



September 1979

Baseline Evaluation: Evaluating Consistency between Federal Standards and Local Provisions

James R. Seaberg
Virginia Commonwealth University

David F. Gillespie
Washington University

Follow this and additional works at: <https://scholarworks.wmich.edu/jssw>

 Part of the [Social Work Commons](#)

Recommended Citation

Seaberg, James R. and Gillespie, David F. (1979) "Baseline Evaluation: Evaluating Consistency between Federal Standards and Local Provisions," *The Journal of Sociology & Social Welfare*: Vol. 6 : Iss. 5 , Article 7.
Available at: <https://scholarworks.wmich.edu/jssw/vol6/iss5/7>

This Article is brought to you for free and open access by the Social Work at ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.



BASELINE EVALUATION: EVALUATING CONSISTENCY
BETWEEN FEDERAL STANDARDS AND LOCAL PROVISIONS*

James R. Seaberg, Ph.D.

School of Social Work
Virginia Commonwealth University

David F. Gillespie, Ph.D.
George Warren Brown School of Social Work
Washington University

ABSTRACT

Baseline evaluation is a form of evaluation procedure in which objectives related to several dimensions of response to a social problem are set externally in the form of federal standards. The standards form the baseline against which local provisions can be compared. The case example giving rise to the development and field testing of the baseline evaluation procedure was new Federal Standards on the Prevention, Identification and Treatment of Child Abuse and Neglect. The baseline evaluation methodology is described and problems encountered are discussed.

Despite considerable variation in the conceptualization of critical elements and procedural formats, most program evaluation methodologies begin with a specification of program objectives by the persons responsible for the implementation of the program (Suchman, 1967; Weiss, 1972; Jones and Borgatta, 1972; Wholey, et al., 1975). It is also typical that the program which is to be evaluated has a limited scope imposed upon it, either by the nature of the target problem or the resources of the program. Evaluation seldom has been attempted where the program objectives have been set externally and a posteriori

*The research on which this material is based was supported by grant #90-C-430 from the National Center on Child Abuse and Neglect, Children's Bureau, Office of Child Development, Office of Human Development, DHEW.

by a governmental agency in the form of standards for responding to a particular social problem or class of social problems. Even less often have evaluators dealt with standards so broad in scope as to include multiple human service systems such as legal, health, social, law enforcement, education and so on. There is good reason for restricting evaluations to particular subsystems and specific objectives. But there is also a need to develop broad-based methodologies for assessing subsystems in relation to external societal standards.

This paper describes a recent experience in the development and field testing of an evaluation procedure for use with new Federal Standards on the Prevention, Identification and Treatment of Child Abuse and Neglect (1976). The purpose of developing this procedure was to provide responsible organizations or advocacy groups at both the state and local level with a mechanism for assessing how their provisions for responding to the problems of child abuse and neglect compared with the provisions prescribed in the Federal Standards. The imposition of externally generated standards on already operative programs raised a new set of evaluation issues and required a new form of evaluation methodology. The development of this methodology, called "baseline evaluation," resolves certain issues and leaves others for future refinement. This label reflects the fact that the standards, which were generated through a process of expert consensus (as most standards are), constitute a baseline or absolute minimum of provisions for an optimal response to the problem of concern, in this case, child abuse and neglect.

Baseline Evaluation

The essential question for a baseline evaluation is simply: Are all of the provisions prescribed as necessary by the standards in place? As will be shown, this seemingly simple question becomes increasingly complex as the scope of the standards is broadened both in terms of the number of human service systems involved and the detail of the items which compose the standards.

Placed in contextual perspective, the notion of baseline evaluation precedes implementation and effectiveness types of evaluation. But it also provides a foundation for these other subsequent forms of evaluation. That is, are the baseline provisions being adequately implemented so the desired effect might be attributed to them, and if the baseline provisions are adequately implemented, does the desired effect result? This paper does not deal with the latter two forms of evaluation, but it is important to recognize that, ultimately, baseline

evaluation is only the first of two prerequisite stages in the assessment of the program effectiveness where the program is prescribed by a set of standards.

