Community Influences on Mental Health Program Evaluation[1]

David R. Atkinson
University of Michigan

Research in community psychology has demonstrated that community environ-ments can play a key role in defining the operations of human service institutions. This study questioned whether significant community groups might indirectly influence the program evaluation record-keeping of mental health outpatient programs to reflect contact with problem areas considered important by those groups. An inventory of program evaluation materials was conducted in three types of mental health programs. The ranked frequencies of record completion for 10 designated problem areas were correlated with staff rankings of im-portance for agency work with problems identified by key community groups. Results indicate that the more important the community demand is to the staff, the more systematically complete will be data sources which document agency performance for that demand. Given the general lack of comprehensive develop-ment of program evaluation in many mental health programs, the results suggest that community interest groups may play a major role in determining what types of information on mental health problems are made available to the public. The resultant lack of balance and scope in these data inevitably limit the range of community responses for program improvement.

This study questioned whether community interest groups might influence the completeness of data available for program evaluation. Following Caplan and Nelson's (1973) thesis, the nature and inferential potential of available data may have a direct influence on the action alternatives for community interventions intended to mitigate or prevent the development of psychosocial problems.

Two major theories have been proposed to explain the systematic bias of previous program evaluation efforts. One is the politicosocial history of evaluation as a field; the other is the adaptation of organizations to their environments. Several researchers have questioned the objective neutrality of social intervention programs which must operate in politically pluralistic environments. Campbell (1969) suggested that such programs may only selectively transmit their operational information to outside groups. Self-preservation interests demand the selection of politically helpful information for public exposure in a potentially hostile environment. Ryan (1971) extended that approach to argue that community intervention programs, and the social forces which support them, often have a vested interest in avoiding information on how social influence processes are involved in the emergence and maintenance of individuals' problems.

Caplan and Nelson (1973) have argued that historical tradition may play a key role in determining the focus of academic or evaluative research. Traditional preferences for individual rather than systems variables led inexorably to documentation of individuals' rather than systems' complicity in the community processes within which these problems develop.

Weiss (1973) has identified how the evaluation industry itself has been shaped by political forces and competitive pressures to focus its efforts on some of the least potentially useful information topics. Evaluative research has often been used by political actors to lend scientific legitimacy to the exposure of inadequacies in politically controversial, experimental programs, while it has systematically ignored many politically entrenched traditional programs.

These arguments imply that community interest groups, especially those segments responsible for sanctioning social interventions, help determine what kinds of information are made public on the operations of social programs. To the extent that the resulting data are used as evidence of the prevalence of problems in community planning, the gaps in these data can limit the range of action alternatives. Unfortunately, research on community interventions has largely failed to document the nature of community influence on the selection of specific types of program evaluation data in specific social programs.

Regarding the second major theory, organizational researchers have shown that organizations and their environments have the capacity to modify each other. Dill (1958) examined comparative environmental influences on organizations' internal communications, goals, structures, and performance measures. Pfeffer (1972; 1973) found that organizations best able to match key demands being placed upon them by outside groups performed better than other organizations less effective in matching such demands. This match was produced by composing the corporate board with individuals best qualified to link it to key regulatory, financial, and other outside groups that could play essential roles in

the organization's growth. Pfeffer proposed that organizations best able to co-opt significant environmental sectors would greatly increase their survival capacity, despite having to modify their internal structures somewhat to meet environmental requests.

If organizational viability is partially dependent upon meeting environmental conditions, then the organization's internal data-collection patterns might also be influenced by community characteristics. This phenomenon would contradict common professional assumptions that mental health programs define their data-collection operations solely according to independent, professional judgments of what information would be rationally useful in studying their problems.

The rationale for this study is that community psychology faces the problem of explaining the specific reciprocal interactions between community interest group influence on mental health programs and the collection of certain data on how program interventions are actually working. What types of demands for agency work, and from which community groups, result in an agency's collection of specific classes of program evaluation data? Before answering such questions, it would first be necessary to show what differences in program evaluation data collection exist among these programs, and whether such differences are linked to discrete requests for program performance by environmental groups seen as significant by each program.

To address this problem, the current study examined three questions: (a) Do agencies supported by or serving different segments of the community differ in the completeness of recording input versus outcome data? (b) Do these agencies vary systematically in the completeness of data on clients *versus* data on community actions and resources? (c) Are problem areas rated most important for key community groups the areas most likely to be the most completely recorded?

