

Journal of Social & Economic Studies,

1981, 9 (2) pp. 221-231.

Mahfooz A. Ansari* and Howard Baumgartel**

The Critical Incident Technique: Description and Current Uses¹

The critical incident technique² was developed by John C. Flanagan and his various collaborators beginning in the summer of 1941. The basic purpose of this development was to improve dramatically the existing methods of describing and evaluating job performance for a variety of purposes such as job description, performance appraisal and the design of more functional training programs in the United States Army Air Force during World War II. The technique became firmly established in the field of industrial and organizational psychology in the year 1954 with the publication in *Psychological Bulletin* of Flanagan's paper, "The Critical Incident Technique." The roots of the technique, according to Flanagan, can be traced back to the writings of Francis Galton and to later developments such as time sampling studies of recreational activities, controlled observation tests, and anecdotal records.

The critical incident technique essentially aims at collecting reports of behavior which are "critical" in the sense that they make a significant difference between effective and ineffective performance in the observed work situation. For example, these incidents must represent actual observations of on-the-job behavior. An incident is "any observable human activity that is self-sufficiently complete in itself to permit inferences and predictions to be made about the person performing the act" (Flanagan, 1954, p. 327). "The incident is acceptable as a critical one only if in the observer's judgment it relates to an important aspect of the work and includes behavior which is outstandingly effective or is ineffective with respect

* Lecturer in the Department of Humanities and Social Sciences, Indian Institute of Technology Kanpur, Kanpur 208 016, INDIA.

** Professor of Psychology and Business Administration, University of Kansas, Lawrence, KS 66045, U. S. A.

to the specific situation" (Flanagan, 1949a, p. 42). Thus the procedure essentially establishes the critical requirements of a job or activity through direct observation by the participants in or supervisors of the job or activity. A critical job requirement is one "which is crucial in the sense that it has been responsible for outstandingly effective or definitely unsatisfactory of an important part of the job or activity in question" (Flanagan, 1949b, p. 420). The *outstandingly effective* performance is referred to as that which leads to especially effective adjustment to the job situation, while the *definitely unsatisfactory* behavior is viewed as inadequate adjustment to a specific situation because of ignorance or other factors (Flanagan, 1950).

Thus, in a nutshell, a critical incident is not an *evaluation* of the person. Rather it is an *observation* of an individual's on-the-job behavior--what happened, what action actually took place, and what were its consequences--in contrast to the various rating scales in use at the time.

The methodological procedures and measurement of the critical incident technique have been described in detail by Flanagan (1952, 1954, 1964). The procedures to be used in setting up the critical job requirements vary from program to program and depend on the specific factors in the local situation--e.g., the relative importance of the program, specific personal qualities, training and background of the supervisor, and the like. Thus for the readers and potential users of the technique, the general principles can be summarized as follows:

The first step is *defining the job and general aim of an activity or program*. It is very difficult to report that a person is either highly effective or highly ineffective in a particular activity or in a particular job situation unless one knows what a person is expected to accomplish. The definition should be very comprehensive. It is only possible through a systematic survey of an adequate sample of representative personnel. The general aim should be a summary statement obtained from the authorities in the field. Specifically, the aim should express in simple terms these objectives to which most people would have agreement.

Plans and specification is the second step. It is essential to give necessary and precise instructions to the observers about the crucial behavior in formulating a functional description of the job activity. For example, "We are carrying out an investigation regarding.... (specific activity should be cited). We believe you are specially the most qualified person to tell us about (specific activity)." Most situations require specification of an activity prior to data collection. The specification must include such information as the place, the persons, the conditions, and the activities. In a complex situation, practical examples should be provided to the observer concerned in order to enable him to decide in a more objective fashion

about whether or not a specific behavior should be observed and recorded; only those observers should be selected who have adequate information about or are familiar with the activity. Probably a supervisor is the most competent judge of the effectiveness of the workers' performance under his supervision.

The critical incident technique is frequently used to *collect the data* (i.e., the third step) on past observations which are reported from memory. It is desirable to include most recent incidents and to motivate the observers to make detailed observations and evaluations at the time the incident occurred.

