A

ABMA, TINEKE A.

(b. 1964, Joure, Holland). Ph.D. Institute for Health Care Policy and Management, Erasmus University of Rotterdam; M.A. Health Care Administration, Erasmus University, Rotterdam; B.Sc. Nursing, Health Care Academy of Groningen.

Since 1990, Abma has worked at the Institute for Health Care Policy and Management, Erasmus University, Rotterdam, and is affiliated with the Institute for Health Ethics, University of Maastricht. She has been involved in numerous evaluations, including responsive evaluations of palliative care programs for cancer patients, rehabilitation programs for psychiatric patients, quality of care improvement in coercion and constraint in psychiatry, and injury prevention programs for students in the performing arts.

Her primary contribution to the field of evaluation is the exploration of how narrative and dialogue can create responsive evaluation approaches that strengthen stakeholders' contribution to policy making in transformational health-care programs. Social constructionist and postmodern thought, hermeneutics, and the ethics of care inform her evaluation practice. Primary influences on her evaluation work include Robert Stake, Egon Guba, Yvonna Lincoln, Thomas Schwandt, Jennifer Greene, Zygmunt Bauman, Rosi Braidotti, Kenneth Gergen, Sheila McNammee, Dian Hosking, Hans-Georg Gadamer, Guy Widdershoven, Joan Tronto, and Margret Walker.

Abma edited the books *Telling Tales, On Evaluation and Narrative, Dialogue in Evaluation,* a special issue of the journal *Evaluation,* and

"Responsive Evaluation," an issue of *New Directions in Evaluation.* Her dissertation was nominated for the Van Poelje Prize by the Dutch Association for Public Administration.

ABT ASSOCIATES

Founded by Clark Abt in 1965, Abt Associates is one of the world's largest for-profit research firms, with over 1000 employees located in nine corporate and 25 project offices around the world. Abt Associates applies rigorous research techniques to investigate a wide range of issues relating to social and economic policy formulation, international development, and business research. Abt clients are found in all levels of government, business, and industry, as well as nonprofit organizations and foundations. Abt Associates is one of the 100 largest employee-owned companies in the United States, with gross revenues exceeding $184 million in fiscal year 2002.

—Jeffrey G. Tucker

ACCESSIBILITY

Accessibility is a common criterion in evaluation, especially of programs, services, products, and information, when the target audience (such as disabled individuals) or the intervention (such as online teaching or Web pages) is presumed to present special challenges of access. These two factors often coincide in evaluations, as for example in evaluating Web-based teaching for persons with disabilities. Many

government agencies have developed guidelines for determining accessibility of a range of services.

ACCOUNTABILITY

Accountability is a state of, or a process for, holding someone to account to someone else for something— that is, being required to justify or explain what has been done. Although accountability is frequently given as a rationale for doing evaluation, there is considerable variation in who is required to answer to whom, concerning what, through what means, and with what consequences. More important, within this range of options, the ways in which evaluation is used for accountability are frequently so poorly conceived and executed that they are likely to be dysfunctional for programs and organizations.

In its narrowest and most common form, accountability focuses on simple justification by requiring program managers to report back to funders (either separate organizations or the decision makers within their own organization) on their performance compared to agreed plans and targets.

In theory, this sounds attractive. It seems likely that such a system will contribute to good outcomes for programs and organizations through providing an incentive system that encourages managers and staff to focus on and achieve better performance and through providing information for decision makers that will enable them to reward and maintain good performance and intervene in cases of poor performance. In practice, as has been repeatedly found, many systems of accountability of this type are subject to several forms of corruption and hence are likely to reduce the sense of responsibility for and quality of performance.

The most common problem is a too-narrow focus on justification through meeting agreed targets for service delivery outputs. In organizations in which this is the case, and in which, additionally, rewards and sanctions for individuals and organizations are tightly tied to the achievement of pre-established targets, goal displacement is highly likely (in goal displacement, people seek to achieve the target even at the expense of no longer achieving the objective). The most notorious example comes from the Vietnam War, during which the emphasis on body counts, used as a proxy for success in battles, led to increased killing of civilians in one-sided and strategically unimportant battles. Public sector examples of this sort of problem abound, but there are also many private sector examples in which senior managers have been rewarded handsomely for achieving specific targets at the cost of the long-term viability of the company.

Another common effect is that what gets measured gets done, as intended, but what is not measured is no longer valued or encouraged. A program may turn to "creaming": selecting easier clients so that targets of throughput or outcomes can be achieved, at the cost of reduced access for those who most need the service. Other important values for the organization, such as cooperation across different units of the organization, may no longer be encouraged because of the emphasis on achieving one's own targets. Finally, there are many reported cases in which such a system encourages data corruption: Reported outcomes are exaggerated or modified to match targets.

Disquiet about the effects of this sort of evaluation is at the heart of concerns about high-stakes testing of children in schools, in which case serious sanctions for children, teachers, and schools follow poor performance in standardized tests.

Even if rewards and sanctions are not tightly tied to these forms of accountability—that is, there are few consequences built into the accountability system— other problems arise: cynicism about the value of monitoring and evaluation and the commitment of decision makers to reasonable decision-making processes.

What are the alternatives? It is not necessary to abandon the notion of being accountable for what has been done but to return to the meaning and focus on systems that both justify and explain what has been done. This requires careful consideration of who is being held accountable, to whom, for what, how, and with what consequences. More thoughtful and comprehensive approaches to accountability should demonstrably support good performance and encourage responsibility. Some have referred to this as *smart accountability.*

The first issue to consider is who is being held accountable. Those being held accountable are most often program managers but could and possibly should include staff, senior managers, and politicians. Politicians often claim that their accountability is enacted through elections, but these are clearly imperfect because they are infrequent, involve multiple issues, and often reflect party allegiances rather than responses to specific issues.

The second issue is to whom these parties are being held accountable and how. They are most often held

accountable to those making funding decisions but could and possibly should be accountable to the community, citizens, service users, consumer advocacy groups, taxpayers, relevant professions, and international organizations for compliance with international agreements and conventions. These different audiences for accountability have been labeled *upwards accountability* and *outwards accountability,* respectively. Parties are most often held accountable through performance indicators shown in regular reports to funders, but may also be held accountable through diverse methods such as annual reports, spot inspections, detailed commentaries on performance indicators, or public meetings and reporting.

Real accountability to citizens would involve making appropriate information accessible to citizens, together with some process for feedback and consequences. Annual reports from government departments, corporate entities, and projects are one method for providing this information but are usually used instead as a public relations exercise, highlighting the positive and downplaying problems. The information availability that has been produced by the Internet has created many more opportunities for reporting information to citizens who have ready access to the Internet if there is a central agency willing to provide the information. So, for example, in Massachusetts, parents seeking to choose a residential neighborhood, and hence a schooling system, can access detailed performance indicators comparing resources spent on education, demographic characteristics, activities undertaken, and test scores in different neighborhoods to help inform their decision. However, there remain difficulties for parents in synthesizing, interpreting, and applying this information to decisions about what would be best for their particular children. Without appropriate analysis and interpretation, there are risks that clients and taxpayers will draw dangerously wrong conclusions. For example, some senior bureaucrats have advocated making surgeons accountable to the public by publishing their rates for complication and death. The obvious problem is that without suitable adjustment for various factors, those surgeons would appear less effective who treated the most severe cases or patients with the poorest health.

The third aspect that needs to be addressed is that for which people are being held accountable. Given the concerns outlined so far, it is clear that consideration must be given to more than simply meeting targets. Other aspects of performance that may need to be

included are coverage (matching actual clients with intended target groups), treatment (providing services in the agreed way and in the agreed amount), fiscal management (spending the money on the agreed inputs, proper controls against fraud), and legal compliance (ensuring procedural adherence to laws, policies, and regulations). It is not reasonable, of course, to expect programs to be able to simultaneously meet unilaterally set targets for all of these. An example would be schools or hospitals that are expected to meet standards for open access to all cases, legal requirements about curriculum or standards of care, and fiscal targets linked to reduced budgets and are punished for not meeting the same outcomes for students or patients as organizations with fewer competing accountabilities.

Accountability needs to be understood not just in terms of reporting compliance and meeting targets but in terms of explaining and justifying legitimate variations, including necessary trade-offs between competing imperatives and accountability. Easy achievement of a timid target can easily be reported. Understandable and legitimate differences between the target and actual performance will require space and a sympathetic audience for a more detailed explanation.

Accountability also needs to go beyond those outcomes that are directly under the control of those who are being held accountable—for example, employment outcomes of school students or long-term family outcomes for children in foster care. Although it is not reasonable to suggest that the performance of these programs and managers should be assessed only by these long-term outcomes, it is also unreasonable to base it only on the completion of units of service that they can totally control. A better form of accountability expects them to be aware of the subsequent causal chain and to be actively seeking to have a beneficial effect on it or to be redeveloping the program so that it is more likely to do so.

The final and most important aspect of an accountability system is the consequences for those providing the reports. Reputational accountability and market accountability, where adverse performance can affect credibility and market pressures, respectively, depend on evidence of performance being available to be scrutinized by relevant parties. More usually, accountability systems focus on reporting discrepancies between targets and performance to funders, the assumption being that they will use this information in future funding and policy decisions. However, accountability

systems rarely provide sufficient information to make it possible for funders to decide if such discrepancies should be followed by decreased funding (as a sanction), increased funding (to improve the quality or quantity of services being provided), or termination of the function.

Accountability requires a much more comprehensive explanation of performance, an incentive system that encourages improvement of performance rather than misreport and distortion of it, and a commitment to address learning as well as accountability. In other words, accountability systems need to be a tool for informed judgment and management rather than a substitute. This is the smart accountability that has been increasingly advocated.

Smart accountability includes demonstrating responsible, informed management; including appropriate risk management, such as cautious trials of difficult or new approaches; and a commitment to identify and learn from both successes and mistakes. The incentive system for accountability needs to reward *intelligent failure* (competent implementation of something that has since been found not to work), discourage setting easy targets, discourage simply reporting compliance with processes or targets, and encourage seeking out tough criticism.

The acid test of a good accountability system is that it encourages responsibility and promotes better performance.

—*Patricia J. Rogers*

Further Reading

American Evaluation Association. (2002). *American Evaluation Association position statement on high stakes testing in preK-12 education.* Retrieved April 23, 2004, from http://www.eval.org/hst3.htm

Mayne, J. (2001). *Science, research, & development: Evaluation workshops. Seminar 1: Performance measurement.* Retrieved April 23, 2004, from http://www.dpi.vic.gov.au/dpi/nrensr.nsf/FID/-341A311BA4C565A7CA256C090017CA16? Open Document#1

Perrin, B. (1999). Performance measurement: Does the reality match the rhetoric? *American Journal of Evaluation, 20*(1), 101-114.

■■ ACCREDITATION

Accreditation is a process and an outcome, and it exists in two main forms. First, there is accreditation offered by an institution or awarding body to individuals on the basis of academic or training credit already gained within the institution or as a direct result of prior or current learning. Second, there is accreditation sought by an institution from an outside awarding body through some form of formal professional review.

The processes of accreditation make credible the autonomous privilege of an organization or body to confer academic or training awards. Furthermore, accreditation is commonly used to form new programs or to open opportunities for a wider adoption of courses leading to the confirmation of awards. In the context of wider participation in education and training, it is not unusual to hear questions about whether academic standards are the same across a nation or state or comparable between institutions capable of awarding academic or training credit. The idea here is not to point the finger at any single phase of education or the quality assurance process but to address difficult questions openly. The reality is a raft of issues concerned with comparing and formally evaluating standards of attainment.

