of British Columbia

Abstract

There is growing awareness that traditional methods of research that focus on quantification and experimentation are inadequate to meet the challenges of counselling practice (Gelso, 1985; Goldman, 1977, 1978; Hayes, 1981; Hill, 1982; Howard, 1985; Ross, 1981). The need to augment the traditional approaches by adding descriptive and qualitative methods of research has been widely recognized (Goldman, 1977), 1978; Hayes, 1981; Hill, 1982; Lecomte, Dumont, & Zingle, 1981; Valle & King, 1978; Van Leeuwen, 1982; Woolsey, in press). At a time when newer research paradigms and a revised philosophy of science encourage pluralism, diversity and innovation in methods of research, it seems
appropriate to restore descriptive inquiry to a legitimate place among the methods of research in counselling (Keeney & Morris, 1985; Manicas & Secord, 1983; Valle & King, 1978). Such qualitative approaches will play a crucial role in developing a research methodology unique to counselling as a discipline (Friesen, 1983). One such qualitative method is the critical incident technique. This technique is an exploratory method that has been shown to be both reliable and valid in generating a comprehensive and detailed description of a content domain (Anderson & Nilsson, 1964). Neglected for many years, the critical incident method is now exciting interest among counsellors.

This paper will describe the critical incident technique and will illustrate its use.

**HISTORY AND PREVIOUS APPLICATIONS**

John Flanagan (1954) developed the critical incident technique during World War II, in order to identify effective pilot performance. The technique consists of a set of simple interview procedures for collecting information from people about their direct observations of their own or others’ behaviour. For example, in early studies he asked combat veterans to report incidents that were significantly helpful or harmful to their mission. Pilots were asked “to think of some occasion during combat flying in which you personally experienced disorientation or strong vertigo” (p. 329) and to describe what they “saw, heard, or felt that brought on the experience” (p. 329). Flanagan (1954) analyzed the descriptions and produced a list of the components critical for task performance. These lists proved more helpful than the vague descriptions which previously had been used for selection and training.

After the war, Flanagan formally developed the critical incident technique and applied it extensively in industry. The technique was used to develop ethical standards for psychologists, to measure task proficiency, to select and classify personnel, to design job procedures and equipment, to identify motivation and leadership attitudes, and to identify factors in effective counselling (Flanagan, 1954).

Since the 1950’s, as the social sciences increasingly emphasized quantification and experimentation, the critical incident method fell into disuse. Certainly it was not completely abandoned over the past thirty years, being used to study the following: group process (Cohen & Smith, 1976), work motivation (Herzberg, Manseur, & Snyderman, 1959), evaluation of clinical practica (Dachelet et al., 1981), psychological aspects of nursing (Rimon, 1979), the American quality of life (Flanagan, 1978), and the cognition-emotion process in achievement-related contexts (Weiner, Russell, & Lerman, 1979). On the whole, however, it was used only occasionally, and even more to the point, it has not been included among standard methods of research training in counsellor education.
METHOD AND ILLUSTRATION

This section of the paper will describe how to do a critical incident study (from Flanagan, 1954), with illustrations from the author’s research. The two basic principles of the critical incident technique are that factual reports of behaviour are preferable to ratings and opinions based on general impressions and that only behaviours which make a significant contribution to the activity should be included. There are five steps to a critical incident study: (1) determining the aim of the activity to be studied, (2) setting plans, specifications, and criteria for the information to be obtained, (3) collecting data, (4) analyzing the thematic content of the data, and (5) reporting the findings.

1. Determining the Aim of the Activity

   The first step in a critical incident study is to identify the aim of the activity to be studied. This aim must be stated in a simple and clear form. For example, I decided to do an exploratory study to identify the characteristics of same-sex social bonds (Woolsey, 1985). As Flanagan (1954) suggested, I consulted the theoretical and empirical literature and then asked experts in the field to identify the main purpose of same-sex social interactions. “Social support” was the answer, but this formulation was still too vague. I decided to ask respondents to describe incidents which significantly strengthened or deepened the bonds between themselves and their closest friends and family. The interactions between respondents and their closest same-sex associates was the activity studied. Strengthening and deepening the relationship was the aim specified. Clearly this wording exemplifies a more precise statement of aim than does the concept of “social support.”