METHODS

Two basic steps comprised the research program. First, a comprehensive inventory was taken of the types and completeness of program evaluation data recorded in three mental health agencies. Data considered relevant included any items recorded on the nature of the clients, their problems, and the attendant community background of such problems. Descriptions of staff activities were omitted. Second, staff members from each agency ranked the importance of perceived demands for program actions being placed upon their agencies by significant community groups. These demands were linked to specific classes of data being recorded.

Setting

The study was done in a Midwestern city of over 100,000 population. Three major types of mental health outpatient programs serve its population: (a) governmentally sponsored community mental health centers (designated Public type); (b) privately sponsored, nonprofit, charity-financed, traditional professional family counseling agencies (Voluntary type); and (c) privately sponsored, charity-financed, nontraditional, volunteer-staffed, minority advocacy/counseling programs (Advocacy type). One agency was selected from each major type on the criteria of similar size (10–20 staff, 300–500 clients), program diversification, and acceptance by the community, Since basically different community segments supported each type of program, it was hypothesized that each type would concentrate on different community problems, hence recording data on different areas.

Measurement of Record Completion

A Task-Data Inventory (Atkinson, 1975) was developed from a review of the program evaluation literature. It was designed to quantify the completeness of recording for data in 10 major categories relevant for program interventions. Five categories encompassed data about clients: client biographical details, descriptions of problem situations, the interpersonal field, client living conditions, and problem development factors. The second group of five categories represents data about the community: other organizations' actions, community resources available, relevant social pressures from the community, descriptions of related community problems, and current systemic changes in the community.

This inventory contained 99 subcategories of potential information which might be gathered by a mental health program. For example, within the living situation category, there were subcategories for housing status, work schedule, transportation problems, etc. For each subcategory respondents were asked: who supplied the information, who recorded it, how often it was updated, and the recorder's discretion in deciding how frequently it would be recorded. Data uniformly gathered at scheduled points were classified as systematically recorded. Data recorded only at each individual worker's discretion were classified as idiosyncratically recorded.

Respondents were interviewed on each subcategory of the Task-Data Inventory. They described in their own words all types of operational data routinely or casually recorded about all phases of their agency's direct services to clients, to related community organizations, and for administrative purposes. Written interview transcripts were reviewed by two independent raters who assigned a yes/no score for recording each of the 99 subcategories separately for both systematically or idiosyncratically recorded data.

A completion percentage score was assigned to each of the 10 data categories for each respondent separately for systematically and idiosyncratically recorded data. Scores were computed by dividing the actual number of areas recorded (e.g., systematically) by the total number of subcategories possible for that category. Mean category completion scores over all respondents for agencies or program elements (e.g., direct services workers) within agencies were computed using percentages for categories. For some statistical operations, mean category completion scores within each agency were ranked from lowest to highest rate of completion.

Interrater reliability was sufficiently high, with Pearson r correlations of .98 or higher for paired yes/no judgments on each item. Reliability was equally high in all categories. Respondents' objectivity appeared high also. It was possible to inspect client-centered and community-centered data records in two of the agencies. Inspection confirmed a written case record for each subcategory of data reported by each respondent.

Measurement of Community Demands

A second semistructured interview was developed to gather agency workers' perceptions of the importance of demands being made upon their agency by key community interest groups. Following Dill's (1958) research, five potentially important environmental sectors were identified, including *suppliers* of resources (money, personnel, equipment); *sanctioners* of community support (professional groups, community service organizations); *regulatory* groups reviewing agency performance (governmental agencies, licensing bodies); *competitors* experiencing conflicts over territorial rights (other human services agencies); and *customers* requesting services (clients and referral groups).

Respondents were asked to list specific agency activities which had been requested by key community groups they identified from each of the five major sectors. Separately for each agency, similar activities were pooled into one overall list of agency activities. Every activity identified by two or more respondents was included in the final list. Examples include organizing interagency planning conferences, providing case consultations to other agencies, providing advocacy services for clients, etc.

A questionnaire containing this pooled list (a different list for each agency) of activities was given to each respondent. They ranked each activity from low to high for its importance to the agency's actual survival and good health. Pooling these rankings within agencies yielded a rank order of perceived importance for task achievement in response to community demand.