STANDARDS AND PHILOSOPHY

The process of accreditation is a complex course of action that attempts to evaluate standards of attainment. Indeed, the definition and evaluation of attainable, objective standards relevant to the discipline or domain are often part of the expressed goals for accreditation. This argument may appear to ascribe greater worth to evaluating the outcomes of a training process rather than understanding the kinds of personal change and development expected during the training. However, whether we are outcome or process driven, it is important to know what principles and ideals drive accreditation.

As suggested in the foregoing argument, a philosophy of accreditation is more often determined by the definition of the concerns and principles underpinning the process. Focusing on such principles should diminish any undue reliance on evaluation of course outcomes as a single focus of assessment. Indeed, confidence in a single overriding criterion for assessment can create a negative model of noncompliance. For example, close matching of prior experiential learning-to-learning outcomes of particular modules within a program can capture some evidence of equivalence of experience, but it is less robust as evidence of corresponding learning. This approach can result in

the accreditation of the matching process rather than learning and leads us to ask the question, "Are we measuring the comparability of standards, the quality of learning, or the similarity of experience?"

ACCREDITED LEARNING

Accreditation of students' or trainees' learning can apply where awards are credit rated, with all program and award routes leading to awards offered within the institution, calibrated for credit as an integral part of the validation process. The outcome of this accreditation process is credit, a recognizable educational "currency" that students can seek to transfer from one award to another, from one program to another, and from one institution to another. Although credit can be transferred, the marks achieved by the student are not usually transferable.

Accredited learning is defined as formal learning, including learning assessed and credit rated or certificated by the institution or an external institution or similar awarding body and learning that has not been assessed but that is capable of assessment for the purpose of awarding credit.

Credit gained in the context of a named award may be transferred to another named award. However, credit transfer across named awards is not automatic. Transfer of credit from one award route to another is dependent on the learning outcomes being deemed by the institution as valid for the new award. To be recognized as contributing credit to an award, the evidence of the accredited learning must be capable of demonstrating the following:

- Authenticity, by evidence that the applicant completed what was claimed
- Direct comparison, by evidence of a matching of the learning outcomes with those expected of comparable specified modules approved by the university for the award sought
- Currency, by evidence that the learning achieved is in keeping with expectations of knowledge current in the area of expertise required
- Accreditation of experiential learning

Experiential learning is defined as learning achieved through experience gained by an individual outside formalized learning arrangements. An applicant may apply for the award of credit on the evidence of experiential learning. Such evidence must be capable of assessment and of being matched against the learning outcomes of the program for which the applicant is seeking credit.

The normal forms of assessment for the accreditation of experiential learning include the following:

- A structured portfolio with written commentary and supporting evidence
- A structured interview plus corroborating evidence
- Work-based observation plus a portfolio or other record
- Assignments or examinations set for relevant, approved modules or units

FORMAL PROFESSIONAL REVIEW

Institutions that are intent on gaining accreditation must freely and autonomously request a formal evaluation from an awarding organization or body. Accreditation involves an external audit of the institution's ability to provide a service of high quality by comparing outcomes to a defined standard of practice, which is confirmed by peer review.

Current accreditation arrangements rely on the assumption that only bona fide members of a profession should judge the activities of their peers, and by criteria largely or wholly defined by members of the profession. Historically, there are some interesting examples of nonexpert examination of the professional practices of institutions (for example, Flexner's examination of medical schools in the United States and Canada in the early 1900s), but this form of lay review is not at all typical of the way modern forms of formal review for accreditation have grown up.

Rather, accreditation of an institution has largely become a formal process of program evaluation by peers, which, if successful, testifies that an institution or awarding body:

- has a purpose appropriate to the phase of education or domain of training
- has physical and human resources, teaching schemes, assessment structures, and support services sufficient to realize that purpose on a continuing basis
- maintains clearly specified educational objectives that are consistent with its mission
- is successful in achieving its stated objectives
- is able to provide evidence that it is accomplishing its mission

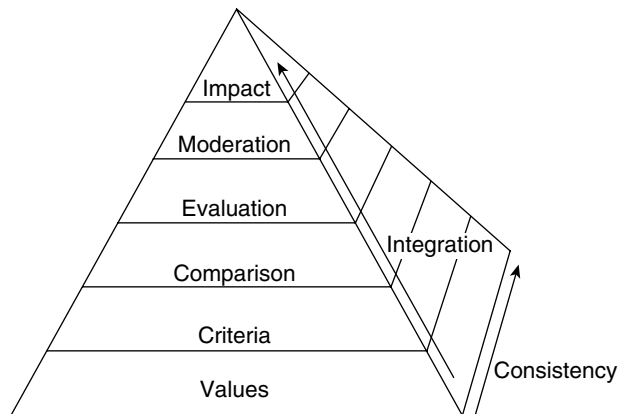**Figure 1**   *Framework of Formal Professional Review*

Figure 2   *Estates of Accreditation*

The requirements, policies, processes, procedures, and decisions of accreditation are predicated on a full commitment to integrity and on institutions dedicating themselves to being learning societies, capable of inclusive and democratic activity. The existence of a single accreditation process in a variety of learning contexts requires adherence to agreed-on evaluation criteria within a common framework of formal program evaluation, and this may not always be possible or desirable. Indeed, it is the development of an adequate accreditation design that is the single most important activity in matching the accreditation process to the context to which it is being applied.

FRAMEWORK OF FORMAL PROFESSIONAL REVIEW

The development of an adequate accreditation design is best placed within a conceptual framework of formal professional review underpinned by a common value system and an agreed-on set of criteria. A model of formal professional review is shown in Figure 1.

Values

Underpinning the review process is an agreed-on set of fundamental principles and values relevant to the awarding body and the institution under review, and these principles and values are shared by all. For example, one of the values might be that the highest duty of all concerned is to preserve life and to labor to improve the quality of life. This is a foundation from which the right to award accreditation is offered and sought.
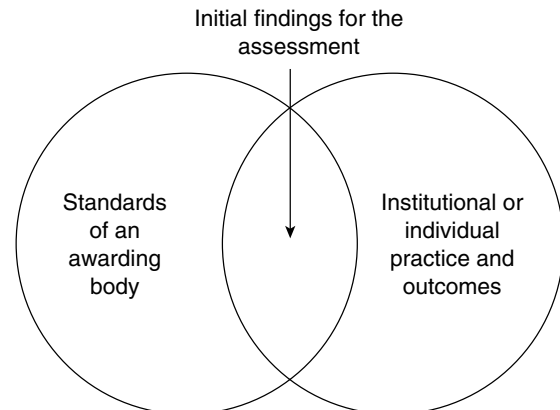
Without such foundations, the review process might well become a somewhat arid inspection of organizational structures, course content, and assessment practices.

Criteria

The formal review is structured around a set of requirements made specific in the form of observable criteria that are grouped in appropriate ways and listed in a convenient form for reference and substantiation. These criteria are embedded in the structure of the value system underpinning the common concern and exist wholly within the parameters of the shared principles. The formulation of criteria affects the organization of the review and the nature of how accordance is recorded.

Comparison

A comparison is made (Figure 2) between two *estates of accreditation:* institutional or individual practice and the standards of the awarding body. Comparisons are made by observation of practice, scrutiny of documentation, and engagement in professional discourse. These processes lie fully within the agreed-on criteria, so when contextual and other interesting matters arise, these are not carried forward into the next stage. The extent to which there is an overlap between the two estates provides initial findings for the assessment.

Evaluation

Reviewers consider the nature and content of the overlap between the two estates and how that might affect their overall judgment. Value is ascribed

and evidence is reviewed to discriminate between discernment, opinion, and critique. Some elements of the comparison may be eliminated from the evaluation process as not pertinent in this context (whether accreditation should be awarded). An argument could be made that evaluation is best carried out by peer professionals, who bring expert experience, knowledge, and understanding to the process, and that this expertise goes beyond commonsense approaches.

Moderation

In this part of the process, the reviewers affirm their judgments and authenticate their findings by grounding them back into the evidence base, verifying the sources of their evidence, and ensuring the rigor of the process that has taken place. If thought necessary or made necessary by some sampling arrangement, alternative opinions are sought and compared with the substantive findings and outcomes. Perceptions are sought of the extent to which practice in assessment and evaluation has been informed by an agreed-on set of fundamental principles and values.

Before the final part of the framework of formal professional review is considered, two dimensions of the framework should be examined that not only form the glue that binds the framework together but make the importance of the impact phase more apparent when they are defined.

Consistency

The consistency dimension is one in which judgments can be made about the extent to which what appears on the face of the institution can be consistently found throughout its structures and organization. At a time when institutions are increasingly positioning themselves as "student facing" or "business facing," it is becoming vital to take a metaphorical slice across the organization to see to what extent attempts to create a compliant façade are supported with scaffolding by the remainder of the organization. Toward the top of the Framework of Formal Professional Review (Figure 1), there is less depth to the organization, and thus approaches to assessment, moderation, and the student experience should be more transparent.

Integration

The division of any process into sections inevitably raises questions about the extent to which judgments

about an institution should contain an element of gestalt. Indeed, the extent to which an organized whole is greater than the sum of its parts is a judgment that needs to be made. The successful integration of the structures, provision, programs, and procedures of any institution is a significant part of its ability to provide evidence that it is accomplishing its mission.

Impact

At the apex of Figure 1 is *Impact.* The impact is often on the individual learner, whose experience is difficult to capture but whose story of personal change and development is pivotal to full understanding of the nature of what takes place in the overlap between the two estates of accreditation. What has happened to this individual and to all the other individuals in the impact section of the framework is of momentous importance to making things better in classrooms, in hospital wards and theaters, in workplaces and homes. It could be argued that securing intelligence of what happens to individual learners and how they are changed by their learning experiences remains an unconquered acme of accreditation.

—Malcolm Hughes and
Saville Kushner

Further Reading
Evans, N. (1992). *Experiential learning: Its assessment and accreditation.* New York: Routledge.
Council for National Academic Awards. (1990). *Accreditation: The American experience.* London: Author.
Fretwell, D. (2003). A framework for evaluating vocational education and training (VET). *European Journal of Education, 38*(2), 177-190.
Stufflebeam, D. L. (2001). The metaevaluation imperative. *American Journal of Evaluation, 22*(2), 183-209.

ACCURACY. *See* Bias, Objectivity, Reliability, Validity

■■ ACHIEVEMENT

In education, *achievement* is the construct of understanding student learning and educational effectiveness. Because what a student knows and can do cannot be measured directly but must be inferred, determining the nature and level of a student's achievement—identifying achievement in real-life cases—is necessarily problematic. Apparent precision in achievement test

scores belies the elusiveness of achievement and maintains public confidence that test scores accurately indicate achievement. Although increasingly prominent since the mid-19th century, standardized high-stakes achievement testing has generated concern and opposition. Thus, as achievement has become ever more crucial in educational accountability and evaluation, determining achievement has become correspondingly controversial.

—Linda Mabry

■ ACTION RESEARCH

The main features of action research are as follows:

- It includes a developmental aim that embodies a professional ideal and that all those who participate are committed to realizing in practice.
- It focuses on changing practice to make it more consistent with the developmental aim.
- In identifying and explaining inconsistencies between aspiration and practice (such explanation may lie in the broader institutional, social, and political context), it problematizes the assumptions and beliefs (theories) that tacitly underpin professional practice.
- It involves professional practitioners in a process of generating and testing new forms of action for realizing their aspirations and thereby enables them to reconstruct the theories that guide their practice.
- It is a developmental process characterized by reflexivity on the part of the practitioner.