   I found this step one of the more difficult in the whole process. It required a good deal of hard thinking to clarify and to focus the aim statement to this extent. This step is less difficult for other topics for which aim of the activity has been well established. Even then, however, there are issues to be dealt with. For example, in Easton’s (1986) study of the grief process, experts disagreed as to whether “recovery” or “progress” was the appropriate aim for counsellors working with grieving persons.

   Another consideration regarding the aim is the wording of the description. Wording is important because respondents will be using the aim statement to select incidents to report. The most effective statements of aims use simple everyday language to convey an obvious meaning.

2. Setting the Plans, Specifications and Criteria

   The second step is to decide: (a) which persons will make the observations; (b) which individuals, activities or groups will be observed, and (c) which of their behaviours or experiences will be observed.
(a) *The Observers.* The person making the observations can be anyone who is familiar with the activity and who can make first-hand observations. People can report about themselves or about others; more than one observer can report. For example, both parents and children could be asked to report parental actions that significantly influenced the children's behaviour. If self-reports are used, as is typically the case for studies in counseling, training of the persons making the observations is usually unnecessary. Even in self-report studies, however, it can be helpful to orient respondents before the interview. For example, in a critical incident study of women's self-actualization, the respondents were given the statement of the aim and a list of interview questions several days in advance so that they could think about them (Woolsey & Adler, 1986).

In studies where people are asked to observe others, training may be needed. Such training would require a review of the aim of the activity and a detailed description of the observations respondents will be asked to make. When the situation to be observed is complex (for example, observations of certain interactions in a classroom situation), the observers will need to be able to differentiate aspects of the situation which are to be observed from those which are irrelevant to the study. In this case, the observers should be given supervised practice in making observations. These should involve actual situations if possible, analogous situations if not. The training should continue until all of the observers have the same understanding about what is to be observed.

In critical incident studies, as in any research, attention must be paid to sampling procedures. Because the critical incident technique is descriptive and exploratory, sampling requirements are much less stringent than for traditional methods of research. The major purpose of a critical incident study is to provide complete coverage of the content domain. In the same-sex bonds study, because I wanted a wide range of respondents to give broad coverage of the content domain, very few limits were set on the sampling. Because quantitative comparisons are not made, it is more important to ensure that some persons possessing each of the salient characteristics of the population be included than it is to have the sample representative in all respects. (The latter situation is, of course, always desirable.)

In general, the characteristics of the respondents determine to whom the results of the study can be generalized. It is helpful to gather relevant descriptive biographical data about the respondents, but these data are used only descriptively in a critical incident study. In the same-sex bonds research, for example, the variability in age, socio-economic class and ethnic group revealed the wide range of persons over which the coverage of the content domain extended.

That sample must not consist of persons selected for characteristics that are related in a systematic way to the activity being studied is
particularly crucial for a critical incident study, because the sample size may be fairly small. Thus the call for volunteers in the bonds study was worded in a way that would attract both people with many close ties to others and those with fewer, weaker ties. In general, threats to validity (Campbell & Stanley, 1963) should be kept in mind while planning sampling procedures.

Size of sample is determined on the basis of number of critical incidents and not number of people. There is no strict test for sample size, but a general rule of thumb is to collect incidents until redundancy appears. Flanagan (1954) states that depending on the complexity of the data, it may require as few as 100 incidents or as many as 2,000 to 4,000. It is suggested that a running count of incidents be kept, after the first 100 and again at the estimated halfway point, of the number of new critical behaviours added to the classification system. When only two or three new critical behaviours are added by 100 incidents, data collection can be discontinued. In the same-sex bonds study, I planned to interview 40 women and 40 men but was able to discontinue after interviewing 35 respondents of each sex, because redundancy was obvious (Woolsey, 1985). In the self-actualization study, 25 respondents provided a sufficient number of incidents to meet the redundancy criterion (Woolsey & Adler, 1986).

(b) The Observations. The plans for the observations need to include a description of the individual, activity or group to be observed. The location, the times and the conditions must also be specified. In the same-sex bonds study, the observers were adult respondents, between the ages of 18 and 40, who reported on any same-sex friends that they defined as close to them, whether these relationships were currently active or not. Retrospective data like these are permissible; in general, however, the more recent and direct the observations, the better (Flanagan, 1954). I originally planned to restrict the time frame to incidents that occurred within the past year. Pilot work showed, however, that adults’ important relationships are long standing and that many important incidents may have occurred early in the relationship. Hence, I modified the time specifications, having found that the observer-respondents still had clear and detailed recall of the incidents because these were so salient.