There has been only one other attempt at developing an evaluation procedure to be used with standards of a national perspective (Standards for Community Agencies, Serving Persons with Mental Retardation and Other Developmental Disabilities, 1973), but there seems to be a trend toward the use of standards, particularly as an aspect of the federal response to various social problems. Assuming this trend continues, the interest in developing evaluation procedures for use with various types of standards will probably increase.

As mentioned above, the case example upon which this discussion is based involved the development and field testing of a baseline evaluation procedure (Seaberg and Gillespie, 1977) for use with new Federal Standards on the Prevention, Identification and Treatment of Child Abuse and Neglect (1976).^{*} These Standards were developed through initial input from a panel of federal government and national experts on the problems of child abuse and neglect. Later, the Standards were circulated to departments of social services in all 50 states and to national experts for review and comment. The Standards were then revised to their current format and substance, Federal Standards for Child Abuse and Neglect Prevention and Treatment Programs and Projects (1978).

The Standards reported in the 1976 draft version were presented in Sections: (1) prevention; (2) reporting; (3) state responsibilities; (4) investigation of reports by local agencies; (5) local agency administrative responsibilities; (6) treatment; (7) prevention, investigation and correction of institutional child abuse and neglect; and (8) court procedure. Each Section contained a varying number of "Standards" which were typically general statements of functions or activities. Each "Standard" was followed by a varying number of "Guidelines" which were more detailed and specific statements of the functions or activities relevant to fulfilling the "Standard." As an illustration of the format, one of the more brief "Standards" and its accompanying

^{*}Prior standards directed at service delivery in response to child abuse and neglect were the Child Welfare League of America Standards for Child Protective Service (1973).

"Guidelines" from Section IV on the Investigation of Reports by Local Agencies follows:

"Standard IV C

THE FOCUS OF THE INITIAL INVESTIGATION SHOULD BE TO DETERMINE WHETHER OR NOT THE CHILD IS SAFE AND WHETHER OR NOT THE REPORT IS VALID.

Guidelines

- . One person should be given responsibility for the investigation. This person should be informed about all information in the report and all other sources.
- . The referral and/or reporting source should be contacted for any additional information that can be provided, with emphasis on information which may indicate that an emergency investigation should be started. (This action will also provide feedback that the report is being investigated).
- . Information on prior occurrences of abuse and neglect to the subject child, his/her siblings, or others in the family should be obtained from the Central Register (if it was not transmitted at the time that a case was assigned by the State agency to the local agency for investigation). Except in extreme emergencies this information should be obtained and reviewed prior to interviewing the parents or visiting the home.
- . The child protective worker assigned to the investigation should see the child and ensure that the child is safe and not in any imminent danger.
- . An assessment should be made as soon as possible of the child's home situation, through a visit to the home and through interviews with the parents, caretakers, and others, such as an examining physician, who can be expected to provide pertinent information.
- . When appropriate, a medical assessment should be obtained. And, if it is felt that the child or his/her siblings may be in danger, arrangements should be made to utilize the appropriate emergency services and/or place (or retain) the child (children) in protective custody.

- . If there is doubt about whether the report is unfounded, or whether the report requires additional investigation, the case should be reviewed with the Interdisciplinary Team. See Standard V B for additional information on the Interdisciplinary Team.
- . If the child protective service determines that the report is unfounded, it should:
 - (1) close the case, and,
 - (2) if appropriate, offer services that it thinks could benefit the child or family, for the family's voluntary acceptance or refusal."

(Federal Standards on the Prevention, Identification and Treatment of Child Abuse and Neglect, 1976: 106-108). This Standard and its Guidelines is typical of the form and substance of the Standards. The exception was a comprehensive set of services which were prescribed to be "adequate" to certain needs. Service adequacy will not be dealt with here because it is a special problem which has been elaborated elsewhere (Seaberg and Bell, 1978). The focus here is only on the nonservice aspects of baseline evaluation.