From an action research perspective, professional practice is a form of research and vice versa.

Good action research is informed by the values practitioners want to realize in their practice. In social work, for example, it is defined by the professional values (e.g., client empowerment, antioppressive practice) social workers want to realize. Professional values are ideas about what constitutes a professionally worthwhile process of working with clients and colleagues. Such values specify criteria for identifying appropriate modes of interaction. In other words, they define the relationship between the content of professional work, practitioners, and their various clients.

Terms such as *care, education, empowerment, autonomy, independence, quality, justice,* and *effectiveness*

all specify qualities of that relationship. Good action research is developmental; namely, it is a form of reflective inquiry that enables practitioners to better realize such qualities in their practice. The tests for good action research are very pragmatic ones. Will the research improve the professional quality of the transactions between practitioners and clients or colleagues? Good action research might fail this particular test if it generates evidence to explain why improvement is impossible under the circumstances, in which case it justifies a temporary tolerance of the status quo. In each case, action research provides a basis for wise and intelligent decision making. A decision to wait awhile with patience until the time is ripe and circumstances open new windows of opportunity is sometimes wiser than repeated attempts to initiate change.

These are not extrinsic tests but ones that are continuously conducted by practitioners within the process of the research itself. If practitioners have no idea whether their research is improving their practice, then its status as action research is very dubious indeed. It follows from this that action research is not a different process from that of professional practice. Rather, action research is a form of practice and vice versa. It fuses practice and research into a single activity. Those who claim they have no time for research because they are too busy working with clients and colleagues misunderstand the relationship. They are saying they have no time to change their practice in any fundamental sense. When practice strategies are viewed as hypothetical probes into ways of actualizing professional values, they constitute the core activities of a research process, a process that must always be distinguished from research on practice by outsiders.

Action research aims to realize values in practice. Practitioner action research may use outsider research, but it always subordinates the generation of propositional knowledge to the pursuit of practical situational understanding.

ACTION RESEARCH DEVELOPS THE CURRICULUM

Good action research always implies practice development. Practice is never simply a set of statements about the content of activities. It always specifies a mode of interaction. If it is specified in terms of specific behavioral objectives, then the message to clients and

colleagues is that the objectives of professional work define them in terms of deficits and that remedies for these deficits may be found only within a dependency relationship. The question then is whether such practices enable professional workers to represent their work in a form that is consistent with the nature of facilitative, caring, or educational relationships.

Professional practice embodies psychological, sociological, political, and ethical theories. It is through action research that the potential of these theories can be assessed. Through action research, practices are deconstructed and reconstructed in both content and form. Practice designs (plans, guidelines, etc.) need not simply determine practice; rather, by means of action research, they themselves can be shaped through practice. Good practice planning will not only specify the content of action but articulate general principles governing the form in which it is to be enacted. In other words, it should specify action hypotheses in the form of strategies for realizing such principles, which practitioners in general can explore in their particular professional contexts. Such strategies need not be confined to immediate interpersonal processes but can refer to the wider organizational practices that shape social relationships and the amount of time available to various participants for working on certain kinds of tasks. A good practice design not only specifies good professional practice but also provides guidance on how to realize it. The outcome of good action research is not simply improvement in the quality of professional work for those engaged in it but the systematic articulation of what this involves and how others might achieve it. Good action research does not generate private knowledge for an elite core of staff. It renders what they have achieved public and open to professional scrutiny.

ACTION RESEARCH IMPLIES *REFLEXIVE* PRACTICE, NOT SIMPLY *REFLECTIVE* PRACTICE

Good action research generates evidence to support judgments about the quality of practice. The evidence is always about the mode of interaction. Practitioner research that focuses on everything other than the interpersonal conditions established by practitioners is not good action research. Good action research is always reflexive, not merely reflective. One can reflect about all manner of things other than one's own actions. Evidence of client outcomes does not, in isolation, constitute evidence of practice quality. Outcomes need to be explained. The quality of immediate practice activities is only one possible explanation for success or failure. Other kinds of evidence need to be collected before the contribution of the individual practitioner's decision making to client outcomes can be judged. Outcome data may provide a basis for hypothesizing about the nature of this contribution, but the hypotheses will need to be tested against other evidence concerning the professional process. Outcome data are very indirect evidence of quality. Judging the quality of outcomes and the quality of practice are different enterprises.

ACTION RESEARCH INVOLVES GATHERING DATA ABOUT PRACTICE FROM DIFFERENT POINTS OF VIEW

Evidence about the quality of practice can be gathered from a number of sources: practitioners' own accounts of their practice, clients' and colleagues' accounts, peer observations of each others' practice, "outsiders'" observational accounts, and video and audio recordings of professional transactions. This process of gathering data from a multiplicity of sources is called *triangulation.* There are three fundamental sources of evidence: from observers and from the major participants, that is, the practitioner and her or his clients and colleagues. In a fully developed action research process, practitioners will be comparing and contrasting the accounts of observers and clients or colleagues with their own.

ACTION RESEARCH DEFINES RATHER THAN APPLIES QUALITY INDICATORS

These are the sources of evidence, but how do practitioners make sense of this evidence? How do they know what to look for? Do they need a precoded checklist of quality indicators—specific behaviors that are indicative of the qualities they want to realize in their practice? The problem with suggesting that they do is that it preempts and distorts what is involved in doing action research. Quality indicators cannot be predefined because it is the task of action research to define them. When practitioners choose

certain courses of action as a possible means of realizing their educational values in practice, they are exploring the question, "What actions are indicative of those values?" Evidence then has to be collected to determine whether the means selected are indicative of the professional values the practitioner espouses. Practitioners may select a course of action in the belief that it will facilitate the empowerment of clients or colleagues and, in the light of triangulation data, discover this belief to be problematic. Clients may, for example, report that they experience the actions as constraints, and such reports may appear to be consistent with observational accounts. The evidence thus renders the practitioners' actions problematic as quality indicators and challenges practitioners to redefine what constitutes good practice in the circumstances they confront.

Professional quality indicators are determined through action research, not in advance of it. Pre-specifications of quality indicators prescribe what practitioners must do to realize professional values. By standardizing responses, they render the responses insensitive to context and substitute standardized assessments of performance in the place of action research. Good action research acknowledges the fact that what constitutes quality in professional practice cannot be defined independently of the particular set of circumstances a practitioner confronts. It can only be defined *in situ* through action research. Practitioners may, through action research, generalize indicators across a range of contexts, but this outcome provides a source of hypotheses to be tested and not a prescriptive straightjacket that preempts practitioners from ultimately judging what actions are indicative of quality in particular circumstances. Good action research involves grounding such judgments in triangulated case data.

—John Elliott

Further Reading

Argyris, C., Putnam, R., & Smith, D. M. (1985). *Action science: Concepts, methods and skills for research and intervention.* San Francisco: Jossey-Bass.

Carr, W., & Kemmis, S. (1986). *Becoming critical: Education knowledge and action research.* London: Falmer.

Elliott, J. (1991). *Action research for educational change.* Milton Keynes, PA: Open University Press.

Kemmis, S., & McTaggart, R. (1988). *The action research planner* (3rd ed.). Geelong, Victoria, Australia: Deakin University Press.

ACTIVE LEARNING NETWORK FOR ACCOUNTABILITY AND PERFORMANCE IN HUMANITARIAN ACTION (ALNAP)

Established in 1997, the Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP) is an international interagency forum dedicated to improving the quality and accountability of humanitarian action by sharing lessons; identifying common problems; and, where appropriate, building consensus on approaches. ALNAP consists of 51 full members and approximately 300 observer members. Member representatives are drawn from the policy, operations, evaluation, and monitoring sections of organizations involved in humanitarian action.

ACTIVITY THEORY

Activity theory is an approach to psychology based on Marx's dialectical materialism that was developed by revolutionary Russian psychologists Vygotsky, Leont'ev, and Luria in the 1920s and 1930s. The focus of analysis is the activity undertaken by a person (subject) with a purpose (object) that is mediated by psychological tools (systems of numbers, language) and often performed in collaboration with others. An activity occurs in a cultural context that includes conventions, such as cultural rules, and forms of relationships, such as a division of labor.

See also SYSTEMS AND SYSTEMS THINKING

Further Reading

Nardi, B. (1996). (Ed.). *Context and consciousness: Activity theory and human-computer interaction.* Cambridge: MIT Press.

ADELMAN, CLEM

(b. 1942, Tring, England). Ph.D. Science Education, London University; B.Sc. Education, London University.

Adelman began teaching about and doing research and evaluation in 1964. Since 1972 he has worked at

the Center for Applied Research in Education (CARE), University of East Anglia, United Kingdom, and has also held appointments at the University of Reading and the University of Trondheim in Norway.

At CARE, Adelman initially worked on the Ford Foundation Teaching Project, which promoted self-evaluation and classroom action research, and then on an ethnographic study of 3- to 5-year-old children in schools. He became increasingly involved in evaluation problems arising from fieldwork and case studies, especially those in which the evaluators aspired to be democratic in their principles and conduct. He has worked on various evaluations, including those concerning arts education; bilingual schooling in Boston, MA; assessment of oral English; school-industry links; and residential care for the handicapped.

His work on the feasibility of institutional self-evaluation has contributed to a broad understanding of how and why educational organizations engage in evaluation. His book (with Robin Alexander) *Self-Evaluating Institution: Practice and Principles in the Management of Educational Change* reflects these contributions. Some of his other books include *The Politics and Ethics of Evaluation; Guide to Classroom Observation* (with Rob Walker); and the edited volume *Uttering, Muttering: Collecting, Using and Reporting Talk for Social and Educational Research.*

Adelman's work has been influenced by fellow evaluators Bob Stake, Barry McDonald, Ernie House, and Gene Glass, as well Bela Bartok, H. S. Becker, Ernest Gellner, Stephen J. Gould, Nat Hentoff, Georg Simmel, and Spinoza.

Adelman has lived and worked in the south of France, Norway, northwest China, and Hungary. He is an accomplished saxophone player.

ADVERSARIAL EVALUATION. *See* JUDICIAL MODEL OF EVALUATION

■■ ADVOCACY IN EVALUATION

Advocacy, defined in many of the debates as taking sides, is seen as one of the more intractable matters in contemporary evaluation. For example, Linda Mabry notes, "Whether reports should be advocative and whether they can avoid advocacy is an issue which has exercised the evaluation community in recent years. The inescapability [from a constructionist stance] of

an evaluator's personal values as a fundamental undergirding for reports has been noted but resisted by objectivist evaluators [who have] focused on bias management through design elements and critical analysis. At issue is whether evaluation should be proactive or merely instrumental." M. F. Smith, considering the future of evaluation, writes, "The issue (of independence/objectivity versus involvement/advocacy) is what gives me the most worry about the ability of our field to be a profession."

Mabry's and Smith's concerns seem well justified. Granted, the issue of advocacy in evaluation is *complex* in that examining one aspect (for example, the nature of objectivity in social science) leads to another aspect (such as methods of bias control), then another (such as the role of the evaluator as an honest broker, a voice for the disenfranchised, or something else), then another as quickly. The issue is *high stakes* in that the approach taken may lead to different evaluation processes, designs, measures, analyses, conclusions, and, probably, consequences. With a few exceptions, there appears to be little systematic comparative study on this point. The issue of advocacy is *divisive* in our field in that beliefs about advocacy seem deeply and sometimes rancorously held. Last, despite the development of evaluation standards and guidelines, in practice, there is no more than chance agreement and *considerable unpredictability* on what experienced evaluators would do in quite a few ethical situations involving advocacy.

