2005

Ethnography and evaluation: Their relationship and three anthropological models of evaluation

Brandon W. Youker, Ph.D, *Grand Valley State University*

Ideas to Consider

Ethnography and Evaluation: Their Relationship and Three Anthropological Models of Evaluation

Brandon W. Youker

Abstract

This paper examines the relationship between ethnographic research methods and evaluation theory and methodology. It is divided into two main sections: (a) ethnography in evaluation and (b) anthropological models of evaluation. Three levels of the leading anthropological models of evaluation are summarized, which include responsive evaluation, goal-free evaluation, and constructivist evaluation. In conclusion, (a) there is no consensual definition of ethnography; (b) in many circumstances, ethnographic evaluation models may be beneficial; and (c) ethnography can be used in evaluation but requires a high level of analysis to transform ethnographic data into useful information for eliciting an evaluative conclusion.

Introduction

Ethnography[1], an applied qualitative social science research method, is often employed in program evaluation. Ethnography, alone and according to its pure anthropological definition, is not a research method capable of being the sole method implemented in an evaluation. Ethnography may prove advantageous to evaluators as an additional method to be employed or considered. However, sound evaluation typically requires multiple data collection methods and a higher level of analysis than ethnography alone can provide. Evaluation synthesizes the narrative and develops an evaluative conclusion. There are various instances when the implementation of an evaluation model that relies heavily on qualitative methods based in the tradition of anthropological research is beneficial. As an evaluator, at minimum, familiarity with these models should be in one's repertoire.

The paper is divided into two main sections: (1) Ethnography and Evaluation and (2) Anthropological Models of Evaluation. The first section presents a summary definition of ethnography, its theories, concepts, and benefits; and the difference between ethnography and anthropology. The author then provides a brief definition of evaluation and discusses the relationship between ethnography and evaluation. There are three anthropological models of evaluation in which the author summarizes, discusses the strengths and limitations, and reflects on their

---

[1] AUTHOR'S NOTE: The author of this paper uses the terms "ethnography," "ethnographic techniques," and often "qualitative research methods" interchangeably. Additionally, the term "program" is used generically, to refer to the *evaluand**. Ethnography in the context of this paper is primarily in regards to program and policy evaluations. Ethnography may also be used in product, personnel, and performance evaluations.—* *"Evaluand:* That which is being evaluated (e.g., program, policy, project, product, service, organization)" (Davidson, 2005, p. 240).

relationship with ethnography. The paper concludes with a synopsis of the author's main impressions and key points.

Ethnography and Evaluation

Ethnography[2] is an applied research method most often associated with anthropology, where it was developed to study cultural interpretation. Ethnography, also called field research, is the process of describing a culture or way of life from a folk peoples' perspective. Anthropologist Clifford Geertz described the ethnographic method as "thick description." It provides detailed notes and descriptions of everything that occurs without attempting to summarize, generalize, or hypothesize. In fact, with traditional ethnography, as a rule of thumb, for every half hour of observation a researcher writes for two hours. The researcher focuses on factual description to allow for multiple interpretations to later infer cultural meaning. To obtain this description of a population's perception, the principle of 'naturalism'[3] is assumed. Thus, trust and rapport are essential between the researcher and the population being studied.

Ethnographers, if following the constructivist[4] philosophy, believe that pure

---

[2] Alternative definitions: Ethnography is "a descriptive study of an intact cultural or social group or an individual or individuals within the group based primarily on participant observation and open-ended interviews. Ethnography is based on learning from people as opposed to studying people" (Beebe, n.d.). Ethnographic research "involves the study of groups and people as they go about their everyday lives" (Emerson, Fretz, & Shaw, 1995). "Ethnography is the art and science of describing a group or culture" (Fetterman, 1989, p. 11).

[3] Naturalism: Leave natural phenomenon alone.

[4] Constructivist philosophy maintains that the researcher manufactures knowledge through her interaction in the field and that there is no objective truth to be uncovered (ontological

objectivity is impossible as: (A) ethnography is an interpretive endeavor by fallible human beings; (B) not all field sites are "foreign" for ethnographers in the same way; (C) ethnography is not replicable; and (D) ethnography is not based on a large number of cases. The epistemology of ethnography is typically a model based on a phenomenologically oriented paradigm, which focuses on multiple perspectives and multiple realities of a phenomenon. Phenomenological inquiry seeks to answer the question: "What is the structure and essence of experience of this phenomenon for these people?" (Patton, 1990, p. 69). Constructivists take a heuristic[5] approach to answering this phenomenological question. According to Fetterman (1989), most ethnographers subscribe to ideational theory, which suggests that change is the result of mental activity—thoughts or ideas—versus materialists who believe that "material conditions—ecological resources, money modes of production-are the prime movers" (Fetterman, 1989, p. 16). The most popular ideational theory is cognitive theory, which assumes we can infer peoples' thoughts from hearing what they tell us.

While many theories, concepts, and methods (e.g., in-depth, open-ended interviews, direct observation, written documents, triangulation) resulting in narrative description commonly recur in the literature, consensus on any one set of fundamental principles of ethnography cannot be found (Genzuk, 2001; Patton, 1990; Payne, 1994). For example, ethnographic theories, concepts, and data collection techniques are also used in non-ethnographic qualitative research and

---

relativism) (Maxwell, 1998 in Bickman & Rog, 1998).