The process by which the Standards were developed resulted in their validity being assumed by their developers. That is, the input of the expert panel along with the comment and revision process were assumed to yield a set of Standards whose provisions represented the best response possible along the most appropriate dimensions (e.g.: reporting statutes, overall protective service administrative procedures and coordination, investigation of reports, treatment planning and coordination, court procedures, service availability for primary and secondary prevention, and others) for an optimal handling of child abuse and neglect problems. Although the validity assumption is necessary, it is also problematic as will be revealed below in the discussion of reliability and validity.

The primary feature which distinguishes baseline evaluation from other forms of evaluation is the existence of a predetermined set of objectives (Standards, in the case example) and sub-objectives (Guidelines, in the case example) imposed on the particular program(s) from sources outside the program(s). Each standard and its components may thus be seen as externally derived program objectives. With the objectives set, the primary consideration becomes how to measure at a given point in time the extent to which existing provisions compare

with the prescribed provisions as specified in particular standards. The implied comparative procedure, however, is deceptively simple. As the standards become more expansive in scope and more comprehensive (as was true with the case example) there is a clear effect on the form of measurement which can be utilized. A tradeoff is forced between the number of variables included in the study and the characteristics and sophistication of the measures which can be used. More specifically, the fewer the variables, the more measurement can attend to multidimensionality and subtle nuances of selected variables. The greater the number of variables (given constraints of cost and comparative unobtrusiveness), the more crude the measures. In the case of the child abuse and neglect Standards, operationalizing their substance into measurable variables or items resulted in 525 items for which information had to be obtained from a great number of sources. The magnitude of the data collection, analysis and reporting problem, then, virtually dictated a measurement procedure which would provide only the most basic information for every item.

In field testing the baseline evaluation procedure, measurement for all of the items, other than service items, was accomplished by the application of a set of four questions to every item. The four questions were cast into two subsets. The first subset, pertaining to the assessment of existing provisions, asked: (1) are existing provisions consistent with those prescribed in the Standards, and (2) if existing provisions are not consistent with the Standards, in what way do they vary? The second subset, which was directed at documenting planned or potential changes in existing provisions which might increase or decrease consistency with the Standards, asked: (3) are there any plans for changing existing provisions or developing new ones related to... (the main themes of clusters of variables derived from the Standards), and (4) to achieve these changes, what problems would need to be dealt with successfully? This second subset of questions served the general function of summarizing planning for new provisions, but it also served the evaluation function of indicating the relative degree of stability associated with the degree of consistency found to exist through the first subset of questions.

Attendant to the basic design of the baseline evaluation procedure was the problem of casting the substance of the standards into measurable units of analysis. In this instance, the reduction of the Standards from their narrative format to key variables was deemed both unwise and impractical; unwise, because of the likelihood of increasing measurement error through loss of critical nuances of meaning in

the course of restatement or content reduction, and impractical because of the great number and range of the Standard components. The process content of the Standards would have been difficult to capture through a content reduction and the validity assumption would have been open to even greater question. Given these constraints, the units of measurement and analysis were derived simply by treating the Guidelines (or their components where more than one existed) as units for measurement. This procedure was judged to result in the least amount of measurement error introduced by the evaluators and, in this case, was greatly facilitated by the presentation of the Standards in a well-structured, outlined format delineating the 525 items. Baseline evaluation, then, involved collecting data appropriate to the four primary questions applied to every one of the 525 items.

A field test of the baseline evaluation procedure was conducted at the state level in the States of Iowa and Washington and at the local level in three counties (which were varied by population from urban to rural) in each of those states.

Sources of Data

The primary task related to data collection for the baseline evaluation was the location of appropriate sources of data. Because of the scope and variation in the substance of the Standards and the multiprofessional, multiorganizational character of the prescribed optimal response to child abuse and neglect, a large number of data sources were necessary.