Often, advocacy issues are posed as questions: What is the role of the evaluator? Is impartiality a delusion? In particular, do the Guiding Principles of the American Evaluation Association place a priority on Principle E (Responsibilities for General and Public Welfare) before Principles A through D, if all cannot be satisfied equally? Does credibility require objectivity? These questions were prominent in the 1994 examination of the future of evaluation, based on statements of more than 25 leading evaluators, and they became even more so in the 2000 statements.

PART OF THE CONTEXT

Relatively powerful groups largely fund evaluations—foundations; large philanthropies such as the United Way; local, state, and federal governments; boards and directors of organizations. The questions the evaluator is to help answer are specified initially by the

organizations funding the evaluation. In some instances, so are the evaluation designs (for example, use of randomized control groups to help rule out alternative explanations of results); so are the constructs to be measured (for example, reading readiness); and, at times, so are the measures themselves, particularly the performance indicators.

Speculatively, all might have been well, if, in the early decades of evaluation, most studies showed reasonably positive effects, leading to more services for more people in need. Evaluations, however, often have yielded macronegative results: no measurable evidence of program benefits. Lacking evidence of benefits, programs may be closed down or revised and underlying policies discredited. If one believes in the evaluation results, this is good utilization of evaluation. Truly ineffective programs waste funds and, worse, deprive service recipients of possibly much more effective assistance if something different were tried: If it doesn't work, programmatically, try another approach.

Many of the programs under fairly constant review, revision, or sometimes closure, however, affected low-income people, minorities, and persons of color; many public expenditures benefiting relatively wealthier persons or groups never got evaluated. Evaluations of such programs tended to focus on service delivery, efficiency, and costs rather than effectiveness. In addition, being perceived as an unfriendly critic or executioner is painful. Understandably, evaluators—particularly those at the local level but also many distinguished evaluation theorists—looked carefully at the adequacy of the evaluation frameworks, designs, measures, analyses, and processes. Could the results derived through evaluation approaches seen as biased toward the most privileged groups be trusted? If not, as a matter of social justice and practicality, another approach should be tried, one more likely to "level the playing field." Thus around the mid-1970s and early 1980s, evaluation began to split along the fault line of the possibility or impossibility of objectivity.

THE POSSIBILITY OF OBJECTIVITY

Some evaluators conclude that evaluator neutrality is at the heart of evaluation; that it is necessary for credibility; that meaningful, reliable, valid information about a situation usually is available; and that

frameworks offering objective, trustworthy answers are possible. Recognizing (a) the need and benefits of listening to stakeholders; (b) the value of including, where appropriate, extensive study of process and context that can help explain how observed outcomes came about; and (c) the value of mixed methods, theorists such as Rossi and Freeman, Boruch, Chelimsky, Scriven, and Stufflebeam have emphasized methodologies they see as offering the most certain, least equivocal answers to client questions.

In this view, social good means spending money on programs that work by criteria seen as their raison d'etre by the funders and by others, backed by strong evidence to rule out alternative explanations or rule in plausible causes. One tries to understand the hopes and the fears for the program, to look for unintended as well as intended consequences, to study the program as it actually happens as well as the program as intended. Pragmatically, because stakeholder involvement can make for better evaluations and better utilization, it is an important part of the evaluation process. Also, in this view, evaluators try to understand, explain, and show what was happening, including context, that best accounts for the findings. They should not, however, take the side of any stakeholder group but strive to be impartial sources of reliable information. Theory-driven evaluation and its application in program logic models (Chen); realist evaluation (Henry, Mark, Julnes); frameworks connecting context, input, processes, and results (Stufflebeam); integrative meta-analyses (Shadish, Lipsey, Light); and contemporary experimental designs (Boruch) evolved from these concerns.

THE IMPOSSIBILITY OF OBJECTIVITY

Other evaluators, such as Lincoln and Guba, Mertens, and House and Howe, conclude that evaluations in which the more powerful have an exclusive or primary say in all aspects of the evaluation are inherently flawed, both as a matter of social justice and because of the impossibility of objectivity. Evaluators should be open and up-front in acknowledging their own value stances about programs; should accept that all sources of information about programs will reflect the value stances of the people with whom they are interacting; and, as a matter of responsible evaluation, should give extra weight to the most disenfranchised. Approaches such as Fourth Generation Evaluation

(Lincoln and Guba), Utilization Focused Evaluation (Patton), and Empowerment Evaluation (Fetterman) emphasize meaning as a social construct; place strongest emphasis on diverse stakeholder involvement and control, particularly the participation of service deliverers and service recipients; and see the evaluator role as "friendly critic" or consultant in helping the learning organization examine itself.

In one variant of this general approach, the evaluator becomes an open advocate for social justice as she or he perceives social justice in each case, for the most disenfranchised and the least powerful. This could mean, for example, making the best possible case for the continuation of a program offering what the evaluator saw as social benefits, such as employment of low-income persons as program staff. It could mean that negative evidence potentially leading to loss of jobs or services for low-income persons might not be reported. It could mean that evidence of what the evaluator perceives as a social injustice (an interviewee tells of an instance of sexual harassment) is reported and the program director threatened, when the line of inquiry was not part of the original study. In other words, the evaluator takes sides for or against the interests of certain, primarily less powerful groups before, during, and after the evaluation.

Ahn offers some information on evaluators' "program entanglements" as they affect practice decisions. Based on interviews of a range of evaluators, Ahn reports,

> Many evaluators considered and identified "the least powerful or marginalized group" in the program, whose voices are rarely heard, as crucial in their work. And their allegiance to this group was addressed in a variety of ways in their program. DM, who viewed "emancipation" as the role of her evaluation, for example, spoke of the importance of evaluators being attentive to the needs of "people who could be hurt the most," giving voices to them and promoting their participation. R was also concerned with questions of power. For example, in reporting her evaluation findings, she devoted the largest section in her report to presenting the perspectives of the least powerful who had the most to lose.

Where the voices of contrasting evaluators can be heard on same evaluation situation, as in Michael Morris' ethics in evaluation series in the *American Journal of Evaluation,* the powerful urge to administer on-the-spot social justice (as an advocate for an individual, program, or principle) is perhaps startlingly clear in ways the theorists probably had not intended.

TOWARD COMMON GROUND

Both broad approaches can be caricatured. No constructionist, for example, advocates making up imaginary data to make a program look good. Most would present negative findings but in ways they believed would be appropriate and constructive, such as informally and verbally. And no "neopositivist" ignores human and contextual factors in an evaluation or mindlessly gets meaningless information from flawed instruments to an obviously biased question.

There is much common ground. The leading proponents of advocacy in evaluation indicate that they mean advocacy for the voices of all stakeholders, not only the most powerful, and advocacy for the quality of the evaluation itself. They are concerned with fairness to all stakeholders in questions, designs, measures, decision making, and process, not with taking sides before an evaluation begins. And so are leading proponents of more positivist approaches.

Nonetheless, too little is known about actual decisions in evaluation practice. Recently, some evaluation organizations have been making all aspects of their studies transparent, even crystalline, from raw data to final reports (for example, the Urban Institute's New Federalism evaluation series). Articles focused on specific evaluations and specific evaluators, such as the *American Journal of Evaluation's* interviews with Len Bickman and with Stewart Donaldson, are showing how some evaluators work through practice choices. This kind of clarity should help us sort out conflicts among standards and principles and establish some agreed-on instances, in a sort of clinical practice, case-law sense.

There seems to be considerable agreement that as evaluators, our common ground is fairness. Our warrant is our knowledge of many ways of fairly representing diverse interests, understanding complexity and context, and wisely presenting what can be learned from a systematic, data-based inquiry. Our need is to anchor our debates on neutrality and advocacy in analysis of specific practice decisions so all can understand what constitutes nobly seeking fairness and what, notoriously exceeding our warrant.

—Lois-ellin Datta

Further Reading

Ahn, J. (2001, November). *Evaluators' program entanglements: How they related to practice decisions.* Paper presented at the annual meeting of the American Evaluation Association, St. Louis.

Datta, L.-e. (1999). The ethics of evaluation neutrality and advocacy. In J. L. Fitzpatrick & M. Morris (Eds.), Current and emerging ethical issues in evaluation. *New Directions for Evaluation, 82,* 77-88.

House, E. R., & Howe, K. R. (1998). The issue of advocacy in evaluations. *American Journal of Evaluation, 19*(2), 233-236.

Mabry, L. (2002). In living color: Qualitative methods in educational evaluation. In T. Kelleghan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 167-185). Dordrecht, The Netherlands: Kluwer Academic.

Scriven, M. (1997). Truth and objectivity in evaluation. In E. Chelimsky & W. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 477-500). Thousand Oaks, CA: Sage.

Smith, M. F. (2002). The future of the evaluation profession. In T. Kelleghan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 373-386). Dordrecht, The Netherlands: Kluwer Academic.

■■ AESTHETICS

Aesthetics is a field of study within the discipline of philosophy that, at least currently and during much of the 20th century, has addressed questions about the nature and function of art. Aesthetic theory became part of the discourse within the evaluation field primarily through the work of Elliot Eisner. Eisner used aesthetic theory from a number of philosophers (such as Susanne Langer and John Dewey) to conceptualize and create a justifying rationale for an approach to evaluation based on art criticism. Among other things, Eisner's connoisseurship-criticism model emphasized the use of literary techniques to capture and communicate the aesthetic dimensions of the phenomena being evaluated.

—*Robert Donmoyer*

See also CONNOISSEURSHIP

Further Reading

Dewey, J. (1980). *Art as experience.* New York: Perigee Books.

Langer, S. (1957). *Problems of art.* New York: Scribner.

■■ AFFECT

Affect refers to observable behavior or self-reports that express a subjectively experienced feeling or emotion. *Affect* and *attitude* are sometimes used interchangeably. Whether affect is independent of cognition is a debated topic. On the one hand is the assertion that people can have a purely emotional reaction to something without having processed any information about it. On the other hand is the assertion that at least some cognitive (albeit not always conscious) processing is necessary to evoke an emotional response. Affective measures are used in evaluation. Self-reports of affect provide information on preferences, although indirect measures of affect are sometimes considered more robust; that is, there is likely to be more stability shown in what people are observed to do than in what they say they do. Because interventions may focus specifically on changing people's affect (for example, emotional responses to gender, race, environmentalism, and so on), it is important for evaluators to establish valid indicators of affect.

■■ AGGREGATE MATCHING

In conditions where participants cannot be randomly assigned to program and control groups, the use of proper nonrandomized control groups is recommended to more accurately assess the effects of the independent variable under study. *Aggregate matching* is a procedure for devising matched controls in quasiexperimental evaluation research. Individuals are not matched, but the overall distributions in the experimental and control groups on each matching variable are made to correspond. For example, as a result of this procedure, similar proportions of characteristics such as gender and race would be found in both the program and comparison groups.

—*Marco A. Muñoz*

■■ AID TO FAMILIES WITH DEPENDENT CHILDREN (AFDC)

Building on the Depression-era Aid to Dependent Children program, Aid to Families with Dependent Children (AFDC) provided financial assistance to needy families from the 1970s to 1996. The federal government provided broad guidelines and program requirements, and states were responsible for program formulation, benefit determinations, and administration. Eligibility for benefits was based on a state's

standard of need as well as the income and resources available to the recipient. In 1996, the Personal Responsibility and Work Opportunity Reconciliation Act replaced the AFDC program with the Temporary Assistance for Needy Families program.