[5] Heuristics is a form of phenomenological inquiry focusing on the personal experiences and insights of the researcher—it considers researcher's experience in addition to other observers that experience the phenomenon.

distinctions between ethnography and other qualitative theories, concepts, principles, and methods is not clearly evident. Instead, there are copious combinations of varying concepts considered fundamental to ethnography from researchers and anthropologists alike.

The key in understanding the differences between ethnography and other qualitative social science research methods is to understand the multiple combinations of techniques, concepts, and data collection methodologies encompassed under the term "ethnography." As with all research methodologies, each philosophical and theoretical decision is located on a spectrum or continuum. Thus, the definition of ethnography and what it entails is idiosyncratic to the ethnographer or researcher depending on her degree of commitment to a hodgepodge of "fundamental" concepts. Past and current literature presents definitions and concepts of ethnography differing by technique, values emphasized, time allotted, data analysis procedures, and commitment to the purist practice of anthropological ethnography. Therefore, ethnographic techniques are qualitative in nature but distinct. Below are a few of the reoccurring concepts specific to ethnography (Fetterman in Bickman and Rog, 1998; Genzuk , 2001; Hall, n.d.):

✓   The focus is on culture and cultural interpretation.

✓   There is an emphasis on an emic[6] perspective.

✓   The holistic perspective is often of greater depth than other qualitative research methods.

✓   Sampling measures are conducted over a longer period of time.

---

[6] Emic perspective is that of the insider and includes the acceptance of multiple realities.

✓   The researcher herself is the primary tool for data collection.

In contrast to ethnographic methodology, *evaluation* methodology commonly comprises the use of both qualitative and quantitative techniques. Commonly defined as the systematic determination of something's merit, worth, or significance (Davidson, 2005; Sanders, 1994; Scriven, 1991). Scriven (1991) claims that evaluation is not only a methodology, but a distinct multi- and transdisciplinary field of study not to be regarded as merely a as a branch of applied social science. As an independent discipline, evaluation may utilize applied social science research methodology, but it is distinct by its unique purpose and methodology (e.g., ranking, grading, and scoring). The determination of merit, worth, and significance of an evaluand requires evaluators to consider the relevant values and to make judgments based on those values. Autonomous subspheres of evaluation are program, policy, product, personnel, performance, and proposal evaluations as well as metaevaluation (i.e., the evaluation of evaluation) and intradisciplinary evaluation. Often but not always, evaluations are based in social science research methods including both qualitative and quantitative data collection procedures. Subcomponents of program evaluations, for example, may include the assessment of context, resources, processes, immediate outcomes (outputs), intermediate and long-term outcomes and impacts considering costs, comparisons to best and worst practices of other programs (Davidson, 2005; Scriven, 1991). Moreover, evaluation may be formative, summative, or ascriptive (Scriven, 2005).

Multiple factors may guide evaluators and researchers alike toward choosing quantitative or qualitative evaluation methodology. In the following, qualitative ethnographic evaluation models will be introduced.

Ethnographic Evaluation Models

Ethnographic evaluation methodologies have been discussed for over thirty-five years and came about as a response to the more traditional evaluation approaches which were overly committed to the scientific paradigm of inquiry. According to Guba and Lincoln (1989), an extreme dependence on the methods of science demonstrated some negative results. For example, reliant on primarily quantitative measurement, evaluands were stripped of their context as if they were not entwined in a highly specific one, resulting in irrelevant or non-useful findings (cf. Seafield Research & Development Services). Moreover, scientific truth is non-negotiable, thus all alternative explanations must be in error.

Ethnographic evaluation methods, in contrast, utilize stakeholders' claims and concerns. For example, Guba and Lincoln (1989) insist upon ethnographic methods for determining what information is necessary in an evaluation and provide five reasons:

1. Stakeholders are placed at risk by an evaluation.

2. Evaluation exposes stakeholders to exploitation, disempowerment, and disenfranchisement.

3. Stakeholders represent an "untapped market" for the use of evaluations that are responsive to self-defined needs and interests.

4. Stakeholder input expands the scope and meaningfulness of the evaluation, in addition to contributing to the dialectic process that is necessary in conducting a sound evaluation.

5. All individuals and parties can be mutually educated toward more sophisticated personal constructions and they may gain enhanced

appreciation of the constructs of other individuals or parties.

Other strengths of implementing ethnographic methods in evaluation are exemplified in connecting quantitative data to observed actual outcomes; the flexibility of design; the 'thick description' of program impactees; the clarification of processes; the study of participation; and the identification of unintended positive and negative side effects. The weaknesses in utilizing ethnographic methods in evaluation consist of such problems as introducing complex threats to validity; increasing the time and cost demands compared to other methods; raising the potential for impactee reactivity to the evaluator; and limiting the ability to compare the data from different measurement instruments.

There are many considerations that will need resolution before deciding if an ethnographic method is an appropriate method for an evaluation. Considerations include the purpose of the evaluation; whether the evaluation is formative or summative; the amount of time allocated for the evaluation; the financial and other resources available; and the level of expertise and competence of the evaluation team. Prior to adopting a specific methodology or model, all the typical issues regarding methodology, conceptual context, validity, ethics, etc. must be discussed.

Relationship Between Ethnography and Evaluation

In evaluation, ethnography should be viewed on a spectrum. One end of the spectrum consists of the pure anthropologically-defined ethnography and on the opposite end are various ethnographic techniques of data collection and methodologies loosely defined, combined, and flexibly implemented. Many researchers and evaluators implement one or a few qualitative data collection methods and then claim their research to be ethnographic. However, most agree that ethnography is defined by the rigor of the data collection procedures.