For the first subset of questions which dealt with the consistency between state and local provisions and those prescribed in the Standards, first preference was given to appropriate documentary sources. These included such documents as the state statute on reporting and investigation of child abuse and neglect cases, the state administrative code, the state department of social services procedure manual and related memoranda, procedure manuals and related memoranda for the local offices of the state department of social services, local juvenile/family court procedure manuals, and so on. The preference for deriving the data on consistency from documentary sources was based on the prerequisite need to concentrate the evaluation on the baseline issue of provision (present/absent) which precedes the move into the implementation stage of evaluation. There tended to be a consistent drift toward the latter when gathering the consistency data through an interview with a key informant.

Interviewing key informants was the selected alternative to using documents on the consistency questions. Informants were drawn upon if there were no appropriate document or if existing documents did not cover all elements of the Standards on a particular topic. While appropriate roles or positions identifying key informants were usually apparent from the substance of the Standards, each role incumbent approached for this purpose was asked to verify that he/she was the most appropriate respondent to the particular Standard items on which the consistency questions were being asked. Appropriateness was defined in terms of access to information confirming the presence or absence of the elements (Guidelines) comprising a Standard. Occasionally this led to another key informant.

The field test revealed that for elements of the Standards which related to local level provisions, there tended to be less formal documentation of provisions as the county became smaller and the social service and court organizations became correspondingly less complex organizationally. The result was a greater reliance on key informants for the consistency questions as the counties varied from urban to rural.

For the second subset of questions (dealing with plans to change existing provisions for responding to child abuse and neglect, a measure of the stability of those provisions), all data were collected through interviews with key informants. These sources for the planning interviews were selected because of their occupancy of positions which would give them access to a broad view of a particular substantive area and planning information related to that area. Again, as a part of the interview, each respondent was asked to self-screen themselves as to whether they were the best respondent for the information being sought. This process did result in a few changes of key informants.

For purposes of these planning interviews the 525 items of the Standards were arranged into clusters based on a common theme(s). The number of items per cluster varied from as few as one to over 100. For each cluster the key informant was asked the two planning questions for each individual item in the cluster. There were few problems with this procedure for the smaller clusters, but for clusters with large numbers of items it turned out to be tedious, highly time-consuming, and because of the close interrelationship of some of the items, occasionally redundant. It was decided, therefore, that for this phase of baseline evaluation some content reduction was necessary. This was accomplished by posing rather comprehensive topical questions which subsumed items of the cluster sharing common themes. For example, for the eight items which made up Standard IV C one question was

composed. It read, in part, "we are seeking information about plans to change or develop the capacity for, timing of, or procedures to investigate reports of child abuse and neglect by the local agency" and was one of nine such comprehensive questions on planning posed to the "local child protective services supervisor" (of the state department of social services). This content reduction presented a very workable procedure which was open-ended enough to allow for input of the appropriate information, but not tedious to the point of precluding active cooperation of the key informants. Through this process the 525 items were reduced to 122 which required 25 key informants.

Reliability and Validity

The validity of data refers to the relevance of the information to the research question. Measurement and data collection procedures are considered reliable if they can be repeated under similar conditions with similar results. The appropriate form for determining reliability is dependent upon the form of measurement or data collection procedure used. The two data collection procedures used in the baseline evaluation procedure required different forms for assessing reliability and validity. (Validity here refers to measurement validity, not to the earlier mentioned issue of the substantive validity of the Standards as the optimal response to child abuse and neglect).

With respect to the prescription/provision consistency, the two straightforward questions and the documentary sources were considered productive of reasonably reliable and valid data. The most likely source of error resided in the interpretation of the written materials relative to the Standards. But even this potential source of error was reduced by the very specificity of the Standards' Guidelines which constituted the items of analysis for the evaluation. To check the reliability of recording data from documentary sources, project staff members independently recorded material from a sample of these sources and produced nearly identical assessments on the consistency questions.