—Jeffrey G. Tucker

ALBÆK, ERIK

(b. 1955, Denmark). Ph.D. and M.A. in Political Science, University of Aarhus, Denmark.

Albæk is Professor of Public Administration, Department of Economics, Politics and Public Administration, Aalborg University, Denmark. Previous appointments include Eurofaculty Professor of Public Administration, Institute of International Relations and Political Science, University of Vilnius, Lithuania, and Associate Professor of Public Administration, University of Aarhus, Denmark. He has been an American Council of Learned Societies Fellow and Visiting Scholar in Administration, Planning, and Social Policy, Graduate School of Education, Harvard University, and in the Science, Technology, and Society program at Massachusetts Institute of Technology. Currently, he is Editor of *Scandinavian Political Studies* and previously was Editor of *GRUS.* Since 1989, he has participated in a variety of cross-national research projects with colleagues throughout Europe and the United States.

His primary contributions to evaluation focus on the history, utilization, and functions of evaluation, as well as decision theory, and his primary intellectual influences are to be found in the works of Herbert Simon, Charles Lindbloom, James March, and Carol Weiss. He wrote *HIV, Blood and the Politics of "Scandal" in Denmark,* cowrote *Nordic Local Government: Developmental Trends and Reform Activities in the Postwar Period,* and coedited *Crisis, Miracles, and Beyond: Negotiated Adaptation of the Danish Welfare State.* In addition, he has authored numerous articles and book chapters in both English and Danish.

ALKIN, MARVIN C.

(b. 1934, New York). Ed.D. Stanford University; M.A.Education, and B.A. Mathematics, San Jose State College.

Alkin was instrumental in helping to shape the field of evaluation through his work on evaluation utilization and comparative evaluation theory. He drew attention to ways of categorizing evaluation theories and provided the discipline with a systematic analysis of the way in which evaluation theories develop. As a professor, Alkin developed novel ways of teaching graduate-level evaluation courses, including the use of simulation and role-playing.

His interest in systems analysis was cultivated through his association with Professor Fred MacDonald in the Educational Psychology Department at Stanford University. His thinking and early writings on cost-benefit and cost-effective analysis were fostered by Professor H. Thomas James, also at Stanford University. He was also influenced by his collegial relationships with other evaluators, such as Dan Stufflebeam, Bob Stake, and Michael Quinn Patton.

Alkin founded and served as the Director of the Center for the Study of Evaluation at the University of California, Los Angeles (UCLA), which gave him the opportunity to expand his thinking about issues related to evaluation theory. He was Editor-in-Chief for the *Encyclopedia of Educational Research* (6th edition) and Editor of *Educational Evaluation and Policy Analysis* (1995-1997). He is also a Founding Editor of *Studies in Educational Evaluation.* He received the American Paul F. Lazarsfeld Award from the American Evaluation Association for his contributions to the theories of evaluation.

He is a grandfather of five grandchildren and a dedicated UCLA basketball fan, having attended the full season every year since 1965.

ALPHA TEST

A term used frequently in software development, *alpha test* refers to the first phase of testing in the development process. This phase includes unit, component, and system testing of the product. The term *alpha* derives from the first letter of the Greek alphabet.

ALTSCHULD, JAMES W.

(b. 1939, Cleveland, Ohio). Ph.D. Educational Research and Development, M.S. Organic Chemistry, The Ohio State University; B.A. Chemistry, Case Western Reserve University.

A professor of education at The Ohio State University, Altschuld has contributed to the field of evaluation, especially in the areas of needs assessment and the evaluation of science education and technology, and has coauthored influential books on both topics. He has worked collaboratively with Ruth Witkin on an approach to needs assessment, a central construct in evaluation practice. Together they have written two books on needs assessment: *Planning and Conducting Needs Assessment: A Practical Guide* and *From Needs Assessment to Action: Transforming Needs Into Solution Strategies.* Along with collaborator David Kumar, Altschuld has developed a model of the evaluation of science education programs, reflected in his edited volume *Evaluation of Science and Technology Education at the Dawn of a New Millennium.* Altschuld has contributed substantially to evaluation as a profession through his position papers on certification of evaluators and in *The Directory of Evaluation Training Programs,* published in 1995.

He considers Ruth Altschuld, his wife, and Belle Ruth Witkin, a colleague, to be the major intellectual influences in his life.

Altschuld received the American Evaluation Association Alva and Gunnar Myrdal Evaluation Practice Award in 2002; in 2000, The Ohio State University College of Education Award for Research and Scholarship; in 1997, the Best Program Evaluation Research Award (with Kumar) from the Society for Information Technology and Teacher Evaluation; in 1990, the Evaluation Recognition Award from the Ohio Program Evaluators' Group; and in 1988, The Ohio State University Distinguished Teaching Award.

Altschuld is a reluctant but regular jogger and a devoted grandfather to Andrew and Lindsay.

AMBIGUITY

Ambiguity refers to the absence of an overall framework to interpret situations. Multiple meanings can exist side by side, and this often causes confusion and conflict. Ambiguity differs from uncertainty. Whereas uncertainty—a shortage of information—can be reduced by facts, new and more objective information will not reduce ambiguity because facts do not have an inherent meaning. Ambiguity poses particular problems for evaluators: How is one to evaluate a policy or program if the underlying concept has no

clear meaning? A possible way out is to conduct a responsive approach that takes ambiguity as a departure point for reflexive dialogues on the evaluated practice.

—*Tineke A. Abma*

Further Reading
Abma, T. A. (2002). Evaluating palliative care: Facilitating reflexive dialogues about an ambiguous concept. *Medicine, Health Care and Philosophy, 4*(3), 259-276.
Weick, K. (1995). *Sensemaking in organizations.* Thousand Oaks, CA: Sage.

AMELIORATION

To engage in evaluation is to engage in an activity that has the potential to improve the evaluand, or the human condition, more generally. Logically speaking, no evaluation in and of itself must necessarily purport to be helpful, and making a value judgment does not entail providing prescriptions, remediation, or amelioration. Michael Scriven has clearly delineated the distinction between doing evaluation, making a value judgment, and making a recommendation. In a theory of evaluation, this is an important distinction. In evaluation practice, however, the making of value judgments and the provision of recommendations becomes blurred because we have come to expect the work of evaluators and the purpose of evaluation to be more than simple rendering of value judgments: The purpose of evaluation is also *to make things better.* Acknowledging the serious logical and conceptual problems of moving from evaluative to prescriptive claims or actions, evaluation is meant to be helpful.

The assertion that evaluation should lead to improvement is in most senses self-evident. What is less evident is how evaluation is expected to contribute to making things better. Indeed, different approaches to evaluation conceptualize *helpfulness* and *improvement* differently. Speaking broadly, amelioration takes the form of progress either through science or through democratic processes. The first sort of amelioration is typical of quasiexperimental, decision-making, and systems analysis approaches to evaluation. The second sort of help is typical of participatory, collaborative, and deliberative approaches to evaluation.

Evaluations that are based on amelioration through science focus on the methods used (quasiexperimental

or at least controlled) because these methods permit the examination of causal hypotheses. It is these causal relationships that are the key to amelioration—if one knows what causes what (e.g., whole-language teaching causes higher reading achievement), then this causal claim can be used to improve programs or services by choices that reflect the causal claim (e.g., adopting whole-language pedagogy). These causal claims might be reflected in their contribution to a general theory (of, say, academic achievement) or program theory (of, say, reading programs).

Evaluations that are based on amelioration through democratic processes assume that the meanings of good and right are socially constructed rather than scientifically discovered. This view suggests that truth claims are not natural causal laws but rather informed, sophisticated interpretations that are tentatively held. These interpretations depend on deliberation and dialogue, and it is this emphasis on evaluation process that flags this perspective of amelioration. Participatory, deliberative, and democratic approaches to evaluation reflect the ameliorative assumption based on faith in inclusiveness, participation, public dialogue, and constructivism as the means to improving or helping. Programs, services, and communities will be better as a result of an evaluation that includes stakeholders in genuine ways, thus enabling self-determination in problem definitions and solutions.

Further Reading

Scriven, M. (1995). The logic of evaluation and evaluation practice. In D. Fournier (Ed.), Reasoning in evaluation. *New Directions for Program Evaluation, 68.*

AMERICAN EVALUATION ASSOCIATION (AEA)

The American Evaluation Association (AEA) is an international professional association of evaluators devoted to the application and exploration of program evaluation, personnel evaluation, technology evaluation, and many other forms of evaluation. The association was formed in 1986 with the merger of the Evaluation Network and the Evaluation Research Society. AEA's mission is to improve evaluation practices and methods, increase evaluation use, promote evaluation as a profession, and support the contribution of evaluation to the generation of theory and knowledge about effective human action. AEA has

approximately 3000 members, representing all 50 states in the United States, as well as many other countries.

AMERICAN INSTITUTES FOR RESEARCH (AIR)

Founded in 1946, the American Institutes for Research (AIR) is a not-for-profit research corporation with a long history of research, evaluation, and policy analysis. AIR's staff of some 800 professionals performs basic and applied research, provides technical support, and conducts analyses using established methods from the behavioral and social sciences. Program areas focus on education, health, individual and organizational performance, and quality of life issues.

—Jeffrey G. Tucker

AMERICAN JOURNAL OF EVALUATION

The *American Journal of Evaluation* (*AJE*) is an official, peer-reviewed journal sponsored by AEA. Between 1986 and 1997, the journal was published under the title *Evaluation Practice.* Prior to 1986, a predecessor publication, *Evaluation News,* was sponsored by the Evaluation Network, one of the two organizations that merged to create AEA. *AJE*'s mission is to publish original papers about the methods, theory, practice, and findings of evaluation. The general goal of *AJE* is to publish the best work in and about evaluation. Blaine Worthen was Editor during the transition from *Evaluation Practice* to *AJE,* and Melvin M. Mark succeeded him in 1999. M. F. Smith and Tony Eichelberger previously served as editors of *Evaluation Practice.*

—Melvin M. Mark

ANALYSIS

Analysis may be defined as the separation of an intellectual or material whole into its constituent parts and the study of the constituent parts and their interrelationships in making up a whole. Analysis has both a qualitative dimension (what something is) and a quantitative dimension (how much of that something

there is). In logic, analysis also refers to the tracing of things to their source, the search for original principles. In evaluation, judging evaluands requires an analysis by identifying important aspects of the evaluand and discerning how much of those aspects is present. The opposite of analysis is synthesis.

■ APPLIED RESEARCH

Applied research refers to the use of social science inquiry methods in situations where generalizability may be limited. Such research provides answers to questions dealing with a delimited group of persons, behaviors, or outcomes. Applied research contrasts with basic research, which has the purpose of addressing fundamental questions with wide generalizability; for example, testing a hypothesis derived from a theory in economics. Both applied and basic researchers can use any of the social science research methods, such as the survey, experimental, and qualitative methods. Differences between the research roles do not relate to methods of inquiry; they relate to the purposes of the investigation. Applied researchers focus on concrete and practical problems; basic researchers focus on problems that are more abstract and less likely to have immediate application.

Evaluation provides many avenues for applied research. For example, an evaluator might perform a needs assessment to determine whether a program aimed at a particular group of clients should be planned and implemented. A community psychologist who surveys directors of homeless shelters to assess the need for a substance abuse counseling program is performing applied research. The range of generalizability is limited to the community being surveyed. The problem being addressed is practical, not theoretical. Formative evaluation activities involve applied research almost exclusively. For example, a director of corporate training might use methods such as surveys, interviews, observations, and focus groups to revise and refine instructional materials. The purpose of the research is to obtain feedback about which aspects of the material should be changed for specific users.