Workers were asked to identify which program evaluation data subcategories, if any, were being used to document their agency's response for each community demand they identified. In each agency, the workers named only

one data category for each agency demand, and these responses were pooled to identify a single data category which documented performance on each community demand for each separate agency. These pairings allowed for measurement of rate of data collection in response to each identified demand.

Respondents

Contact persons within each agency explained its role structures. They developed a list of potential employee respondents at all functional levels, including direct work with clients, supervision of clinical staff, general administration, and board membership. A random sample of two staff members was chosen from each functional level within each agency. These selection levels included differences inherent to different work shifts as described by agency contact persons. At some levels, particularly administrative roles, random sampling was impossible when only one or two persons worked at a given level.

Thus, the final number of respondents for each agency was a function of the degree of differentiation of roles. For example, the Public agency, having the greatest differentiation, required a larger number of respondents to ensure coverage of each appropriate agency level: 34 respondents were interviewed on the Task-Data Inventory, including 17 from the Public agency, 8 at the Voluntary agency, and 9 at the Advocacy agency.

A second random selection was made for the second series of interviews on Community Demands. However, it was not possible to interview only new respondents for the second interviews and still contact all relevant organizational levels, due to the small staff size of each agency. Hence, only 13 of the 34 second interview respondents were new respondents. A problem of dependency could have arisen between the results of the two interviews. Perceptions of the agency's data-recording patterns could have influenced the subsequent interviews on perceptions of community demands. However, that dependency appears highly unlikely because individual workers' estimates of which data categories would be most completely recorded did not correlate with actual agency data-recording practices.

RESULTS

The results were grouped into three major areas, the relationships between community segment served and: (a) input and outcome data; (b) data about client and data about community; (c) priority of environment task demand and record completeness.

For the first area, input data were defined as all data entered on each data subcategory the first time it was recorded. Similarly, outcome data were defined

Table I. Pooled Staff Perceptions of Percentage Completion of Input Vs. Outcome Data Categories

Recorded data categories	Public agency		Voluntary agency		Advocacy agency	
	Input	Outcome	Input	Outcome	Input	Outcome
Client biographics	.57	.14	.70	.48	.06	.00
Problem characteristics	.31	.16	.52	.45	.38	.22
Interpersonal field	.19	.11	.31	.16	.11	.04
Living situation	.19	.03	.27	.16	.02	.00
Problem development	.28	.03	.34	.19	.03	.00
Organizational actions	.11	.15	.19	.13	.40	.30
Community resources	.16	.16	.17	.05	.79	.69
Community pressures	.05	.01	.06	.02	.70	.61
Community problem statistics	.14	.09	.17	.05	.21	.14
Community dynamics	.11	.03	.21	.21	.61	.43
$df = 9$	$t = 2.77^a$		$t = 5.21^a$		$t = 5.34^a$	

[a] $p < .01$, one-tailed t tests.

as any subsequent data documenting the later status of previously recorded data, such as change in problem situation, or the appearance of a new resource person.

All three agencies recorded significantly more input than outcome subcategories across the Task-Data Inventory. Results shown in Table I give the pooled percentages of subcategories completed within each major category for each agency.

The sector served by each agency appeared to have only minor influence upon differences in completion of input *versus* outcome data. Both the Voluntary and Advocacy agencies reported an elevated completion rate for outcome data in two or three categories, while no elevated rates were reported by the Public agency.

The differences in record completion were actually even greater than indicated in Table I. All respondents who collected outcome data reported that recorded outcome data were never as complete as the input data already collected. Time pressures, fear of further intrusion into client privacy, and lack of cooperation by other agencies were cited as reasons for their judgment of much lower quality in the recorded outcome data.

Finally, the agencies' own workers were the primary sources for almost all outcome data subcategories. In their words, these data were ordinarily collected only in response to high priority demands upon the agency. Even when outcome data were gathered from interviews with clients, family members, other agencies, etc., the workers recorded their own information summaries. The general lack of outside sources for outcome data and the low quality rating of such data cast serious doubts on the objective reliability of outcome data in these agencies.