Fetterman (1982) identified a study that called itself ethnographic although the researchers were on site for only five days. Deneberg (1969) and Fetterman (1982) claim that these researchers are fickle to scholastic fads and refer to them as "Zeitgeister-Shysters." Zeitgeister-Shysters become involved in research that is a hot topic or trendy and the result is superficial research. Such researchers contribute minimally to the field and often tarnish the reputation and credibility of ethnography. In describing the Zeitgeister-Shysters, Fetterman stated, "rather than conducting ethnographies, they are simply using ethnographic techniques" (Fetterman, 1982, p.2). Wolcott (1980) concluded that "much of what goes on today as educational ethnography is either out and out program evaluation, or, at best, lopsided and undisciplined documentation" (p.39). Fetterman warns that the adoption of random elements of ethnography without emphasis on the whole, results in "the loss of the built-in safeguards of reliability and validity in data collection and analysis" (Fetterman, 1982, p.2). Researchers often use anthropological tools (ethnography) without understanding the values and cosmology underlying the ethnographic techniques. Wolcott (1980) reminds the reader that the purpose of ethnography is cultural interpretation and this requires the researcher to examine the whole trait complex rather than a few single traits. Still many evaluators study single traits and call their evaluation 'ethnographic'.

The importance of ethnographic data sources in the evaluation of social programs and policies is rarely argued (Agar, 2000; Fetterman, 1982; Fetterman, 1984; Guba & Lincoln, 1989; Hopson, 2002; Patton, 1997; Posavac & Carey, 1997; Scriven, 1991; Swartzman, 1983; Shadish, Cook, and Leviton, 1995; Stake, 1975; Wholey, Hatry, & Newcomer, 2004; Wolcott, 1982; and Worthen, Sanders, & Fitzpatrick, 1997). Hopson (2002), for example, cites a report by Nastasi and Berg (1999) who urge evaluators to "capture views of program participants about their experience of

a program, its acceptability, and whether or not they were influenced to modify behavior or thinking" (p. 45). This has always been a consideration for evaluators, as it pertains to, or affects the program's quality, significance, or merit. Experienced evaluators typically employ several qualitative data collection methods in an evaluation in hopes of understanding some of these cultural issues, albeit less in depth than with pure ethnography.
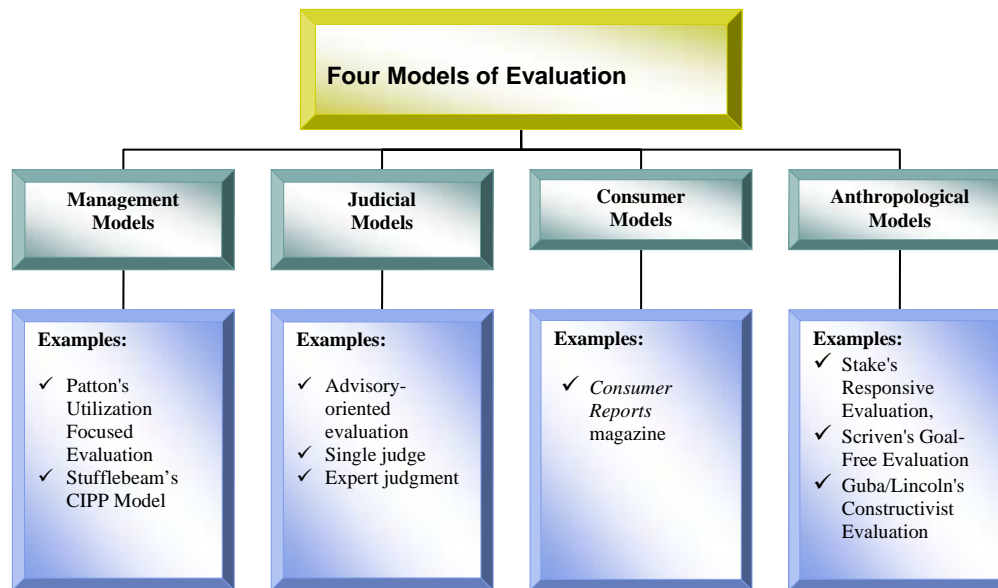
Focusing on context is crucial in all evaluations and the utilization of qualitative methods is fundamental to any good program or educational evaluation; however, the title 'ethnographic evaluator' may be problematic or misleading. Many readers may assume that the term "ethnographic evaluator" implies the use of ethnography in conducting an evaluation. This is false and arguably not possible. Ethnography is a social science research method that emphasizes cultural interpretation. The product of ethnography is a non-judgmental description of context and then a cultural interpretation of the program.

Evaluation is the systematic process of determining the merit, worth, significance or importance of the evaluand. To evaluate something, the relevant values are determined and used to place judgments regarding the overall quality of the program. Ethnography and other qualitative research methods are instrumental in collecting data for determining the most important values to use as criterion for success. Ethnography may uncover unanticipated costs, processes, and outcomes; however, other qualitative methods may reveal similar side-effects but take much less time. There is a point of saturation when a researcher gets the sense that it is unlikely that further study will uncover significant new information that will be important to include in the evaluation. Extended time in the field may not be necessary or feasible for many evaluations.