For data generated through interviews with key informants, a reasonably high level of reliability was also determined. The ideal test of reliability, of course, would have been the comparison of data from multiple sources for each item. Somewhat weaker but acceptable would be the comparison of data simultaneously recorded by multiple recorders using the same source at the same time. The former of these two possibilities was not possible given the scope of the data collection problems. The latter was employed on a limited sample basis, particularly with regard to the planning survey questions. Two recorders

were used to document responses for a sample of sources and their records were subsequently compared. The two sets of data were virtually identical which, again, demonstrated a high level of inter-reporter reliability.

Valid measures tend to be reliable as well (Nunnally, 1967). The collection of valid data was achieved by (1) use of documentary sources where possible, (2) use of key informants who had task role positions relevant to the Standard substance, (3) asking key informants to verify their role as the best key informant (candor and cooperation were the norm in this regard), and (4) summarizing the recorded responses and asking verification or correction by the key informant. Through these several mechanisms a reasonable case for the collection of valid and reliable data can be made.

Interpretation of Data

The data accrued in response to the first of the four primary questions (are existing provisions consistent with those prescribed in the Standards?) were of the form "yes" or "no." Data generated for the other three questions were in brief narrative form. With 525 items, even a simple presentation of the data in these forms could be voluminous and tedious to use. Since, however, the four questions were highly interrelated, it was possible to develop a synthesized and easily interpreted format for the presentation of results.

The primary concern of a baseline evaluation is reflected by the first subset of questions dealing with (1) consistency with the Standards and (2) the form of variation from consistency. When the form of variation from consistency was considered, it was apparent that the yes-no dichotomy on consistency was too stringent and that a mechanism was needed for taking into account the degree of variation from consistency. For example, one of the items dealing with the central registry system prescribes that closed cases should be expunged after five years, but in one of the field-tested states closed cases are expunged after six years. A literal interpretation of this guideline would render the state inconsistent with this provision of the Standards, but, on the other hand, the presence of an expunging procedure and the trivial difference of a year between the state's practice and the Standard's prescription suggest that the state is basically consistent with that item of the Standards. To deal with situations of this type, categories of relative consistency were developed as cells of a matrix formed by the dimensions "Consistent with Technical Detail

of Standards" and "Consistent with Intent of Standards" with "Yes" and "No" elements for each dimension. (See Figure 1).

Figure 1

Degree of Consistency Matrix

		Consistent with Intent of Standards	
		YES	NO
Consistent with Technical Detail of Standards	YES	Consistent	∅
	NO	Technically Inconsistent	Inconsistent

The result was three usable categories for categorizing consistency: consistency with both the technical detail and the intent of the Standards was labeled "consistent"; non-consistency with the technical detail, but consistency with the intent of the Standards was labeled "technically inconsistent"; and, non-consistency with both the technical detail and the intent of the Standards was labeled "inconsistent." The fourth cell of the matrix was determined to be an empty set; that is, no instance could be conceived of where there would be consistency with the technical detail and not the intent. The scope and specificity of the Standards ensured that any application of the baseline evaluation procedure would reveal numerous technical inconsistencies. A literal interpretation of each item would render these inconsistent and produce a misleading assessment insofar as the intent of the Standards was concerned.

Applying the consistency categories to the raw data on consistency and variation from consistency, summary statements were developed for each Section of the Standards. Recall that the Standards were organized into eight Sections dealing with major topical areas each of which contained a varying number of Standards and Guidelines, e.g. prevention, reporting, state responsibilities, investigation of reports by local agencies, and so on. The summary statements reviewed the specific areas of technical inconsistency and

inconsistency between existing provisions and provisions prescribed by the Guideline items. Consistencies were not mentioned specifically, but the statements were concluded with a summary judgment of the relative degree of inconsistency with the Standards. For example, for Section IV for Grant County Washington this concluding statement read: "Overall, the Grant County CPS investigative procedures are nearly identical to the guidelines laid out in the Standards. The few inconsistencies that are present are due to the non-existence of the Inter-Disciplinary Consultation Team, and to the local agency's contention that it is always inappropriate to disclose the location of a child who has been placed in protective custody." (Gillespie et al., 1977:54.) The raw data on the consistency questions was presented as an appendix to the report.