For the second area, it was assumed that the professional orientation of both the Public and Voluntary agencies would lead to a primary focus on collecting data about individuals instead of data about the community. The reasoning follows Caplan and Nelson's (1973) thesis for problem bias caused by the political characteristics of their supporting groups. Because the Advocacy agency is a counterprofessional agency, the reverse relationship was expected, a concentration on community-centered data. Staff members in each agency classified

Table II. Means of Staff Perceptions for Proportion of Data Categories Systematically Completed, Grouped by Data Focus

Agency type	Client data	Community data	t Score (df = 4)
Public	.3088	.1154	2.68[a]
Voluntary	.4284	.1605	3.16[a]
Advocacy	.1207	.5424	−3.37[b]

[a]$p < .05$.
[b]$p < .025$

Table III. Rank Order Correlations Between Environmental Task Importance and Data Category Completion

	Data recorded ($df = 9$)	
Agency	Systematically	Idiosyncratically
Public	$r = .66^a$	$r = -.41$
Voluntary	$r = .68^a$	$r = -.68^a$
Advocacy	$r = .81^b$	$r = -.67^a$

[a] $p < .025$.
[b] $p < .01$.

each data category as either client or community data. Each staff member named the same five categories with a primary focus on clients, the other five with a community focus.

The Public and Voluntary agencies concentrated much harder on gathering client data, whereas the Advocacy agency focused on community data. Table II presents the means for pooled staff perceptions of record completion for the two groups of five categories for each agency. The Voluntary agency recorded significantly higher percentages of client data than did the Advocacy agency, $t = 2.93$, $p < .025$, $df = 4$. Conversely, the Advocacy agency recorded significantly more community data than did either the Public agency, $t = 3.96$, $p < .01$, $df = 4$, or the Voluntary agency, $t = 3.50$, $p < .01$, $df = 4$.

The third study area assumed that the more important the demand made by an environmental group for specific tasks, the more data will be recorded to document agency involvement with those tasks. As external demand lessens, the quantity of systematically recorded operations data relevant to that demand will decrease.

For each agency, the mean rank order of completion for systematically recorded data in the 10 data categories was compared with the pooled rank order of importance for the 10 tasks demanded by community groups. Each data category documented agency work on one of those tasks. The Spearman-rho test was used to compare whether the order of record completion for the 10 categories corresponded to the order of perceived importance of record categories for meeting community demands.

Positive correlations were found between these two rank orders for each agency (Table III).

Within agency organization levels in the Public and Voluntary agencies, 14 of the 16 correlations between rank of community demand and rank of data completion were significant at the .05 level or above.

The results indicate that agency workers were most familiar with the types of community demands and agency record-keeping which related to their specific job duties, e.g., documenting individual clients' problems and requests

for treatment. Workers at all agency levels tended to know very little about either specific community demands or data-recording activities relevant to other divisions within their agency.

A reverse relationship appeared when task priorities were compared to completion levels of data which were reported as idiosyncratically recorded (i.e., recorded at each individual worker's discretion and not summarized for publication). Within agencies, ranks were assigned to the results of the Task-Data Inventory for the percent of idiosyncratically recorded data subcategories. Correlations between importance ranks for environmental tasks and idiosyncratic recording ranks were negative (Table III). The less important the community demand for work to be routinely recorded in a given task area, the more likely for data on that work to be nonsystematically recorded. Individual workers viewed such recording as primarily for their own personal evaluation purposes, not for public evaluation.

The longevity of staff did not appear to influence the amount of record completion. Whether a category of data was systematically recorded did not appear to influence which categories were idiosyncratically recorded. No single category could be rated as both systematically and idiosyncratically recorded. Correlations among these variables did not approach statistical significance. Worker reports of actual interest group demands for agency activity did not represent their own preferences for activity priorities, which preferences, they noted, would be substantially different. Most stated also that they were not recording the kinds of data which they might prefer or view as most useful.

The functional level of each worker appeared to account for most of the difference in kinds of data recorded and knowledge of external community demands. Workers at the direct service levels recorded client-oriented data and knew the most about the demands being made by clients, their families, and by workers in allied human services agencies. Supervisory level staff recorded little client data and knew the most about interorganizational issues and program evaluation problems, probably because their roles required them to maintain almost daily contact with supervisory staff members in many agencies. Administrative personnel and board members limited their primary contacts to regulatory, licensing, financing, and sanctioning groups, and hence recorded most of that data. Workers often were unaware as to precisely what data recording or community demands were occurring at different levels.