To summarize, ethnography is a research method and evaluation uses multiple research methods to collect information for determining the merit or worth of a program. As Fetterman (1984) points out, the distinction between ethnography and evaluation is regarding the level of analysis and objective. Evaluators take the ethnographic data to a higher level of analysis by extracting data which is relevant to some standard; comparing it with data from other methods and sources; and judging the program accordingly. Therefore, I conclude that "ethnographic evaluation" is a misnomer or false label for what some evaluators do. Moreover, evaluations claim to use ethnographic methods while in reality, they simply employ varying degrees of qualitative methods. Anthropologically, pure ethnography may serve useful when analyzed further by an evaluator to examine actual processes and outcomes. Anthropological evaluation techniques may be best when conducted independently of the more quantitative research methods, similar to Scriven's (1991) goal-free evaluation. Therefore, in an evaluation which uses multiple research methods, ethnography serves as a way of triangulating these methods. Furthermore, ethnographic data is useful in triangulating data sources adhering to the *principle of critical multiplism* (c.f. see Shadish, 1994). An examination of three evaluation models which are based in anthropology will further illustrate the relationship between ethnography and evaluation.

Anthropological Models of Evaluation

Payne (1994) categorizes 4 evaluation models, the fourth of which contains anthropological approaches (see Figure 1).

Four Models of Evaluation

Management Models	Judicial Models	Consumer Models	Anthropological Models
Examples: ✓ Patton's Utilization Focused Evaluation ✓ Stufflebeam's CIPP Model	**Examples:** ✓ Advisory-oriented evaluation ✓ Single judge ✓ Expert judgment	**Examples:** ✓ *Consumer Reports* magazine	**Examples:** ✓ Stake's Responsive Evaluation, ✓ Scriven's Goal-Free Evaluation ✓ Guba/Lincoln's Constructivist Evaluation

Figure 1. *Models of Evaluation* **(adapted from Payne, 1994)**

The anthropological models of evaluation—responsive evaluation, goal-free evaluation, and constructivist evaluation—have many similarities. They tend to be qualitative, exploratory, highly descriptive, and take an inductive approach to understanding the program under evaluation. Each model was created in the post-positivist value-pluralist perspective, focusing on the question: *whose values and methods should shape or have shaped the evaluation?*

The anthropological models protect against any of the evaluator's personal opinions from being used to determine the values and methods emphasized in the evaluation. However, Scriven separates goal-free evaluation from the other two anthropological models by contending that the stated goals of the client should also not be known or utilized by the evaluator. The three models re-examine the ontology[7] of evaluative interpretations. In both responsive evaluation and constructivist evaluation, the selection of relevant values and the determination of

---

[7] Ontology: The nature of the real.

the merit of outcome measures are decided by the program impactees and stakeholders. Evaluators are partners with the stakeholders in the creation of data and they orchestrate the consensus building process. By contrast, in goal-free evaluation, program success is decided by examining change relative to the identified needs through a comprehensive needs assessment. Lastly, all three models rely on an evaluator with significant commitment to and experience with ethnographic and qualitative methods.

The remainder of the paper will discuss each anthropological evaluation model and illustrate its relationship to ethnography and the qualitative research paradigm of evaluation.

Responsive Evaluation

Stake (1975) called his approach to evaluation responsive evaluation to stress flexibility and responsiveness to the concerns and issues of the program stakeholders. Responsive evaluation is less reliant on formal communication such as the statement of goals, objective tests, standards of program personnel, and research-type reports. Rather, it focuses on gathering the observations and reactions of the program stakeholders, which as Stake claims, is the way people naturally evaluate things. Stake believes this and other qualitative methods are not frequently employed in evaluation due to "subjectivity." Responsive evaluation is poorly suited for evaluating formal contracts, and there lies potential to uncover negative side effects or raise embarrassing questions.

Stake suggests examining a program by organizing the evaluation into four components: environment, workspace, output, and support (see Figure 2).

Environment

- Quantity (investigate for quantity including the counting of frequencies, occurrences, products, performances, participants, resources, etc.).
- Diversity (diversity in artistic products, performances and participants).
- Excellence (refers to technique or quality of execution/performance; has a varying threshold of acceptability).
- Originality (separate from quantity and diversity; referring more to creativity and inventiveness; the ability to make someone "catch their breath"; best measured by degree on a variable range).
- Vitality (changeability of physical environment measured over time; encourages regular review of the physical conditions and aesthetics of environment).

Workspace

- Space and content - suitability and accessibility
- Quantity and quality of equipment and supplies

Output

- Measure outputs with careful consideration of the threshold of acceptability
- Incorporate experts in the field

Support

- Within the program and from the community, the school or organization as a whole
- Investigates how outputs are regarded and rewarded

Figure 2. *Four Components of Evaluation: Environment, Workspace, Output, and Support* **(adapted from Stake, 1975)**

Exemplifying educational evaluation, Stake states,

> [A]n educational evaluation is a *responsive evaluation* if it orients more directly to audience requirements for information; and if the different value-perspectives present are referred to in reporting the success and failure of the program.