It should be noted that as a means of summarizing the degree of consistency with the Standards, there may be an inclination to report the number and/or percentages of consistencies, technical inconsistencies and inconsistencies. Such a summary procedure is not valid, however, since (a) there is no comparability between the relative importance of the specific items used to elaborate each Standard, and (b) the number of items used to elaborate each Standard was at best arbitrary. More important though is the fact that (c) evaluation of the type being described here is essentially qualitative rather than quantitative. That is, the description of the substance of provisions for responding to child abuse and neglect and variation in those provisions from that prescribed by the Standards is the critical point of analysis. A notable example of the inappropriate application of quantitative analysis to what is essentially a qualitative problem occurs in Local Child Welfare Agency Self-Assessment Manual (Turem et al., 1977).

The summary statement on consistency for each Section of the Standards was followed by a summary statement on the stability of existing provisions. This consisted of a summary of the planning survey data and an interpretation of the potential effect of planned development or change on the status of the consistency summary. These potential effects were qualified as appropriate by the data on the problems anticipated in achieving the planned developed or change. For example, if a planned change related to providing trained staff for a local CPS unit involved substantial budgetary appropriations and possibly reclassification of positions which in turn involved a

state employees union, the prospect of that change occurring in the immediate future was considered slim and the current provisions stable. On the other hand, a minor change in procedure for local investigation of reports of child abuse and neglect which could be achieved through an edict of the administrator of the organization containing the local CPS unit, was considered much more probable and, hence, the current provision would be judged unstable.

In combination, then, the two summaries provided a comprehensive, but brief statement for each major substantive section of the Standards which compared state and local provisions for responding to child abuse and neglect with those prescribed in the Standards.

Discussion

The development and utilization of standards as an expression of social policy in the human problem/human service area seems to be increasingly more common. The objective of the research reported here was to facilitate the utilization of comprehensive, multi-system Federal Standards on the Prevention, Identification and Treatment of Child Abuse and Neglect (1976) which were intended to "...assist communities in evaluating existing efforts and planning new ones" (Child Abuse and Neglect Reports, 1976). To this end, an evaluation procedure was developed and field tested which, while specific to these Standards, should have generalizable implications for the development of similar evaluation procedures for standards in other human service areas. In the process of this experience it was necessary to clarify the type of unique evaluation problem confronted, the evaluation methodology appropriate for responding to the problem, and the relation of that methodology to other classes of evaluation methodology. Because the problem was unique, the methodology developed and its place among other evaluation methodologies is unique. For these reasons, this type of evaluation procedure was labeled "baseline evaluation"; the major distinguishing characteristic being that the focal concern was limited to the comparison of existing provisions for responding to a social problem and those prescribed by standards as optimal, with no concern for the implementation and effectiveness of those provisions.

While most of the major technical problems in developing and utilizing a baseline evaluation procedure were dealt with successfully to a reasonable degree in relation to this specific case example, one problem remains which is as much of a problem for the developers of

standards in their conceptualization process as it is for evaluators. This problem has to do with the overall validity of the standard substance as it pertains to the local situation. Validity in this instance is at a conceptual level rather than at a measurement level. The issue can be more clearly presented by reference to the case example upon which this discussion is based. The Standards for responding to child abuse and neglect have features which are oriented toward investigative and service organizational units which are complex and where specialized functions are appropriate. These include, for example, items for provision of child protective service units organized separately from and administered independently of the rest of the local agency, and routine evaluation of the CPS and community-wide programs by a special team. This structure may be appropriate in the sense that the greatest number of child abuse and neglect cases are responded to by such specialized units, but nationally the greatest number of responding organizational units are relatively small. Of course, it makes no sense for a person in a rural setting who is responsible for responding to an occasional incident of child abuse or neglect on as little as a quarter-time basis to be organizationally isolated as a special unit. Neither would the activities of this person warrant the attention of a special evaluation unit. These are other examples of situations where a standard may be inappropriate or not valid, given the local situation.