In cases where the observers are reporting on others or on events less important to them, recency is of much greater significance. The criterion for accuracy of reporting is the quality of the incidents themselves. If the details are full and precise, the information can be taken as accurate, whereas if the reports are vague, some of the data may be incorrect (Flanagan, 1954).

It was not necessary to specify either location or conditions in the same-sex bonds study because personal relationship interactions can
occur anywhere and under varying conditions. In some studies, such as job analyses, such specifications would be important.

(c) The Specific Behaviours or Experiences. Finally, the planning must include a description of the general type of activity and the specific behaviours or experiences to be observed. The criteria for importance and relevance of the incidents to the general aim must be specified. For example, the activity studied in the same-sex bonds study was that of closest same-sex friendship and family relationships, at times when these relationships were especially close. The specific incidents “when your friend did something or when something happened between you that significantly strengthened or deepened the bond between you.” The interviewers then ascertained whether this incident had affected the relationship in any of the ways specified as criteria for relevance and importance to the aim (see Table 1).

### TABLE 1
*Interviewer Guide and Sample Questions*

**Establishing the Aim**

- **Focus:**
  “Please focus on your relationship with your best friend or with a close friend of the same sex.”

- **Friendship criteria checks:**
  “What are some of the things you do together?”
  “About how often do you see each other?”
  “What are the main things you talk about?”

- **Context:**
  “Think of a particular time in your relationship when you were very close, when the ties between you were very strong.”
  “What was happening then?”

**The Critical Incidents**

- **Incident:**
  “Please describe a particular incident or incidents when your friend did something or when something happened that significantly strengthened or deepened the bond between you.”

- **Effect of incident criteria checks:**
  Spent more time together. Shared more activities. Exchanged more confidences. Other appropriate criteria.

- **Additional information to clarify incidents:**
  “In what ways did this strengthen the bond?”
  “For how long were you close after this?”
  “How did this affect your relationship as a whole?”

- **Search for any other incidents.**
3. Collecting the Data

Collecting the data generally is done by means of an interview. If written responses are used, many of the guidelines for collecting the interview data still can be applied.

It is helpful to write a detailed interview guide to train interviewers. The questions should be pilot tested in advance. For example, some male respondents in my study were irritated by the original wording “your best friend” because they did not have any one “best” friend. So, the wording was modified to include “close friend,” a term which suited them better. A lengthy and detailed interview guide was used to train interviewers for pilot work. Afterwards, they were given a shortened version which they used during the interview (see Table 1 for sample questions from the latter). It is very important that the interview questions (or directions for written responses) indicate exactly the kind of incidents desired or the data will be diffuse and difficult to categorize.

In my studies, the interviewers used empathic listening and perception checking to be sure that they were correctly understanding and fully capturing the essence of what the respondents were reporting. After the first few incidents were reported, most respondents became adept at fully first few incidents were reported, most respondents became adept at fully reporting incidents, with little prompting.

Interviews should be tape-recorded. Transcribing is helpful, but not essential. I trained research assistants to extract the incidents from the transcribed interviews and put them on cards. The three research assistants each independently extracted incidents from the same interview. Their initial attempts were corrected by my comparing them with the incidents I had taken from that same interview. This process was repeated until all three were consistently obtaining the same incidents and including the same amount of information. I continued to spot check their work by doing my own independent ratings of portions of subsequent interviews. Later, I found that I could have them work directly from the tape recordings. One advantage of working with tapes is that vocal nuances are not lost, as they are on transcripts. The additional nonverbal cues provided by the voice sometimes make the incident clearer.

4. Analyzing the Data

Anyone planning to do a critical incident study should be prepared for the fact that analyzing the data is the most difficult and frustrating part of the method. This phase consists of an analysis of thematic content, arrived at by inductive reasoning. It is necessarily subjective, although some steps can be taken to ascertain whether similar conclusions would be reached by other judges, as discussed later. The objective of data analysis is to provide a detailed, comprehensive and valid description of the activity studied. There are three steps in the process:
(a) selecting a frame of reference, (b) forming categories and (c) establishing the level of specificity-generality to be used in reporting findings.