> (Stake, 1975, p. 14)

It is not critical to be explicit about purpose, scope, or causation in determining worth, according to Stake. In conducting responsive evaluation, the evaluator observes the program to gather narrative and descriptive information from program stakeholders; and negotiates values in which to judge the program. An evaluator should not presume that only the measurable outcomes provide evidence of the program's worth. Outcome evaluations tend to negate the idiosyncratic and unique

ways people benefit from involvement with the program and among each other; furthermore, they are not sensitive to changes in program purpose. Stake cites Scriven (1967) and suggests that it may be preferable to evaluate the "intrinsic merit of the experience rather than the more elusive payoff" (p. 27). Stake feels that less emphasis on preconceived notions of success will allow for increased stakeholder flexibility in determining the purposes of the evaluation and criteria by which to measure success. In a responsive evaluation, the evaluator has the ability to respond to emerging issues, rather than sticking to a strict evaluation plan or structure. This ultimately leads to an increase in the evaluation's utility to the program stakeholders. Recurring events in responsive evaluation (Stake 1975):

1. Talking with clients, program staff, and audiences.

2. Identifying program scope.

3. Providing an overview of program activities.

4. Discovering purposes, concerns.

5. Conceptualizing issues, problems.

6. Identifying data needs regarding issues.

7. Selecting observers, judges, and instruments if any.

8. Observing designated antecedents, transactions, and outcomes.

9. Providing a theme; preparing portrayals, case studies.

10. Winnowing, match issues to audiences.

11. Formatting for audience use.

12. Assembling formal reports, if any.

Data is collected through direct personal experience or the second best option,

vicarious experience. Observations are not only conducted by the evaluator, but the evaluator enlists program stakeholders according to the issues being studied and the audience being served. Having multiple observations and observers increases data reliability; observations continue to be subjective but through replication random error is reduced. The bias of direct or vicarious experience decreases as repeated observation and diverse points of view are attained. The evaluator produces portrayals typically featuring descriptions of persons, such as a five-minute script, a log, scrapbook, multi-media or audience role-plays. The small number of case studies is often criticized for sampling error, but Stake attests that the error may be minimal and that it is a small price to pay for potentially substantial improvements in communication. Moreover, Stake assumes that case studies of several students are more interesting and representative of a program than a few measurements on all program participants. Therefore, the reader benefits by a more comprehensive understanding of the program.

The evaluation encounters two pluralisms of values: (1) in context, or in Stake's terms the "antecedent condition in which the program is found" (p. 23) and (2) the personal outcomes or outcomes of the program. The evaluation team should not impose its values on the "actors," "spectators," and/or "critics" of the program during the consensus building process. Stake identifies two measures of the value of evaluation: its increment of added experience and its enhancement of responsive alternatives.

Strengths of responsive evaluation include it being flexible, adaptable, and good in providing cultural explanation and recognition of diversity. It may be particularly useful in evaluating programs where the stakeholders generally agree on the intrinsic value rather than the instrumental value of the program. For example, many people will discuss the importance of music and art "because they're good

things to do" (ibid, p. 16). Furthermore, in formative evaluations, responsive evaluation is useful in monitoring the program and to identify positive and/or negative side effects. It is helpful in summative evaluations by giving the reader an understanding of the program's activities, its strengths and weaknesses, and by providing a vicarious experience in the evaluation.

Limitations of responsive evaluation include the difficulty in making comparisons to standards; it serves the immediate audience and may not fulfill distant or future needs. In today's world, funding constraints on arts education programs, for example, has led to an increased demand for quantifiable outcomes and results which are not emphasized in responsive evaluation. Moreover, responsive evaluations may be less objective, reliable, and generalizable as compared to traditional evaluations, or as Stake calls them preordinate evaluations. Responsive evaluation is not useful when it is important to measure goal attainment, whether promises were kept, or in cases where predetermined hypotheses are to be examined.

Ethnography, in the more traditional sense, has compatibility problems with responsive evaluation but there may be potential for combining them. A primary distinction is that with responsive evaluation, the evaluator solicits the observation of stakeholders, thus making the stakeholders part of the evaluation team and adding them as additional data collection instruments. Nevertheless, with some concessions on both sides, the two may be combined.

Goal-Free Evaluation

The evaluator, in a goal-free evaluation (Scriven, 1991), intentionally enters the field without being aware of the specific stated goals and objectives of the program. The evaluator learns about the program and its outcomes inductively.

This means that all program materials are screened either by a non goal-free evaluator on the evaluation team, an administrative assistant, or by the client to ensure that none of the stated goals or objectives are described to the goal-free evaluator. The purpose of this is:

> …finding out what the program is actually *doing* without being cued as to what it is *trying* to do. If the program is achieving its stated goals and objectives, then these achievements should show up; if not, it is argued, they are irrelevant.

(Scriven, 1991, p. 180)

Goal-free evaluations can assist in determining whether the objectives are worthwhile; instead of "are the objectives being achieved?" It is similar to the double-blind pharmaceutical study; just like the drug evaluator, the goal-free evaluator does not have to know the direction of the intended effect or the intended extent of the outcomes (Scriven, 1973). The evaluator intends to find the program's actual outcomes and then works backward to determine if the effects were caused by the program. The goal-free evaluator is like the crime scene investigator who tries to eliminate rival explanations which may have led to the outcome under investigation. Information regarding the stated goals of the program is withheld from the evaluator. However the evaluator is able to review some program documents, budgets, schedules, recorded observations, profiles of participants and staff, etc. as long as there is no implication of any stated goal.