The implications of this type of validity problem pose a dilemma for the evaluator implementing a baseline evaluation. As in the example cited above, local provisions will be inconsistent with the Standards, but for the local situation the Standards are inappropriate. There appear to be only two ways in which this dilemma can be resolved, either (1) by rewriting the standards to provide for appropriate variance on these matters, or (2) by building into the evaluation procedure a mechanism for local determination of the appropriateness of each standard item to the local situation. The latter option contains the inherent danger that the subjects of the evaluation could carefully structure the evaluation criteria so that only a favorable result could accrue. Since establishing limits on such a redefinition would probably be highly arguable, the former option of rewriting the Standards is preferable. In the instance of the case example, the development and field testing of the baseline evaluation procedure was conducted in relation to a draft of the Standards which meant some rewriting was anticipated anyway. In the generic sense, rewriting the standards in such a way as to be relevant to all local situations seems to be the preferred approach to resolving such problems and accompanying dilemmas related to the conceptual validity of the Standards.

While this discussion of baseline evaluation is limited to the experience accrued from only one case example, and other case examples might reveal additional problems and issues which attend this form of evaluation, a firm start in the development of baseline evaluation has been made. Replication of baseline evaluation on other sets of standards in the human needs/human services area are needed in order to advance this methodology. Attention also must be directed to the development and refinement of implementation evaluation as the other complement to existing effectiveness evaluation methodology.

REFERENCES

Child Abuse and Neglect Reports. Washington, D.C.: National Center on Child Abuse and Neglect, Children's Bureau, Office of Child Development, Office of Human Development, DHEW, (February 1976), p. 9.

Child Welfare League of America Standards for Child Protective Service. New York: Child Welfare League of America, 1973.

Federal Standards for Child Abuse and Neglect Prevention and Treatment Programs and Projects. Washington, D.C.: National Center on Child Abuse and Neglect, Children's Bureau, Administration for Children, Youth and Families, Office of Human Development Services, DHEW, Draft - March 1978.

Federal Standards on the Prevention, Identification and Treatment of Child Abuse and Neglect. Washington, D.C.: National Center on Child Abuse and Neglect, Children's Bureau, Office of Child Development, Office of Human Development, DHEW, Draft - October 1, 1976.

Gillespie, D. F., et al. A Baseline Evaluation Procedure for Federal Standards on the Prevention, Identification and Treatment of Child Abuse and Neglect, Volume II: State of Washington Field Test. Seattle: Center for Social Welfare Research, School of Social Work, University of Washington, 1977.

Jones, W. C. and Borgatta, E. F. Methodology of Evaluation. In E. F. Mullen and J. K. Dumpson, Editors, Evaluation of Social Intervention. San Francisco: Jossey-Bass, 1972.

Nunnally, J. C. Psychometric Theory New York: McGraw-Hill Book Company, 1967.

Seaberg, J. R. and Gillespie, D. F. A Baseline Evaluation Procedure for Federal Standards on the Prevention, Identification and Treatment of Child Abuse and Neglect, Volume I: Development, Field Testing and Recommended Procedure. Seattle: Center for Social Welfare Research, School of Social Work, University of Washington, 1977.

Seaberg, J. R. and Bell, M. A. Evaluation of Broad-Scale Service Networks. Social Work Research and Abstracts, 14, (Spring 1978).

Standards for Community Agencies, Serving Persons with Mental Retardation and Other Developmental Disabilities. Chicago: Accreditation Council for Facilities for the Mentally Retarded, Joint Commission on Accreditation of Hospitals, 1973.

Suchman, E. A. Evaluative Research. New York: Russell Sage Foundation, 1967.

Turem, J. S., et al. Local Child Welfare Agency Self-Assessment Manual (Draft). Washington, D.C.: The Urban Institute, March 1977.

Weiss, C. H. Evaluation Research: Methods of Assessing Program Effectiveness. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Wholey, J. S., et al. Federal Evaluation Policy. Washington, D.C.: The Urban Institute, 1975.