Workers in each agency were asked to rate the degree to which community interest groups made routine influence in both their everyday agency activities and the development of policy. The amount of influence by such outside groups into both categories was reported as very low (less than 25% of the time) by both the Public and Voluntary agencies. The Advocacy agency reported a higher level of community influence on everyday activities, due to its participatory model of governing structure, but rated the amount of environmental input as only moderate (25–50% of the time) on policy matters.

Despite both the low frequency and intensity of community interest group demands as reported by the workers, these influences still appear directly related to the types of program evaluation data collected. If that relationship were to hold under conditions of more intense, frequent, and broad influence, one would anticipate a higher and more comprehensive rate of data collection.

DISCUSSION

This study reported that three types of outpatient mental health facilities record more input than outcome data, that they tend to concentrate on either client-centered or community-centered categories of data depending upon the sector of the community they serve, and that significant correlations exist between community demands upon agencies and the degree of completion of certain agency program evaluation records.

These empirical results correspond with interview comments. Workers in the Public and Voluntary agencies commented on the large amount of time spent on documenting the individual problems of clients as compared to time spent gathering background data on community processes. They reported that almost all the pressures from regulatory agencies, financing groups, and community committees encouraged them to change the recipients of services rather than to change social factors.

Workers from the Advocacy agency, however, reported an opposite perception. Because the Advocacy agency was founded as a minority group's alternative to traditional counseling programs, its primary community supporters actively encouraged it to try to change social, organizational, and community influences upon individuals. By documenting the existence of such forces and the problems they produced for individual clients, the Advocacy agency was encouraged to build even broader public support for alternative approaches to providing services.

These results tend to confirm Campbell's (1969) hypothesis that social political climates may have a differential impact on the kinds of evaluation data produced in social intervention programs. The inferential potential (whether individual or system blaming) of program evaluation data appears to depend in part on the kinds of demands made by community groups for interventions.

Partly because of the scientific aura of social program evaluation, the inferential potential of these tacitly selected data is often left unchallenged. Because the data produced can be linked to direct influence attempts by specific community groups, there is serious danger that such data may help maintain the self-interests of these groups. For example, data showing a steady increase in the number of clients seen per dollar spent suggest favorable conclusions about the governmental officials whose public role is to monitor spending in

these programs. Unfortuantely, such data alone say nothing about the quality or adequacy of services rendered.

As Mulder (1971) has pointed out, unequal access to key operational data may tend to increase the power of recognized authorities (i.e., the mental health professionals and program administrators) over other community participants when cooperative planning is mandated. Comprehensive evaluative research on programs' performance, conducted either internally or externally, may be severely hampered by a lack of community encouragement broad enough to stimulate a comprehensive system of program actions. It would be unreasonable to expect social service agencies like these not to be responsive to the needs of important community groups upon whom they must depend for their everyday existence. These interviews suggested, however, that there is wide room for additional community influences on these programs due to the overall low level of environmental demands. Additional planned community requests could reshape the nature of their program-evaluation efforts to include presently ignored areas. The continuation of a low level of community demands would be expected to prolong the inhibition of community participation, as program evaluation fails to provide useful alternative data for decision-making.

REFERENCES

Atkinson, D. *Community influences on the generation of program evaluation data in mental health services.* Unpublished doctoral dissertation, University of Michigan, Ann Arbor, 1975.

Campbell, D. Reforms as experiments. *American Psychologist, 1969, 24,* 409-429.

Caplan, N. & Nelson, S. On being useful: the nature and consequences of psychological research on social problems. *American Psychologist, 1973, 28,* 199-211.

Dill, W. Environment as an influence on managerial autonomy. *Administrative Science Quarterly,* 1958, *2,* 409-433.

Mulder, M. Power equalization through participation? *Administrative Science Quarterly,* 1971, *16,* 21-38.

Pfeffer, J. Size and composition of corporate boards of directors, the organization and its environment. *Administrative Science Quarterly,* 1972, *17,* 218-228.

Pfeffer, J. Size and composition, and function of hospitals boards of directors: a study of the organization-environment linkage. *Administrative Science Quarterly,* 1973, *18,* 349-363.

Ryan, W. *Blaming the victim.* New York: Random House, 1971.

Weiss, C. Where politics and evaluation meet. *Evaluation, 1973, 1,* 37-45.