An evaluation can sometimes have dimensions of both applied and basic research. For example, a summative evaluation can have both practical uses and implications for theory. A demonstration project on preschool education could simultaneously reveal the merit of the project and test a theory about the impact of instructional activities on the school readiness of 4-year-olds. Although evaluation is associated more with applied research than with basic research, the latter has strongly influenced some evaluation frameworks. A prime example is theory-driven evaluation, which focuses on structuring the evaluation to test whether predicted relationships among variables are verified by the program. An evaluator using this approach might employ statistical models, such as path analysis, that require a set of hypothesized relationships to be tested with empirical data derived from program participants.

—Joseph M. Petrosko

■ APPRAISAL

Appraisal, sometimes used as a synonym for *evaluation,* refers most specifically to estimating the market or dollar value of an object, such as a property or a piece of art or jewelry. The term is also used in valuing intangibles, such as investments; business and industry value, solvency, and liability; and even psychological traits such as motivation.

■ APPRECIATIVE INQUIRY

Appreciative inquiry is a method and approach to inquiry that seeks to understand what is best about a program, organization, or system, to create a better future. The underlying assumptions of appreciative inquiry suggest that what we focus on becomes our reality, that there are multiple realities and values that need to be acknowledged and included, that the very act of asking questions influences our thinking and behavior, and that people will have more enthusiasm and motivation to change if they see possibilities and opportunities for the future. Appreciative inquiry is based on five principles:

1. Knowledge about an organization and the destiny of that organization are interwoven.

2. Inquiry and change are not separate but are simultaneous. Inquiry is intervention.

3. The most important resources we have for generating constructive organizational change or improvement are our collective imagination and our discourse about the future.

4. Human organizations are unfinished books. An organization's story is continually being written by the people within the organization, as well as by those outside who interact with it.

5. Momentum for change requires large amounts of both positive affect and social bonding—things such as hope, inspiration, and sheer joy in creating with one another.

Appreciative inquiry is often implemented as a "summit" that lasts from 2 to 5 days and includes 20 to 2500 people. During their time together, participants engage in a four-stage process of discovery, dream, design, and destiny, during which they respond to a series of questions that seek to uncover what is working well, what they want more of, and how the ideal might become reality. Appreciative inquiry questions might include, "As you reflect on your experience with the program, what was a high point?" "When did you feel most successful in terms of your contributions to the project?" "What are the most outstanding moments or stories from this organization's past that make you most proud to be a member of this organization?" "What are the things that give life to the organization when it is most alive, most effective, most in tune with the overarching vision?"

Appreciative inquiry and participatory, collaborative, and learning-oriented approaches to evaluation share several similarities. For the most part, they are catalysts for change; emphasize the importance of dialogue and, through questioning, seek to identify values, beliefs, and assumptions throughout the process; are based on the social construction of reality; stress the importance of stakeholder involvement; embrace a systems orientation; and reflect an action orientation and the use of results.

Appreciative inquiry is being used to evaluate a wide variety of programs and services around the world. While some evaluators use appreciative inquiry as an overarching framework (as with utilization-focused or empowerment frameworks), others are adopting appreciative inquiry principles to construct interview protocols and procedures. Using appreciative inquiry for evaluation may be particularly useful (a) for framing and implementing developmental and formative evaluations, (b) as a method to focus an evaluation study, (c) as an interviewing technique, and (d) as a means to increase an organization's commitment to engaging in evaluation work.

—Hallie Preskill

Further Reading

Hammond, S. A. (1996). *The thin book of appreciative inquiry.* Plano, TX: CSS.

Watkins, J. M., & Cooperrider, D. (2000). Appreciative inquiry: A transformative paradigm. *OD Practitioner, 32*(1), 6-12.

Watkins, J. M., & Mohr, B. J. (2001). *Appreciative inquiry: Change at the speed of imagination.* San Francisco: Jossey-Bass.

■ APPROPRIATENESS

In the focus on effectiveness, efficiency, and economy, the appropriateness of programs, projects, policies, and products is often ignored—to the detriment of sound judgments about worth. To evaluate appropriateness, one of two comparisons is made. The program may be compared to the needs of the intended clients, using any of the techniques of needs analysis. Alternatively, the program can be evaluated in terms of its compliance with process. In health, for example, some evaluations focus on appropriate care, including treatment of conditions (heart disease) or events (childbirth). Appropriateness can be determined through expert review of individual cases.

—Patricia J. Rogers

■ ARCHIVES

An *archive* is a place in which past and current records and artifacts of ongoing value are protected and made available to people such as evaluators. Such material often forms part of historic memory and can enhance understandings of cultures, organizations, and programs. Archives may contain primary or secondary sources such as memorabilia (e.g., photographs, documents, letters), equipment, newspaper articles, rare books, minutes of meetings, and records. Locations of archives vary: They can be found in government departments, libraries, museums, newspaper offices, universities, private companies, and religious organizations.

From archives, evaluators may gain valuable insights into organizations and programs that may not have been apparent before and that could not have been discovered in any other way. Frequently, evaluators must go to the archival site, although some archives are now being made available through microfiche and the Internet.

—Rosalind Hurworth

■■ ARGUMENT

In an evaluation context, *argument* is the framework, or methodological reasoning, used to persuade an audience of the worth or value of something. Rarely do evaluations find definitive answers or compelling conclusions; most often, evaluations appeal to an audience's reason and understanding to persuade people that the findings of an evaluation are plausible and actionable. Different audiences want different information from an evaluation and will find different constructions of that information compelling. Evaluators use information and data collected during the evaluation process to make different arguments to different audiences or stakeholders of the evaluation. Evaluation is a process of deriving criteria for what counts as important in the evaluation (either as an expert, in concert with program managers or staff, or through participatory processes for deriving these criteria); collecting, analyzing, and interpreting data; and presenting those analyses and interpretations to audiences of interested parties.

Argumentation (methodological reasoning) is an inherent element of designing and carrying out an evaluation; the points of decision that confront evaluators as they work to understand a program or policy and put together the most compelling description possible of their findings are similar to the decision points necessary to construct a valid argument. Evaluators buttress their arguments with data from the evaluation process. However, persuasion comes into play when presenting those facts to audiences: For example, school boards and parents may find measures of student achievement most compelling; educators and school administrators may find evidence of heightened interest in learning, gathered through student interviews, most compelling. An evaluator's job is to present the information in as unbiased a fashion as possible and make the most credible, persuasive argument to the right audiences. Rather than convince and demonstrate, an evaluator persuades and argues; rather than amass and present evidence that is compelling and certain, an evaluator strives to present evidence that is widely accepted as credible.

—*Leslie K. Goodyear*

Further Reading

House, E. R. (1980). *Evaluating with validity.* Beverly Hills, CA: Sage.

■■ ARTISTIC EVALUATION

Artistic evaluation is a general term that could be used to refer to a number of ideas and activities within the evaluation field. Three possible meanings are discussed here.

EISNER'S ARTISTIC EVALUATION MODEL

Possibly the most obvious referent would be the approach to educational evaluation that Elliot Eisner has fashioned using art criticism as a model. Eisner's connoisseurship-criticism model can be considered an artistic approach to evaluation for at least four reasons.

First, Eisner draws on a number of aesthetic theories, including those of John Dewey and Susanne Langer, to conceptualize and justify his educational connoisseurship and criticism model. Second, the model emphasizes the use of artistic and literary forms of discourse to describe the program that is being evaluated. (Only evocative, somewhat poetic language can capture the aesthetic dimensions of what is being studied, Eisner argues.) Third, the model emphasizes the eclectic use of social science theory to interpret the phenomena being evaluated, and this eclectic use of theory can be thought of as being artistic, at least in a metaphorical sense, because it is not rule governed or systematized. Finally, Eisner argues that social phenomena are like works of art in a number of respects, including their complexity, and he applies John Dewey's dictum about evaluating works of art to evaluating educational and other social phenomena: The worth of complex phenomena cannot be assessed by applying a predetermined standard; rather, the evaluator's judgment must be employed.

ARTISTIC FORMS OF DISPLAYING EVALUATION DATA AND RESULTS

The term *artistic evaluation* might also be used to reference the growing use of art forms and artistic techniques to display evaluation findings. Evaluators' use of alternative forms of data display—including forms that are rooted in the arts—normally is motivated by a desire to communicate more effectively with different evaluation audiences, including audiences that are unlikely to respond positively to (or, for that matter,

even to read) traditional social science–sounding evaluation reports.

An example of an artistic data display technique that has been used to communicate evaluation findings in contracted evaluation studies is readers' theater. Readers' theater is a stylized, nonrealistic form of theater in which actors hold scripts and audience members are asked to use their imaginations to visualize details that, in realistic theater, would be provided for them by the scenery and costume designers. In this respect, readers' theater productions are like radio plays; the difference, of course, is that, in a readers' theater production, the actors are visible to the audience and may engage in some visually observable behavior to symbolize certain significant activities (e.g., an actor might turn his back to the audience to symbolize that he or she has left the scene).

Robert Donmoyer and Fred Galloway used readers' theater in an evaluation project funded by the Ball Foundation. The project assessed the foundation's educational reform project in a Southern California school district. Specifically, the readers' theater data display technique was used to communicate evaluation findings to an audience of teachers and administrators involved with the reform. Teachers and administrators in the district also served as readers and actors in the production of the readers' theater script.

The readers' theater script titled *Voices in Our Heads: A Montage of Ideas Encountered During the Evaluation of the Ball Foundation's Community of Schools Project in the Chula Vista Elementary School District* was constructed from quotations excerpted from interview transcripts. The script frequently juxtaposed different and, at times, antithetical points of view about the foundation-funded reform initiative in the district.

Those present judged the presentation to be both an enjoyable and an effective way to communicate important ideas emerging from the evaluation to an audience that was unlikely to read the rather lengthy written report that the two evaluators had produced. Foundation officials, in fact, were so pleased with this artistic approach to reporting evaluation findings that they contracted with the two evaluators to evaluate the annual conference the foundation sponsored for the schools and school districts it funded and to report the results of that evaluation in another readers' theater production.

Before proceeding, it should be noted that the use of alternative modes of displaying data—including modes rooted in the arts—is hardly a novel idea. Both Robert Stake, in his discussions of his "responsive" approach to evaluation, and Elliot Eisner, in discussing his educational connoisseurship and criticism model of evaluation, endorsed the use of artistic modes of data display in evaluation work as early as the mid-1970s.

THE PROCESS MEANING OF ARTISTIC EVALUATION

There is at least one other meaning that could be associated with the term *artistic evaluation.* The term might refer to the fact that all forms of evaluation have a serendipitous, unplanned element. Even when an evaluator is intent on rigidly following predetermined and prespecified standard operating procedures and making the evaluation process systematized and rule governed, things do not always turn out as planned, and, consequently, an improvisational element always creeps into evaluation work. To state this point another way: Effective evaluation work inevitably requires a degree of artistry, metaphorically speaking.

For some, the artistic dimension of evaluation work is not seen as a necessary evil. Indeed, at times, the artistic element of evaluation work is lauded and brought front and center in discussions about the designs to be employed in evaluation studies. A positive view of the use of artistry in designing and conducting evaluation studies is on display in Michael Quinn Patton's book *Creative Evaluation,* for example. Whether or not one decides to embrace and cultivate the serendipitous aspects of evaluation work as Patton clearly does, these dimensions will be present in some form and to some degree in all evaluation work. In this respect, all evaluations can be considered to be, to a greater or lesser extent, artistic evaluation.