A comprehensive, fair, and accurate needs assessment is essential in conducting a goal-free evaluation. Merit is determined by comparing the actual program outcomes to the relevant needs of those impacted, instead of to the program goals or consumer wants or desires. The program is evaluated according to the level of fulfillment of the consumers needs. Scriven believes by keeping the goals vague, a

less pure goal-free evaluation still makes finding outcomes difficult and encourages the evaluator to connect program effects to recipients' needs instead of the stated goals of the program. Altschuld and Witkin (2000) state that the needs at the primary level (i.e., recipients of the program) are the most critical concern, and from there the needs assessment can considers the needs of the service deliverers and the program delivery system. They argue that the primary needs are the "raison d'être" or the "rationale for the existence" of the service deliverers and delivery systems (Altschuld & Witkin, 2000, p. 10).

There are also relative degrees to which an evaluation may be goal-free. Goal-free evaluations may be combined, in full or in part with other evaluation methods (e.g., "qualitative versus quantitative, survey versus experiment, multiple perspectives versus one right answer, etc.", Scriven, 1991, p. 182). Additionally, an evaluation may begin goal-free and then become goal-based; the reverse is not possible. It is also suggested that goal-free evaluation can be used as a supplement to a traditional outcomes evaluation conducted by a separate evaluator. The evaluator implementing the goal-free evaluation collects exploratory data to supplement and provide context to another evaluator's goal-oriented data. Goal-free evaluators observe the program in an attempt to understand the culture, meanwhile considering needs, processes, and outcomes. Below, the author provides a simplified illustration of a goal-free evaluation using a physical education and training program.

The evaluator of a physical education and training program enters into the evaluation without any prior knowledge of the program's goals. She would likely be capable of directly observing changes in health-related knowledge, strength, and endurance, which are the program's stated goals. However, the goal-free evaluator might also discover changes in endurance, flexibility, physique, changes in

behavior, social status, networking with other students, finding new supportive workout partners, sharing of dietary and nutrition tips, increased self-esteem, etc. all of which were not original goals of the program and would be considered positive, unintended side-effects. They would likely have been missed if the evaluation solely examined the stated or preordained goals.

Arguments for the utilization of goal-free evaluation include (Scriven, 1991):

- It may identify unintended positive and negative side-effects and other context specific information.

- As a supplement to a traditional evaluation, it serves as a form of triangulating both data collection methods and data sources.

- It circumvents the traditional outcome evaluation and the difficulty of identifying true current goals and true original goals, and then defining and weighing them.

- It is less intrusive to the program and potentially less costly to the client.

- It is adaptable to changes in needs or goals.

- By reducing interaction with program staff, it is less susceptible to social, perceptual, and cognitive biases.

- It is reversible; an evaluation may begin goal-free and later become goal-based using the goal-free data for preliminary investigative purposes.

- It is less subject to bias introduced by intentionally or unintentionally trying to satisfy the client because it is not explicit in what the client is attempting to do; it offers fewer opportunities for evaluator bias or corruption because the evaluator is unable to clearly determine ways of cheating.

- For the evaluator, it requires increased effort, identifies incompetence, and enhances the balance of power among the evaluator, the evaluee and client.

Scriven and other users of goal-free evaluations have provided minimal direction regarding operational methodology in conducting the model. The only known attempt to develop an operational methodology for goal-free evaluation was by Evers (1980) in a doctoral dissertation. Evers outlined a goal-free evaluation methodology consisting of six components each of with comprising several sub-categories. The six main components were: (1) Conceptualization of Evaluation; (2) Socio-Political Factors; (3) Contractual/Legal Arrangements; (4) The Technical Design; (5) Management Plan; and (6) Moral/Ethical/Utility Questions. The success of a goal-free evaluation is dependent upon the quality of the needs assessment. If there is not an accurate comprehension of the program participants' needs then the entire evaluation may be at jeopardy.

A goal-free evaluation could feasibly be ethnographic. However, goal-free evaluation focuses on using observation to connect needs to actual program activities, rather than for thick description. Furthermore, traditional ethnography focuses on culture which is always goal-free in nature.

Constructivist Evaluation

Guba and Lincoln's (1989)[8] fourth-generation or constructivist evaluation approach

---

[8] The new meaning of constructivist methodology: Truth is determined by consensus building among informed constructors, not of correspondence with an objective reality. Facts are meaningless without a value framework; therefore, no proposition can be objectively assessed. Causes and effects do not exist; accountability is relative and implicates all interacting parties equally (Guba & Lincoln, 1989).

outlines five generations of evaluation: (1) measurement (e.g., IQ testing); (2) description (e.g. formative evaluation of programs); (3) judgment of merit and worth; (4) *constructivist* (negotiated co-creations of social reality); and (5) meta-evaluation (the evaluation of an evaluation). The constructivist evaluation was created in response to the perceived failure or critical flaws of the first three generations of evaluation. The fourth-generation evaluator may use any of the earlier evaluation techniques as appropriate. Carney (1991, p. 35) reports that the underlying method in fourth-generation evaluation is known by other names:

> British scholars call it 'human inquiry' (inquiry conducted in human ways for humane ends); Americans scholars call it 'action research' (research which aims to produce action on or through it[s] findings, and third world or developmental evaluators call it 'developmental evaluation' (evaluation which develops the understanding, and resources to respond, of those evaluated). A common generic term for it is 'collaborative inquiry' (which simply describes what goes on when you use the method).

In constructivist evaluation, evaluation is:

a. A process that combines data collection and data valuing (interpretation) into one inseparable.

b. A local process.

c. A sociopolitical process.

d. A teaching and learning process.

e. A continuous, recursive and divergent process.

f. An emergent process.

g. A process for sharing accountability.

   h. A hermeneutic dialectic relationship.