(a) The Frame of Reference. Selecting the frame of reference is done on the basis of the intended use of the results. For example, the categories would be developed somewhat differently, depending on whether the data were to be used to specify a job description to be used for job classification or to generate procedures for training new personnel. In the same-sex bonds study, the model was intended to be used by counsellors to identify the strength and closeness and of same-sex friendship and family bonds. In the vocational choice study, Friesen and Young (1985) used a theoretical model as a frame of reference for developing categories. In this instance, it should be noted that the choice of such a specific frame of reference sets limits on the exploratory and heuristic use of the method.

(b) The categories. Formulation of the categories is done inductively, by sorting the incidents into clusters that seem to group together. As Flanagan (1954) notes, category formation requires "insight, experience, and judgment" (p. 344) and is unavoidably subjective. In the same-sex bonds study, the incidents were first transferred onto 3 x 5 cards and then sorted into piles. These groupings formed the initial categorizations. Finding these first attempts lacking in richness and distinctiveness, I began again, adding information to the incident cards under the following headings: the general context (the background of the relationship and the circumstances), the specific context (what immediately led up to the incident), the source (what really made the incident so helpful), the agent (who made the difference, the respondent, the friend, or both) and the outcome (what happened to the relationship as a result).

What finally proved satisfactory was to categorize the incident itself, using the source and context information to make additional distinctions. (The agent and outcome information proved, in this case, to be unhelpful for category formation and so that information was disregarded.) Before categorizing an incident, it was helpful to review the interview as a whole. In particular, reviewing the information about the relationship to which the incident referred aided categorization. This kind of trial-and-error procedure seems typical of critical incident studies (Flanagan, 1954).

In the self-actualization study, we were unable to categorize specific incidents and instead had to categorize descriptions of relationships; for it was not a single incident or even a type of incident repeated over time that these respondents found meaningful, but the overall kind of relationship formed by the specific experiences. Friesen and Young (1985) encountered another problem. They found that a theoretical orientation was needed to make their data comprehensible and amendable to classification. So they began with the theoretical framework and used
that to build the categories of their data. In yet another context, Borgen and Amundson (1984) categorized critical incident data by using emotional shifts as an indicator of change, looking for points in the stories where respondents reported new or different emotions.

(c) The Level of Generality. The headings and subheadings under which the data are reported establish the level of generality. Flanagan (1954) provided a useful list of practical, commonsense considerations to aid in deciding on headings. Thinking it of greater significance, however, to decide the level of specificity-generality on the basis of maximizing richness and distinctiveness of categories, I drew upon newer work in cognitive psychology on natural categories and prototypes. These are commended to the reader's attention (Cantor & Mischel, 1979; McCloskey & Glucksberg, 1978; Rosch, 1978). Using these newer principles of categorization is more effective than following Flanagan's (1954) guidelines. In my studies, for example, it did not appear necessary to have the same number of incidents in each category as Flanagan suggested. Forcing the data into superficially "equal" categories would have distorted them in other, more significant ways.

The same-sex bonds study yielded three levels of categorization: categories, subcategories, and facets. When the basic categories and subcategories had been arrived at, as described above, something still seemed to be missing. As I struggled with the data, it became apparent that three facets cut across all of the categories: the characteristics of circumstances, of persons and of relationships. Once these facets were identified, the whole system became clearer. The basic categories (e.g., activity-based bonds, support in time of need and deep sharing during grief/trauma) clustered naturally into circumstance, person and relationship-centred bonds. The whole system then showed a progression from more superficial bonds to deeper and stronger ties. There was an intuitive sense of "rightness" when this categorization system was arrived at. It was as if I had discovered some underlying structure and not simply imposed an artificial order. Although highly subjective, this description illustrates something of what the process of categorization is like. It is important to continue working with the category system until a point such as this is reached—a kind of "aha" response.

When the categorization is complete, independent judges are asked to sort the incidents into the categories, to see whether the categories can be replicated. There is no established criterion for the level of agreement necessary, but Andersson and Nilsson (1964) suggest that it is acceptable if independent raters can correctly classify 75% to 85% of the incidents into the categories and 60% to 70% into the subcategories. It is also advisable to compare the category system with other categorizations of the activity available in the scholarly literature.
5. Reporting the Findings

The report should use the vivid and evocative description necessary for qualitative research (Smith, 1981). The categories and subcategories need to be given self-explanatory titles. Simplicity, brevity and clarity are essential here. Descriptions of categories should be rich, though not lengthy, vividly conveying a picture of the kind of incidents included in the category. When writing the category descriptions, it is helpful to focus on a prototypical incident, for the more peripheral incidents do not illustrate the category as well as do the prototypical ones. Example incidents in the respondents’ own words make the writing more evocative. The descriptions also must enhance the distinctiveness of the categories.