—Robert Donmoyer

See also CONNOISSEURSHIP

Further Reading
Donmoyer, R., & Yennie-Donmoyer, J. (1995). Data as drama: Reflections on the use of readers' theater as a mode of qualitative data display. *Qualitative Inquiry, 1,* 397-408.
Patton, M. Q. (1988). *Creative evaluation.* Thousand Oaks, CA: Sage.

■ ASSESSMENT

From the Greek, "to sit with," *assessment* means an evaluative determination. Roughly synonymous with *testing* and *evaluation* in lay terms, *assessment* has become the term of choice in education for determining the quality of student work for purposes of identifying the student's level of achievement. A more important distinction is between the terms *assessment* and *measurement* because educational constructs such as *achievement,* like most social phenomena, cannot be directly measured but can be assessed.

—Linda Mabry

■ ASSOCIATIONS, EVALUATION

A hierarchy of evaluation organizations is slowly emerging. At the global level are organizations such as the International Organization for Cooperation in Evaluation (IOCE) and the International Development Evaluation Association (IDEAS). The IOCE is a loose coalition of some 50 regional and national evaluation organizations from around the world. The mission of the IOCE is to legitimate and strengthen evaluation societies and associations by promoting the systematic use of evaluation in civil society. Its intent is to build evaluation capacity, develop principles and procedures in evaluation, encourage the development of new societies and associations, procure resources for cooperative activity, and be a forum for the exchange of good practice and theory in evaluation.

The initiative to establish IDEAS arose from the lack of an international organization representing the professional interests and intellectual needs of development evaluators, particularly in transition economies and the developing world. IDEAS seeks to fulfill that need by creating a strong body of committed voluntary members worldwide, particularly from developing countries and transition economies, that will support essential, creative, and innovative development evaluation activities; enhance capacity; nurture partnerships; and advance learning and sharing of knowledge with a view to improving the quality of people's lives.

The regional level of the hierarchy is made up of evaluation organizations that have a geographical focus that spans two or more countries. The regional

evaluation organizations registered to attend the IOCE inaugural assembly include the African Evaluation Association, the Australasian Evaluation Society (AES), the European Evaluation Society, the International Program Evaluation Network (IPEN) of Russia and the Independent States, and the Program for Strengthening the Regional Capacity for Evaluation of Rural Poverty Alleviation Projects in Latin America and the Caribbean. Most, if not all, regional evaluation organizations are based on individual memberships. To a lesser extent they serve as umbrella organizations for the national evaluation organizations that lie within their geographical focus. Like their national counterparts, regional evaluation organizations offer many benefits to their memberships.

The national level of the hierarchy is made up of evaluation organizations that operate throughout a single country. The national evaluation organizations registered to attend the IOCE inaugural assembly include the American Evaluation Association (AEA), Associazione Italiana di Valutazione, Canadian Evaluation Society (CES), Egyptian Evaluation Society, Eritrean National Evaluation Association, Ghana Evaluators Association, Israeli Association for Program Evaluation, IPEN-Georgia, IPEN-Ukraine, IPEN-Russia, Kenya Evaluation Association, Malawi Network of Evaluators, Malaysian Evaluation Society, Rede Brasileira de Monitoramento e Avaliaçao, Réseau nigérien de Suivi Evaluation (ReNSE), Réseau de Suivi et Evaluation du Rwanda, Sociedad Española de Evaluación, Société Française de l'Evaluation, South African Evaluation Network, Sri Lankan Evaluation Association, United Kingdom Evaluation Society, and Zimbabwe Evaluation Society.

The structure of regional and national evaluation organizations arises organically in response to contextual variables that are unique to their respective areas of geographic focus. In democratic countries, a common structure is the association or society. (A notable exception is Utvärderarna, a very loose evaluation network in Sweden.) In nondemocratic countries, informal networks whose memberships are not registered with the government have often been found to be preferable. There also appears to be a natural progression to the development of regional and national evaluation organization structure. National organizations often begin as informal networks. As the organizations mature and contextual variables change, oftentimes the networks begin to formalize. Eventually, some networks take the step of becoming legal

organizations that are formally recognized by their governments.

Regional and national evaluation organizations provide many benefits to their members. Conferences are a common benefit. These meetings provide an opportunity for professional development and networking. Some organizations seek to connect members who have common interests. The American Evaluation Association does this by grouping members into 32 interest groups that deal with a wide variety of evaluation topics. Journals are another common benefit. Journals provide members with news of the profession as well as information about the latest approaches and methods. Some regional and national evaluation organizations (e.g., AEA, AES, CES, IPEN, etc.) have defined ethical codes to guide the conduct of their members. Finally, several regional and national evaluation organizations (e.g., Swiss Evaluation Society, German Evaluation Association, African Evaluation Society, Australasian Evaluation Society) are in various stages of publishing program evaluation standards. In many cases, these standards are based on the Joint Committee for Educational Evaluation's Program Evaluation Standards.

The subnational level of the hierarchy is made up of evaluation organizations that operate within a region, state, or province of a country. Examples of these types of organizations include the Société québécoise d'évaluation de programme, the Societe Wallonne de l'Evaluation et de la Prospective, the province chapters of the CES, and the local affiliates of the American Evaluation Association. The IOCE has created an enabling environment that permits international cooperation at the subnational level. For example, the Michigan Association for Evaluation (one of AEA's local affiliates) and the Ontario Province Chapter of CES have been exploring opportunities for cooperation such as joint workshops and conferences, professional exchanges, shared publications, and so on. It may be at the subnational level that individuals most directly experience the benefits of international cooperation.

—*Craig Russon*

Further Reading

International Organization for Cooperation in Evaluation. (2002). [Home page]. Retrieved April 27, 2004, from http://home. wmis.net/~russon/ioce/

ATTITUDES

Attitude is a predisposition to classify objects, people, ideas, and events and to react to them with some degree of evaluative consistency. Inherent in an attitude is a judgment about the goodness or rightness of a person, thing, or state. Attitudes are mental constructs that are inferred and are manifest in conscious experience of an inner state, in speech, in behavior, and in physiological symptoms. Measurements of attitudes are often evaluation data, as they may reflect responses to an evaluand.

See also AFFECT

AUDIENCE

An *audience* is the identified receiver of the findings or knowledge products (evidence, conclusions, judgments, or recommendations) from the evaluative investigation. One can think of a hierarchy of audiences—primary, secondary, and so on. The primary audience is an individual or group to which the findings are directed during the evaluation. Early identification of a key individual who has influence in an organization can significantly affect the utilization of evaluation findings in that organization, and there is evidence that influential groups can be similarly influential.

—*John M. Owen*

AUDITING

Broadly defined, *auditing* is a procedure in which an independent third party systematically examines the evidence of adherence of some practice to a set of norms or standards for that practice and issues a professional opinion. For several years, this general idea has informed the process of metaevaluation—a third-party evaluator examines the quality of a completed evaluation against some set of standards for evaluation. In addition to this generic way of thinking about the nature of auditing and its relevance for evaluation, more specifically, one can examine the relations between the practices of program evaluation and program and performance auditing at state and national levels. For many years, these activities have existed side by side as distinct practices with different

professional cultures, literatures, and academic preparation (e.g., training in financial and performance auditing or in disciplines of social science research). During the last several decades, each practice has gone through substantial changes that have influenced the dialogue between the two practices on issues of purpose and methodology. As defined by the Comptroller General of the United States, a performance audit is "an objective and systematic examination of evidence of the performance of a government organization, program, activity, or function in order to provide information to improve public accountability and facilitate decision-making." A program audit is a subcategory of performance auditing, in which a key objective is to determine whether program results or benefits established by the legislature or other authorizing bodies are being achieved.

Both evaluation and performance auditing share an interest in establishing their independence and in warranting the credibility of their professional judgments. Furthermore, both practices are broadly concerned with assessing performance. However, some observers have argued that evaluation and performance auditing differ in the ways they conceive of and accomplish that aim. Some of the differences between the two practices include the following: Auditors address normative questions (questions of what is, in light of what should be), and evaluators are more concerned with descriptive and impact questions. Auditors work more independently of the auditee than evaluators do with their clients. Auditors are more exclusively focused on management objectives, performance, and controls than are evaluators. Auditors work with techniques for generating evidence and analyzing data that make it possible to provide quick feedback to auditees; evaluations often (though not always) have a longer time frame. Although both auditors and evaluators base their judgments on evidence, not on impressions, and both rely on an extensive kit of tools and techniques for generating evidence, they often make use of those tools in different ways. For example, auditors plan the steps in an audit, but an evaluator is more likely to establish a study design that may well take into account examining related evaluation studies and their results. Both study designs and reporting in evaluation are likely to include great detail on methods; for example, interview schedules, the process of selecting interviewees, and the conditions of interviewing. Evaluators also often draw on multiple methods of generating and analyzing data, moreso than auditors. Finally, auditors operate under statutory authority; evaluators work as fee-for-service consultants or as university-based researchers. Other observers have argued that the practices of auditing and evaluation, although they often exist independently of one another, are being blended together as a resource pool for decision makers responsible for public programs. In this circumstance, an amalgamated picture is emerging of professional objectives (e.g., placing high value on independence, strict attention to documentation of evidence), purpose (e.g., combining normative, descriptive, and impact questions), and methodologies (e.g., making use of a wide range of techniques).

—Thomas A. Schwandt

Further Reading

Chelimsky, E. (1985). Comparing and contrasting auditing and evaluation: Some notes on their relationship. *Evaluation Review, 9,* 485-503.

Wisler, C. (Ed.). (1996). Evaluation and auditing: Prospects for convergence. *New Directions for Evaluation, 71.*

■■ AUTHENTICITY

Authenticity is defined as a report writer's attempt to present the voices of the evaluands integrated with the context of the action in a way that seems to correspond with and represent the lived experience of the evaluands. Claims of authenticity are best corroborated by the evaluand's supportive feedback on the report, perhaps as part of the pursuit of construct validity. Criteria to be satisfied include whether the report is accurate, has appropriate coverage, and is balanced and fair. Audio and visual recordings of the action do not in themselves provide authenticity when transcribed and otherwise interpreted by the evaluator, for they lack the experiential accounts and are, for all their detail, partial, like all other data. Evaluators who have employed the concept of authenticity in their work include Elliot Eisner, Rob Walker, Terry Denny, and Saville Kushner. The origins of the concept of authenticity can be found in *The Confessions* of J. J. Rousseau and also in the works of Camus and Heidegger.

—Clem Adelman

See also CONTEXT

❖ AUTHORITY
OF EVALUATION

Authority in evaluation is contingent on myriad reciprocal and overlapping considerations, some obvious and some subtle, as well as on starkly different perspectives as to what constitutes evaluation as a professional practice.

WITHIN THE PROFESSIONAL COMMUNITY

Broad distinctions between stakeholder-oriented and expert-oriented evaluation approaches have implications for authority. Expert-oriented evaluation approaches invest authority in a professional evaluator (e.g., Stufflebeam's context-input-process-product, or CIPP, model), a connoisseur (i.e., Eisner's connoisseurship approach), or a group of professionals or technical experts (e.g., evaluation or accreditation teams, advisory and blue-ribbon panels). In contrast, stakeholder-oriented approaches may confer authority on program personnel, participants, and relevant others or may involve shared authority. Less explicitly, a stakeholder-oriented evaluator may retain authority as well as closely attend to stakeholder aims, priorities, and criteria.