In collaborative inquiry the people being evaluated participate as informed collaborators rather than research subjects. The purpose of a constructive evaluation is to attain a deeper comprehension of all the issues encountered by all the stakeholders and consumers; while the goals comprise mutual education, improved awareness, and increased motivation to utilize the evaluation results. Most constructivist evaluators are relativist and implement qualitative methods; however, the evaluation is conducted in a disciplined manner and it produces an audit trail to ensure transparency and credibility of its findings. The realities discovered by the constructivist inquiry are the constructions of the reality proposed by the evaluees themselves. They develop into co-constructions and subsequently reconstructions, as both evaluators and evaluees mold them. The constructivist evaluation assumes that evaluators are unable to maintain distance from the evaluees. Therefore, it accepts a hermeneutic dialect. Guba and Lincoln continue by rejecting the positivist assumptions, which they claim are embedded in most evaluation methodology. They use "validity" as an example of a term that evaluators are socialized into accepting as the positivist definition. Furthermore, they feel that the relationship between the evaluator and the program managers is often characterized by disenfranchisement and disempowerment. The constructivist evaluation, in the same vein as Scriven's goal-free evaluation, aims to restore the balance of power.

The process of Constructivist Evaluation can be illustrated in nine steps:

   1. Identify all relevant stakeholders.

   2. Elicit from each stakeholder group their construction and concerns regarding the issue at hand.

3. Provide a context and methodology which allows for multiple constructions, claims, concerns, and issues that can be comprehended, critiqued, and factored in the evaluation as appropriate.

   - Conduct this methodology within each stakeholder group.

   - Cross fertilize each group with the constructions, claims, concerns, and issues identified by other stakeholder groups, or issues drawn from the literature or other sites. All view points are taken into account as long as they are open to critique and criticism.

4. Generate consensus.

5. Prepare an agenda for negotiation on items where there is little or no consensus.

6. Collect and provide the information requested in the agenda for negotiation.

7. Establish and facilitate a forum of stakeholder representatives where negotiation can occur.

8. Develop a report, or several reports, that communicate any consensus on constructions and resolutions. Additionally, the report should communicate the pertinent issues raised by other stakeholder groups.

9. Recycle the evaluation to continue working on unresolved constructions.

The main limitations of Guba and Lincoln's model is that it minimally acknowledges the fundamental role of evaluation in determining the merit, worth, significance, value, quality, or importance of the program, which are core elements within the definition of evaluation (see for examples Scriven, 1991; Davidson, 2005; and Sanders, 1994). Guba and Lincoln claim to offer a formative evaluation

model placing little emphasis on making an evaluative conclusion and more on program improvement through consensus-building; however, to suggest improvement, they must determine deficits in the evaluand, thus they do actually evaluate. A second weakness with the constructivist model is in assuming that stakeholders will always offer the reliable, valid, and honest information. There may be many factors contributing to a stakeholder's knowledge, ability, and candor that must be weighed relative to the observed program impacts in providing a valid evaluative conclusion.

Conclusion

Ethnography is an applied social science research method, while evaluation incorporates various research methods, one of which may be ethnography. The purpose of ethnography is thick description and cultural interpretation; evaluation's aim is to systematically judge a program's merit and develop an evaluative conclusion. The qualitative evaluation approach has demonstrated benefits for evaluators, and three of these approaches are epitomized in the anthropological models of evaluation. Responsive evaluation, goal-free evaluation, and constructivist evaluation have conceptual and methodological similarities. An evaluator should be able to recognize when one of these ethnographic or anthropological models may be feasible and appropriate in evaluating a program. The evaluator should then present the model and its strengths and limitations to the program stakeholders to be considered when selecting the most appropriate evaluation methodology. Sound evaluation typically requires the employment of both quantitative and qualitative research methods. Ethnography and the anthropological models of evaluation may be best suited as a supplement to the quantitative components of an evaluation and serve as a way of triangulating data

collection methods and data sources. A competent evaluator should be informed of these ethnographic techniques and the anthropological models of evaluation.

References

Altschuld, J. W. and Witkin, B. R. (2000). *From needs assessments to action: Transforming needs into solution strategies*. Thousand Oaks, CA: Sage Publications, Inc.

Beebe, J. (n.d.). *Rapid assessment process*. Gonzaga University website. Retrieved January 27, 2005 from http://208.164.121.55/reference/SOME/Outlines/rapid_assessment _process.html

Beebe, J. (1995). Basic concepts and techniques of rapid appraisal. *Human Organization, 54*(1), 42-51.

Carney, T. (1991). Fourth generation evaluation. *Canadian Journal of Communication,* 16(2).

Davidson, E. J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation.* Thousand Oaks, CA: Sage Publications, Inc.

Denzin, N. K. & Lincoln, Y. S. (Eds). (2000). *Handbook of qualitative research* (2nd ed.)*.* Thousand Oaks, CA: Sage Publications Inc.

Deneberg, V. (1969). Prolizityies A. Zeitgeister. *Psychology Today,* 311(50).

Evers, J. W. (1980). *A field study of goal-based and goal-free evaluation techniques.* Unpublished doctoral dissertation, Western Michigan University, Kalamazoo.

Fetterman, D. M. (1982). Ethnography in educational research: The dynamics of diffusion. *Educational Researcher, 11*(3), 17-29.