How can the critical incidents be obtained? There are a number of ways of collecting the data. Among them four are most popular--interviews, group interviews, questionnaires, and record forms. While using the interviews, it is essential to tell the interviewees the sponsorship and purpose of the study. The purpose of the study should be cast in an informal manner as, for example, "We wish to find out what makes a teacher effective," or "We are trying to learn in detail just what successful performance as an executive includes." Sometimes group interviews are conducted to reduce the cost and save time. This method provides for a check on the data supplied by the interviewees. The introductory remarks are similar to the individual interviews. The size of the group which can be handled effectively varies with the situations. When the group is large, the group interviews are handled with a questionnaire procedure. The mailed questionnaire method has given results not very different from those obtained by interview methods. The final procedure, the best of all in the opinion of Flanagan (1964), is the record forms. Generally, there are two ways of recording incidents by this method. One is to record the details of incidents as they happen. This situation is quite similar to the interview procedures except that observation is delayed following the introductory remarks and presentation of questions until an incident is observed to happen. The second way is to record such incidents on forms by putting a check mark in the appropriate place. The form describes most of the possible types of incidents.

Before using any of the procedures discussed above, it is advisable to try out the questions on a small sample, i.e., pilot run. *What should be the size of the sample?* There is no simple answer to this question. Flanagan (1954) suggests that for simple activities 50 or 100 incidents are sufficient. But some complex activities may require several thousand incidents for an adequate statement of requirements,

How can the incidents be analyzed? The aim of *data analysis* is to summarize and describe the information in an efficient way to make the incidents

effective for any practical purposes. Generally, the critical incident technique involves three problems. First is the problem of *frame of reference*. At this point one has to start with a very simple question "Critical requirements for what?" That is, one has to decide whether the use of requirements will be related to selection, training, development, or any other aspects. If it is for selection purposes, for example, one has to formulate a set of headings related to psychological traits that are associated with or utilized in selection process. The second problem lies in the *category formulation*. Of course, this procedure is very subjective. No hard and fast rules are available. The quality of the categorization depends largely upon the skill and sophistications of the formulator. Flanagan (1954) suggests the users to sort a relatively small sample of incidents into piles that are related to the frame of reference selected, and then define these tentative categories and classify additional incidents into them. Next, one may modify the tentative categories until all the incidents have been classified. Finally, a re-examination of the categories in terms of the actual incidents is suggested. The last problem of data analysis is related to the *general behaviors*. This step determines the most appropriate level of specificity-generality to use in reporting the data. Flanagan (1954, 1964) has reported several considerations to take into account in establishing the headings for major areas and in stating critical requirements at the selected level of generality: (a) the headings should indicate a clear-cut and logical organization, (b) the headings should convey meanings in themselves, (c) the list of statements should be homogeneous, (d) the headings should all be of the same general magnitude, and (e) the list of headings should be comprehensive and cover all incidents having significant frequencies.

In the early 1960's, the critical incident technique was used to develop the behaviorally anchored rating scale--a significant turning point. The well-known scale of this type was developed by Smith and Kendall (1963). Dunnette (1966) has summarized their procedure in several steps. Several groups of head nurses discuss the use of evaluation in improving nursing performance. Still others (head nurses) provide information by mail. Each of the groups outlines the major qualities of successful nursing. The critical incidents are collected and classified to describe the examples of behavior related to each quality. The group also formulates statements defining high, low, and acceptable performance for each quality illustrated by incidents. Incidents were then judged by another group of head nurses on a scale ranging from 0.0 to 2.0 according to the most desired behavior of nursing. The scale finally disclosed six qualities: knowledge and judgment, conscientiousness, skill in human relationship, organizational ability, objectivity, and observational ability. As to the consistency of scale judgments,

Smith and Kendall reported reliabilities ranging over .97. In short, their procedure is a variant of critical incident methodology that requires appropriate organizational personnel to consider in detail the components of performance for the job in question and to define anchors for the performance continua in *specific behavioral terms*.

It should be noted that the Smith and Kendall (1963) scale, following the retranslation of expectations procedure, showed a fairly high average reliability coefficient and content validity in rating situations. Perhaps the most systematic studies in the reliability and validity of the critical incident technique were reported in 1964 by Andersson and Nillson. They applied this technique in analyzing the job of store managers in a Swedish grocery company. They obtained over 1800 incidents through interviews and questionnaires. The incidents were classified into 86 sub-categories, 17 categories and three areas--relation to customers, relation to personnel, and relation to store and its sales. Andersson and Nillson provided a methodological check for the critical incident technique by collecting the incidents from four categories of personnel--superior, store manager, assistant, and customer--which give a positive impression of this method. The recategorization system was used in order to ensure the high stability of the subcategory, i.e., reliability. The content analysis of training literature was compared with the analysis of questionnaire ratings to ensure the truthfulness of this technique. The findings clearly showed that the method covered the essential points in the job, i.e., content validity. Thus, they concluded that the information collected by this method is both reliable and valid.