An important consideration in the first half of the 20th century has been the variety of approaches to evaluation, roughly corresponding (but not limited) to the so-called models of evaluation, of which there are about a dozen, depending on categorization schemes. Even the question of whether the evaluator is obliged to render an evaluative judgment regarding the quality, worth, merit, shortcoming, or effectiveness of a program depends, in part, on his or her approach. According to some, the evaluator's basic responsibility is to define foci and questions, determine data collection methods, establish analytic criteria and strategies, and report evaluative conclusions (e.g., Scriven's approach). In other approaches, such decisions and activities are undertaken collaboratively with clients. In articulating the continuum (controversially), Scriven has decried as "not quite evaluation" approaches in which the evaluator leaves authority for final judgment to program participants (e.g., Stake's responsive evaluation) and as "more than evaluation" (i.e., consultation) those approaches in which the evaluator continues to work with clients after an

evaluation report to ensure its useful implementation (e.g., Patton's utilization-focused evaluation) and approaches in which overarching goals include improved working relationships among program personnel (e.g., Greene's participatory evaluation) or their improved political efficacy (e.g., Fetterman's empowerment evaluation).

In any approach—perhaps appropriately, perhaps inevitably—some stakeholder aims, priorities, and criteria will be emphasized over others. For example, all approaches are vulnerable to threats of managerialism, the prioritizing of managers' interests and concerns, because managers are often the contracting agents. All approaches are also susceptible to clientism (attempts to ensure future contracts by giving clients what they want) and, to varying degrees, to the emergence of relationships between evaluators and stakeholders. The day-by-day conduct and internal politics of an evaluation demonstrate that authority in evaluation both varies and fluctuates.

Locus of authority is also affected by evaluation purpose. With its general intent to assess program accomplishment at a specific point in time (e.g., the end of a granting agency's designated funding period), summative evaluation may tend toward the investment of authority in an evaluator. By comparison, formative evaluation (monitoring to improve program implementation) may tend toward shared authority, as ongoing use of evaluation findings by program personnel is the aim. Purposes more specific than these broad strokes also shade evaluation authority.

Authority in an evaluation is influenced by the nature of the evaluator's position vis-à-vis the program itself or the organization of which the program is a part. External evaluation by an outside professional contracted for the purpose of conducting a specific evaluation is often considered more credible than internal evaluation conducted by program personnel or by an organization's internal evaluation unit because of the external evaluator's presumed independence and lack of investment in the program. Independence suggests that external evaluation may tend toward locating authority in the evaluator. Internal evaluators may be subject to particularly intense encroachments on their authority. Even where there is formal provision for independence, they must report findings to their superiors and colleagues who, if threatened or dissatisfied, may actively undermine their credibility or status. Internal evaluators may find themselves caught between reporting fully and

accurately, expected in competent practice, and keeping their jobs. This complication of independence suggests that authority may tend to be more diffuse in internal evaluation.

Calls for metaevaluation (e.g., by Scriven and Stufflebeam), especially in large-scale, high-impact, and expensive projects, imply that even well-placed and well-exercised authority in evaluation guarantees neither competent practice nor credibility to clients. Metalevel review of evaluation projects further disperses authority in evaluation, signifying that even the most autocratic evaluator may not be the final authority regarding his or her own evaluation studies.

BEYOND THE PROFESSIONAL COMMUNITY

Clients may feel that authority should be theirs regarding the focus of an evaluation, dissemination of results, and perhaps other matters, such as identifying key contacts and informants and selecting methods (e.g., specifying an interview study or a population to be surveyed). Having commissioned evaluation for their own reasons, clients may insist, blatantly or subtly, on their authority, as they would with other types of outsourced services. Wrangles may ensue over how an evaluation should be conducted and over publication of findings, even after good-faith contract negotiations. Difficulty may be unavoidable when evaluators believe they have an ethical responsibility to report publicly or to right-to-know or need-to-know audiences but clients fear that broad dissemination may damage their programs by revealing hurtful deficiency. In this way, ethical consideration of the public interest may give the public an authoritative influence in evaluation.

Funding agencies frequently require evaluations of the programs they support and constitute important audiences for evaluation reports. Consequently, they may exercise real authority over some aspects of an evaluation, as when their review panels accept or reject proposals on the basis of whether they conform to the funding agency's interests (e.g., the World Bank's interest in economic aspects of educational projects). Even an applicant's awareness of the evaluation designs of successful prior proposals may give funding agencies oblique authority over evaluations.

Where program managers and program funders are distinct entities rather than one and the same and where the interests of program managers and funders clash, their struggles for authority over evaluations are predictable. For example, formative first-year evaluation reporting may be expected to serve program managers' interests in improving the program and simultaneously serve funders' decisions about funding continuation or termination. Full reporting of early program difficulties may be opposed by program managers because of the risks to personnel but be demanded by funders, each tugging for authority over reporting.

AUTHORITY IN MEANING MAKING

Who should determine the criteria or standards of quality against which a program will be judged, and whether the criteria for determining the quality, worth, merit, shortcomings, or effectiveness of a program should be preordinate or explicit, are issues of authority regarding meaning making, framing the interpretation of data and the development of findings. These issues have been debated in the professional community, sometimes as a matter of criteriality. Partly reflecting evaluators' models or approaches to evaluation, the responsibility to identify or devise criteria is seen by some as an obligatory part of the evaluator's role. Alternatively, clients may justifiably feel that the authority to determine criteria of quality should be the responsibility of professional organizations in their respective fields (e.g., early childhood education standards, rehabilitation standards) or should be shared, giving them a right to comment on whether the standards are sufficiently sensitive for their programs or appropriate for programs of their kind.

Other evaluators consider that useful, sensitive criteria are not well captured in formal, explicit statements, even standards promulgated within the program or within its field of endeavor. Ill-advised criteria, as preordinate criteria determined at the start of an evaluation when relatively little is known about the program may be, could mislead an evaluation, most seriously if taken as guideposts from the outset, exercising authority over all aspects of the study from initial focus to final analysis. From this perspective, emergent criteria intuited by the evaluator during the conduct of the evaluation as the program becomes familiar may seem more nuanced, more appropriate, and more useful. However, explicit, preordinate criteria can foster client awareness and encourage client input in ways not available to unstated intuitive criteria, rendering claims about the greater sensitivity of intuitive criteria vulnerable to challenge.

Audiences may unknowingly exercise authority regarding evaluation results as evaluators strive to achieve accessibility in reporting and to encourage appropriate use. Evaluators often try to be conscious of prospective audiences as they design and conduct their studies, considering what types of data may be meaningful to readers and hearers, for example, and, as they analyze data, considering what types of interpretations may be useful and how to present them comprehensibly. Audiences take what they can and what they will from an evaluation report. Their background (educational level, cultural values, personal experiences, socioeconomic status, and the like) will affect what they glean from anything they read. Audiences' varied agendas also affect their understanding and use of information. Accordingly, evaluators are commonly admonished to consider accessible report formats and styles and to try to predict likely uses and misuses of reports.

Ultimately, those empowered to make decisions affecting programs assume authority for the implementation of evaluation-stimulated program changes. Such decisions can occur at many program levels, with one group of stakeholders citing an evaluation report to press for changes opposed by another group of stakeholders citing the same report. For example, from their reading of a report, program managers may move to expand intake of information about their clients' backgrounds to fine-tune service delivery, and beneficiaries may better understand their rights from reading the same report and resist deeper incursions into their privacy. Stakeholders' common failures to implement evaluation findings, failures to implement them appropriately, and attempts to suppress reports and discredit evaluators have raised enduring issues of utility and responsibility. They also raise issues of authority because a stakeholder might know better than an evaluator what would constitute appropriate interpretations of data, appropriate report dissemination, and appropriate implementation of findings. An evaluator, however conscientious regarding appropriate distribution of authority, cannot claim omniscience or prevent error.

POSTMODERN CONCERN ABOUT AUTHORITY

One way to think about these issues is to see them as leading inexorably to troubling postmodern perspectives regarding authority. Postmodernism is characterized by doubt and denial of the legitimacy of the understandings and cultural practices underlying modern societies and their institutions—including evaluation. Postmodernism views science as error prone, methods as value laden, logic and rationality as hopelessly discontinuous with reality, and strategies for social improvement as the beginnings of new failures and oppressions by new authorities.

In its implacable skepticism about truth and knowledge, postmodernism implies that all evaluators abuse their authority, misrepresenting to greater or lesser degree the programs they intend to document. Postmodernists view evaluators (and others) as unable to escape their personal or professional ways of knowing and unable to compel language, if meanings vary across contexts and persons, to convey fully what they think they know. Falling especially on outsiders such as external evaluators, postmodern opposition to metanarratives and to the hidden presence of authors (and their values and agendas) in their texts implies that evaluation reports are necessarily insensitive to insiders, diminishing real persons, who are reconfigured as mere stakeholders. Postmodernism implies condemnation of evaluators' presumptions of authority, as authors whose claims of expertise regarding the quality of the program override and overwrite program participants' understandings, which are based on lived experience. Postmodernism also recognizes that authorizing evaluation is an exercise of power, with evaluators both assuming power and serving the powerful, contributing to a status quo that undeniably privileges some and oppresses others.

In this dark recognition, it is not clear how evaluators, heirs of the 18th-century Enlightenment, might revise their professional practices so as to promote appropriate and legitimate authority in evaluation. Encouraging clients and stakeholders to demand a share or a balance of authority would not necessarily better protect the accuracy, sensitivity, or utility of an evaluation or report. Fuller negotiation or sharing of authority implies the possibility of consensus, but clarifying or redrawing boundaries and expectations might promote dissensus instead. If, for example, the criteria for judging program quality were open to input by various interested parties, the many values dear to the many stakeholders might prove so diverse that standards could not be selected or devised without neglecting or offending some.

Similarly, favoring readerly texts, which explicitly leave interpretive options open to readers for their own meaning making, over writerly texts, disparaged

by postmodernists as improper usurpations of authority by authors, might engender stakeholder confusion and professional reproach, especially from colleagues who hold that provision of evaluative interpretations or findings is requisite to competent practice. When authority for conducting evaluation is dispersed beyond reclamation and divisiveness becomes entrenched, the evaluation of a program can be thwarted. It is not clear that absence of authority would serve stakeholders or society better than modernist evaluation does.

THE RELATIONSHIP BETWEEN AUTHORITY AND RESPONSIBILITY IN EVALUATION

Has the evaluator the authority or the responsibility to decide the focus, methods, findings, and audiences for an evaluation? The considerations offered here regarding this question imply new questions: Has the evaluator the authority or the responsibility to advocate for a particular focus, methods, interpretation, audience, or stakeholder group? Has the evaluator the authority or the responsibility to advocate for evaluation as a profession, for evaluation's professional codes of conduct, or for social science generally? Has the evaluator the authority or the responsibility to advocate for his or her personal views of social improvement or equity?

If the credibility and social utility of the profession are to be preserved, being a good professional may very well require something like professional disinterest rather than advocacy for any group or value. However, if being a good professional requires professional disinterest rather than advocacy for any group or value, then evaluators foreclose on their obligations as citizens by restricting their participation in the democratic fray within programs or within the political contexts in which programs exist. If being a good professional requires professional disinterest rather than advocacy for any group or value, then evaluators similarly foreclose on their obligations as human beings by restricting challenges to inequity they might otherwise take up.

Appropriate exercise of authority in evaluation is not a simple matter but is contingent on methodological views, interpersonal politics, contextual exigencies, and philosophical perspectives, none of which are absolute, unequivocally clear, or static. In a given situation, either an exercise of authority or a refusal to exercise unilateral authority might be justified on similar grounds. Appropriate exercise of authority is ultimately a judgment call in an environment where different judgments can be made and defended—and must be.

—Linda Mabry