Fetterman, D. M. (1986a). *The ethnographic evaluator*. Stanford University. Retrieved January 24, 2005 from http://www.stanford.edu/~,davidf/230c1ass/ethnographicevaluator.html

Fetterman, D. M. (1986b). Conceptual crossroads: Methods and ethics in ethnographic evaluation. In D. D. Williams (Ed.), *Naturalistic evaluation: New directions for program evaluation.* San Francisco, CA: Jossey-Bass.

Fetterman, D. M. (1998). Ethnography. In L. Bickman & D. J. Rog (Eds.) *Handbook of applied social research methods* (pp. 473-504). Thousand Oaks, CA: Sage Publications, Inc.

Genzuk, M. (2004). *A synthesis of ethnographic research*. University of Southern California-Center for Multilingual, Multicultural Research. Retrieved February 3, 2005 from http://www-ref.usc.edu/~genzuk/Ethnographic_Research.pdf

Guba, E. G. & Lincoln, Y. S. (1989). *Fourth generation evaluation.* Newbury Park, CA: Sage Publications.

Hall, B. (n.d.). *Methods: What is ethnography?* Center for Public Interest Anthropology at University of Pennsylvania. Retrieved January 24, 2005 from http://www.sas.upenn.edu/anthro/CPIA/METHODS/Ethnography.html

Hopson, R. K. (2002). Making (more) room at the evaluation table for ethnography: Contributions to the responsive constructivist generation.

Exploring evaluation role and identity (pp. 37-56). Information Age Publishing.

Jessor, R., Colby, A., & Shwedler, R. A. (1996). *Ethnography and human development: Context and meaning in social inquiry.* Chicago, IL: The University of Chicago Press.

Maxwell, J. A. (1998). Designing a Qualitative Study. In L. Bickman & D. J. Rog (Eds.) *Handbook of Applied Social Research Methods* (pp. 69-100). Thousand Oaks, CA: Sage Publications, Inc.

McLean, L. D. (1975). Judging the quality of a school as a place where the alis might thrive." In R. Stake (Ed.), *Evaluating the arts in education: A responsive approach* (pp.41-58). Columbus, OH: Charles E. Merrill Publishing Company.

Nastasi, B. K. & Berg, M. J (1999). Using ethnography to strengthen and evaluate intervention programs. In J. J. Schensul, M. D. LeCompte, G. A. Hess, B. K. Nastasi, M. J. Berg, L. Williamson, J. Brecher, & R. Glassner (Eds.). *Ethnographers toolkit.* Walnut Creek, CA: Altamira Press.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Sage Publications Newbury Park, CA.

Payne, D. A. (1994). *Designing educational project and program evaluations: A practical overview based on research and experience.* Boston, MA: Kluwer Academic Publishers.

Sanders, J. (1994). *The program evaluation standards* (2nd Ed.). Thousand Oaks, CA: Sage Publications, Inc.

Scriven, M. (2004). Zen and art of everyday evaluation. *Journal of MultiDisciplinary Evaluation*, *1*. Retrieved July 20, 2005 from http://www.wmich.edu/evalctr/jmde/content/JMDE_Num_001_Part_I.htm

Scriven, M. (1991). *Evaluation thesaurus (4th ed.).* Newbury Park, CA: Sage Publications, Inc.

Scriven, M. (1973). Goal-free evaluation. In E. R. House (Ed), *School evaluation: The politics and process*. Berkeley, CA: McCutchan Publishing Corporation.

Scriven, M. (1967). The methodology of evaluation. *AERA Monograph Series on Curriculum Evaluation, Vol. 1* (pp. 39-83). Chicago, IL: Rand McNally.

Seafield Research & Development Services. (n.d.). *Fourth generation evaluation*. Retrieved January, 27, 2005 from http://www.srds.ndirect.co.uk/4th.htm

Shadish, W. R. (1994): *The guiding principals of evaluation.* Boston, MA: American Evaluation Association.

Stake, R. (1975). *Evaluating the arts in education: A responsive approach.* Columbus, OH: Charles E. Merrill Publishing Company.

United States General Accounting Office. (2003). *Ethnographic studies can inform agencies' actions*. GAO-03-455.

Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (Eds.) (2004). *Handbook of practical program evaluation* (2nd ed.). San Francisco, CA: Jossey-Bass.

Wikipedia (n.d.). Hermeneutics. Retrieved March 22, 2005 from http://en.wikipedia.org/wiki/Hermeneutics.htm

Wolcott, H. P. (1980). How to look like an anthropologist without really being one.

   Practicing Anthropology, *3*(2), 56-59.

Wolcott, H. F. (1982). *Ethnographers sans ethnography: The evaluation compromise*. Bloomington, IN: Agency for Instructional Television.

Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (1997). *Program evaluation: Alternative approaches and practical guidelines* (2nd ed.)*.* White Plains, NY: Longman Publishers.

About the Author

Brandon W. Youker obtained his Bachelor of Arts in Social Work from Michigan State University, Master of Science in Social Work from Columbia University in the City of New York, and is a former post-graduate advanced clinical social work fellow at Yale Child Study Center-Yale School of Medicine. Currently, Brandon Youker is a doctoral student in Interdisciplinary Evaluation at Western Michigan University and an evaluator at Western Michigan University's Evaluation Center. His academic interests include evaluation theory, methodology and design; international program evaluation; evaluation of social work practice and programming; the evaluation of human service programs; and the evaluation of arts programs.