Campbell, Dunnette, Arvey, and Hellervik (1973) studied behaviorally-based rating scales in developing the criterion measures. They obtained critical incidents from 537 department managers. Specifically, they were interested in whether such scales would yield less leniency and halo errors and whether they would exhibit significant convergent and discriminant validity. They found that such errors as halo and leniency were not severe for the method of scaled expectations but were rather pronounced for summated ratings. In order to assess the convergent and discriminant validity, they followed the multi-trait-multi-method approach. From this perspective then, they produced a 36 x 36 multi-traits (9 performance dimensions) multi-method (summated ratings *vs.* scaled expectations), multi-rater (store manager *vs.* assistant store managers) matrix. All the entries in the validity diagonal with their corresponding row and column entries in the validity diagonal were significantly different from zero at $\alpha < .001$, showing a high convergent validity. Discriminant validity could be indicated in two ways--(a) by comparing entries in the hetero-trait hetero-method triangles, and (b) by comparing the validity diagonal

entries (same trait but different methods) with the corresponding row and column entries in the hetero-trait-mono-method triangles. They suggested that the scales can serve as criterion against which to evaluate predictors for selection and promotion decisions in addition to performance appraisal and review system.

The usefulness of these scaling techniques has been questioned by some investigators. For example, Kay (1959) comments, ". . . Judges throughout actually were incapable of assessing the degree of likelihood of effective, average, and ineffective foreman doing that which was described in the critical incident" (p. 270). But at the same time supportive evidences are also available. For example, Baylie, Kujawski, and Young (1974) note, "In modified forms it is an important element in the development of behaviorally anchored rating scales and forced choice rating forms" (p. 4). A recent advocate of this technique is Levinson (1976) who suggests some specific behavioral data to be used in the critical incident process. These are: feedback data, promotion data, coaching data, and long term data. He asserts, "The critical incident process compels the superiors to face subordinates, a responsibility too many shirk" (p. 6). Recently, Smith (1976) again recommended the use of this technique "with personnel on whom a large number of incidents can be observed, as in the military" (p. 752). Korman (1977) also advocates that ". . . the best way to get an adequate measure of an individual's performance is to keep a record of these unusual or critical behaviors, both good or bad, so that a periodic recording of them on a person's record can provide a ready-made evaluation of performance" (p. 372).

In sum, behaviorally-based rating scales incorporated with the critical incident technique offer a number of advantages (Baylie, Kujawski, & Young, 1974): (a) The scales are developed through extensive participation by organizational personnel, the potential users of the scales. They apparently provide a beneficial learning experience for the personnel participating in the development program. (b) Generally three types of errors are commonly associated with the traditional graphic and other rating procedures of job analysis--halo, leniency, and central tendency. A halo error is the tendency to allow one characteristic of rating object to dominate over others. The tendency to rate all objects in a relatively favorable direction is referred to as a leniency error. The central tendency error suggests the tendency to rate all objects around the middle point of the continuum. As we noted a little earlier (Campbell, Dunnette, Arvey, & Hellervik, 1973), the critical incident approach to behaviorally based rating scales is less likely to suffer from these errors. (c) The scales serve to reveal rather than obscure the complex behavior which contribute to successful performance by individual employees on a particular

job. (d) They can be used for providing feedback about specific aspects of individual job performance and for communicating expectations of job performance to new employees. (e) They provide fairly high reliability and validity which give a positive impression of these scales and critical incident methodology.

It is worthy of mention that these days Flanagan is extending the use of critical incident technique in the area of quality of life too. He (1978) studied three age groups (30, 50, and 70) and reported predictors of life satisfaction in adulthood. His survey identified 15 types of critical incidents. Each of the age groups was asked how important each of the 15 dimensions was to their quality of life. These "importance" dimensions were then compared with ratings of how well the group felt their needs and wants were actually being met. The best predictors turned out to be in the area of material comforts (money) and health, followed by work and active recreation. Also significant were learning, expressing oneself creatively, close friends, understanding oneself, socializing and close relationships with spouse (except for 70-year old women). He notes, "It was clear that the things that people considered when asked to rate their overall quality of life were not identical to the items they rated as most important and for which they reported their needs as being well met an activity that lacks deep commitment and satisfaction may still be very important contributor to one's overall feeling of well-being" (p. 7).

The Current Uses⁴

Looking over the above description as well as the published literature on it, we noted that the technique has been used in a wide variety of contexts quite often with positive impression. However, descriptions of the current uses of this technique are not being reported in the professional literature.