The amount and type of information provided depends on the purpose of the report. For a journal article, Flanagan (1978) simply listed the categories and gave a brief description of each. Borgen and Amundson (1984) were more selective, highlighting the category heading and using direct quotes from respondents to describe the category. Level of detail provided varies. Subcategories may or may not be reported, as appropriate. A job analysis for industry, for example, requires greater detail than does a research report for a scientific journal. It is important that all significant aspects of the activities be represented because complete content coverage is central to the critical incident technique.

The report should clearly stipulate the limitations of the method (Flanagan, 1954). With respect to the limitations, some data on reliability and validity are available. Andersson and Nilsson (1964) found the critical incident technique to be valid in representing the content domain, so much so that other methods of assessing the same domain added no new information. After two-thirds of the incidents had been classified, 95% of the content categories appeared. In addition, the subcategories were found to be stable. The number and structure of the incidents were affected only slightly by different methods of data collection and by different interviewers. They concluded that the method is both reliable and valid.

STRENGTHS AND APPLICATIONS IN COUNSELLING

The critical incident methodology is highly flexible. It can be used to study a wide range of phenomena, for example, relationships, decision-making, self-actualization, vocational choice, and group process. It can be modified to collect data on factual happenings (rather than restricting its use to “critical” incidents), and on qualities or attributes; to use prototypes to span the various levels of the aim or attribute (low, medium, high); and critical or factual incidents to explore differences or turning points.
The applications of the critical incident methods are several in the present context of counselling psychology. The technique can be used for foundational and exploratory work, opening and clarifying a new domain for further research. Borgen and Amundson’s (1984) work on unemployment, using a modified critical incident methodology is an example of this kind of work, as is Friesen and Young’s (1985) work on parental influence on children’s vocational choices and Woolsey’s (1985) work on same-sex bonds. Bronfenbrenner (1979) stated that the chief obstacle to an adequate theory of development is the lack of a fundamental taxonomy of molar activities; it is hoped that these critical incident studies will contribute to such a taxonomy.

Critical incident studies are particularly useful in the early stages of research because they generate both exploratory information and theory or model-building. As such, they belong to the discovery rather than to the verification stage of research (Rice & Greenberg, 1984). Thus the critical incident method was used in the first phase of the same-sex bonds research both to generate a theoretical model and also to serve as a basis for writing questionnaire items for test construction. Later stages of this research will use experimental methods of research to verify and to refine the model. The critical incident method can be used in other areas in the way that task analysis has been utilized to explicate the processes of psychotherapy (Rice & Greenberg, 1984). The critical incident technique has also been used successfully for criterion development in industrial psychology and therefore has potential for criterion development in counselling process and career development research. In addition, the critical incident method is entirely consistent with the skills, experience and values of counselling psychology practitioners (Woolsey, in press) and thus can inspire counsellors with enthusiasm for research. Finally, and perhaps most important, if counsellors learn to use qualitative methods like the critical incident technique, this expertise will contribute to the development of a unique methodology for the discipline of counselling.
References


*About the Author*

Lorette K. Woolsey is a faculty member in the Department of Counselling Psychology at the University of British Columbia. Her professional and research interests include the psychology of women. She is currently conducting research on same-sex friendship and kinship relationships.

Requests for reprints may be addressed to Lorette K. Woolsey, Department of Counselling Psychology, Faculty of Education, University of British Columbia, 5780 Toronto Road, Vancouver, B.C. V6T 1L2.

*Acknowledgements*

An earlier version of this paper was presented at the Counsellor Educators’ meeting at the annual meeting of the Canadian Guidance and Counselling Association, Fredricton, N.B., June 1983.

Preparation of this article was supported in part by the Social Sciences and Humanities Research Council of Canada, Grant No. 410-83-0707 and by a grant from the Canadian Research Institute for the Advancement of Women.

The assistance of Dr. N. Amundson, Marjorie Sinei, and three anonymous reviewers is gratefully acknowledged.