In view of the lack of recently published literature about the critical incident technique, we planned and carried out a small mail survey to obtain additional information on this subject. The objective of this survey was to learn about where and how the technique has been used in recent years or is being used currently. For the mail survey, a true random sample of 50 members of Division 14 of the American Psychological Association was drawn from the 1978 APA Directory and a true random sample of 149 members of the list of 1978 OD Network members currently employed by business or industrial concerns was drawn. Personally addressed and signed letters accompanied by a short questionnaire were sent to each of the respondents. Forty per cent of the respondents answered the enquiry--46

per cent of the Division 14 sample and 34 per cent of the OD Network sample. A summary of the mail survey results is shown on Tables I and II.

Table I: Summary Results of Mail Survey About Critical Incident Technique

<i>Respondents Questionnaires:</i>		<i>Respondents:</i>		<i>Respondents:</i>		<i>Respondents:</i>	
<i>Mailed</i>	<i>Returned</i>	<i>Have Used</i>	<i>Have not Used</i>	<i>Fami-liar</i>	<i>Unfami-liar</i>	<i>Recently used or using</i>	
		<i>Per cent</i>		<i>Per cent</i>		<i>Per cent</i>	
<i>Div 14 APA</i>	50 23	47.8	52.2	65.2	34.8	39.1	
<i>OD Network</i>	149 51	15.7	84.3	37.3	62.7	17.7	

Table II: Respondent Report of Current Use By Area of Application

<i>Area</i>	<i>APA Members Per cent</i>	<i>OD Members Per cent</i>
<i>Performance Appraisal</i>	17.4	2.0
<i>Scaling</i>	04.4	3.9
<i>selection</i>	17.4	0.0
<i>Training</i>	04.4	3.9
<i>Conflict Resolution</i>	04.4	0.0
<i>Team-building Activities</i>	00.0	2.0
<i>No Response</i>	00.0	4.0
<i>Total</i>	48.0	15.8

Though the survey was based on a small number of respondents, it readily shows that over 65% of the Division 14 is familiar with the Flanagan's method. Of these about 48% have used this technique sometime in their professional careers, and most of these (39%) are making current use of this technique. Content analysis of the responses shows that those psychologists who are recently using this technique are mostly concerned with performance appraisal and selection (17.4% each): Some are engaged in incorporating this technique with behaviorally anchored rating scales in line with the Smith and Kendall scale and others with training issues (4.4% each). Yet others (4.4%) are expecting this technique to be helpful in conflict resolution in business organizations.

The responses of OD Network members are somewhat different. Most of the respondents (65.2%) are not familiar with the technique (a significant comment on OD specialists). Only about 16% of the members have used this technique in their present organizations. But, currently, about 18% of the members are using it, most of them for training purposes and in the development of rating scales. One respondent described that he is using this technique in team-building activities in his organization. He has found this technique viable thus far; however, it is possible that he was using something other than Flanagan's technique.

We also asked the respondents: "How do you feel about the results or success of using the critical incident method?" Interestingly, most of the respondents (psychologists = 30.4%; OD Network = 21.6%) reported that they have found it to be an excellent and successful tool. About 13% of the APA members are of the opinion that although this technique is time consuming, it is highly successful in collecting the specific behaviors of an individual on the job. Only 4.4% of the APA sample and 2.0% of OD Network members reported that they have not found favorable results in using this technique. But these respondents did not mention any specific reasons. It is interesting to note that some of the companies (11.8%) represented by OD Network members have been using the critical incident method over the years, but the respondents do not know whether this is the technique developed by John C. Flanagan or some other technique.

Comparing the responses of the APA and OD Network samples, one can safely conclude that Division 14 members of the APA are more aware of the technique and are more active in using it than are the OD Network members. It is quite natural to have such a trend in the results because many OD Network members are less involved in the traditional problems and methods of industrial and organizational psychology.

DISCUSSION AND CONCLUSIONS

Looking over the foregoing discussion, it appears that a great deal of attention was paid to the critical incident technique during the late 1940's and early 1960's. After 1954, the researchers have reported fewer studies using this technique which may indicate that it has not maintained the earlier pace of development--over 50% of the research literature was published before 1954. It is also evident from our review of the literature (not reported in detail here) and the questionnaire survey that there is little disagreement about the usefulness of this technique, although some few have found negative results. The main criticism of the technique-based scales lies in the fact that they require considerable time and commitment of

manpower for their preparation. Flanagan has suggested many alternatives to save time. For example, it is not always necessary to use the performance record forms. One could use group interviews as well as mail questionnaires to reduce complexity. Smith and Kendall (1963) have successfully used the "conference method" in obtaining the incidents.

The trends in the published literature show that people were more concerned with performance appraisal, selection, and training until 1963. A significant turning point came with the publication of Smith and Kendall's (1963) paper, which described the use of this technique by incorporating it with their behaviorally anchored scale. In addition to these developments, the technique is extending its usefulness in other areas such as team-building activities, conflict resolution, and studies of the quality of life. Thus the dream which Flanagan had in the 1950's seems to be coming true. Finally, it should be noted that the technique is being used in many organizations by professional experts but there is very little published research about these current issues. One might conclude that its use has become institutionalized and, hence, research interest has waned.

So far as future is concerned, we expect the technique to be further developed for the prediction and assessment of managerial success. The definition of executive success continues to be a perplexing problem (see Argyris, 1953). The main difficulty lies in the development of criterion measures (Stark, 1959; Ansari, 1979b). Thus, the extended use of the critical incident method may help to identify the behavior critical for successful and unsuccessful executives taking into account the newer contingency approaches to the understanding of organizational behavior (Ansari, 1979a).

REFERENCES

- Andersson, B.E. and Nilsson, S. G. 1964 "Studies in Reliability and Validity of the Critical Incident Technique", *Journal of Applied Psychology* **48**, 398-403.
- Ansari, M. A. 1979a *The Fit of Personal Orientation with Organizational Climate in Predicting Managerial Success: A Contingency Approach*, Unpublished Master's Thesis, University of Kansas, U. S. A.
- Ansari, M. A. 1979b *Some of the Determinants of Success on Executive Positions*, Unpublished Doctoral Dissertation, Patna University, India.
- Argyris, C. 1953 "Some Characteristics of Successful Executives" *Personnel Journal*, **32**, 50-55.

- Baylie, T. N., 1974 "Appraisal of People Resources", In D. Yoder and Kujawski, C. J. H. G and Young, D. M. Heneman (eds.) *ASPA Handbook of Industrial Relations*, The Bureau of National Affairs, Washington, D. C., U. S. A.
- Campbell, J. P., 1973 "Development and Evaluation of Behaviorally Based Rating Scales", *Journal of Applied Psychology*, **57** 15-22.
- Dunnette, M. D., Arvey, R. D., and Hellervik, L. V. 1966 *Personnel Selection and Placement*. Wadsworth Publishing Company, California.
- Flanagan, J. C. 1949a "A New Approach to Evaluating Personnel", *Personnel*, **26**, 35-42.
- Flanagan, J. C. 1949b "Critical Requirements: A New Approach to Employees Evaluation", *Personnel Psychology*, **2**, 4 19-425.
- Flanagan, J. C. 1950 "The Critical Requirements Approach to Educational Objectives", *School and Society*, **71**, 321-324.
- Flanagan, J. C., 1952 "Principles and Procedures in Evaluating Performance", *Personnel*, **28**, 373-386.
- Flanagan, J. C. 1954 "The Critical Incident Technique", *Psychological Bulletin*, **51**, 327-458.
- Flanagan, J. C. 1964 *The Critical Incident Approach to Personnel Selection*. American Institute for Research, Pittsburgh, U. S. A.
- Flanagan, J. C. 1978 "A Blueprint for Quality", *American Psychological Association Monitor*, **9**, 1-64.
- Kay, B. R. 1959 "The Use of Critical Incidents in a Forced-Choice Scale", *Journal of Applied Psychology*, **43**, 269-270
- Korman, A. K. 1977 *Organizational Behavior*, Prentice Hall, New Jersey U, S. A.
- Levinson, H. 1976 "Appraisal of *What* Performance?" *Harvard Business Review*, **54**, 30-48.
- Smith, P. C. 1976 "Behavior, Results, and organizationl Effectiveness", in M. D. Dunnette (ed.) *Handbook of Industrial and Organizational Psychology*, Rand McNally, Chicago, U.S. A.
- Smith, P. C., and Kendall, L.M. ° 1963 "Retranslation of Expectations : An Approach to the Construction or Unambigucus Anchors for Rating Scales", *Journal of Applied PsycholoRy*, **47**, 149-153.
- Stark, S. 1959 "Research Criteria of Executive Success", *Journal of the University of Chicago*, **32**, 1-14.

FOOT NOTES

1. The authors thankfully acknowledge the assistance of the Department of Psychology and the School of Business, the University of Kansas, for the financial support and resources which made this study possible. In addition, we thank those members of Division 14, APA and the OD Network who completed our research questionnaire.
2. This section is based largely on Flanagan's paper (1954).
3. A complete bibliography is available on request.