5 First Approaches: Expertise and Consumer-Oriented Approaches

Orienting Questions

1.

What are the arguments for and against using professional judgment as the means for evaluating programs?

2.

What are the different types of expertise-oriented approaches? How are they alike and how do they differ?

3.

Why is accreditation of institutions of higher education controversial today? How do these controversies reflect the controversies that frequently arise in many evaluations?

4.

How is the consumer-oriented evaluation approach like the expertise-oriented approach? How is it different?

5.

How do these approaches influence the practice of evaluation today?

Everyone evaluates. As we discussed in Chapter 1, we all form opinions or make judgments about the quality of things we encounter. Such evaluations include everything from the meal we just finished eating or the movie or concert we saw last week to more serious endeavors—the program to help students at risk of dropping out at our high school or the parent contact program for parents new to our school. Our focus here is not on our individual judgments of something, but on evaluations that are more formal, structured, and public. We connect these personal evaluations with the more formal ones here, though, because the earliest evaluation approaches were concerned, almost exclusively, with judging the quality of something. Those judgments were often derived by a group of individuals coming together to consider their criteria and the program or product to be judged.

The first modern-day approaches to evaluation were expertise-oriented and consumer-oriented evaluations. These approaches continue to be used today, though not so widely in the professional evaluation field. However, they have influenced the ways we think of evaluation and its purposes and methods. We will review each briefly, with a focus on the most widely used current method—accreditation—to illustrate the key principles of these approaches and how they affected, and continue to affect, evaluation practices.

The Expertise-Oriented Approach

The expertise-oriented approach to evaluation is probably the oldest type of formal, public evaluation and, as its name implies, it relies primarily on professional expertise to judge the quality of an institution, program, product, or activity. For example, the merits of a leadership training program for school principals could be assessed by experts from various fields including leadership, educational administration, and training who would observe the program in action, examine its materials and underlying theory, perhaps interview some trainers and participants, or, in other ways, glean sufficient information to render a considered judgment about its value.

In another case, the quality of a hospital could be assessed by looking at its special programs, its operating facilities, its emergency room operations, its in-patient operations, its pharmacy, and so on, by experts in medicine, health services, and hospital administration. They could examine facilities and equipment/supplies of the hospital, its operational procedures on paper and in action, data on the frequency and outcomes of different procedures, the qualifications of its personnel, patient records, and other aspects of the hospital to determine whether it is meeting appropriate professional standards.

Although professional judgments are involved to some degree in all evaluation approaches, this one is decidedly different from others because of its direct, open reliance on professional expertise as the primary evaluation strategy. Such expertise may be provided by an evaluator or by subject-matter experts, depending on who might offer most in the substance or procedures being evaluated. Usually one person will not own all of the requisite knowledge needed to adequately evaluate the program, institution, or agency. A team of experts who complement each other are much more likely to produce a sound evaluation.

Several specific evaluation processes are variants of this approach, including doctoral examinations administered by a committee, proposal review panels, site visits and conclusions drawn by professional accreditation associations, reviews of institutions or individuals by state licensing agencies, reviews of staff performance for decisions concerning promotion or tenure, peer reviews of articles submitted to professional journals, site visits of educational programs conducted at the behest of the program’s sponsor, reviews and recommendations by prestigious blue-ribbon 127128panels, and even the critique offered by the ubiquitous expert who serves in a watchdog role.

To impose some order on the variety of expertise-oriented evaluation activities, we have organized and will discuss these manifestations in four categories: (1) formal professional review systems, (2) informal professional review systems, (3) ad hoc panel reviews, and (4) ad hoc individual reviews. Differences in these categories are shown in Table 5.1, along the following dimensions:

1.

Is there an existing structure for conducting the review?

2.

Are published or explicit standards used as part of the review?

3.

Are reviews scheduled at specified intervals?

4.

Does the review include opinions of multiple experts?

5.

Do results of the review have an impact on the status of whatever is being evaluated?

Developers of the Expertise-Oriented Evaluation Approach and Their Contributions

It is difficult to pinpoint the origins of this approach, since it has been with us for a very long time. It was formally used in education in the 1800s, when schools began to standardize college entrance requirements. Informally, it has been in use since the first time an individual to whom expertise was publicly accorded rendered a judgment about the quality of some endeavor—and history is mute on when that occurred. Several movements and individuals have given impetus to the various types of expertise-oriented evaluations.

Elliot Eisner, an early evaluator discussed later in this chapter, stressed the role of connoisseurship and criticism in evaluation, roles that required expertise in the subject matter to be evaluated. James Madison and Alexander Hamilton took on the role of “expert evaluators” in discussing and elaborating on the meaning and merits of the newly proposed Constitution in The Federalist Papers. (They were experts because they were both present and active at the Constitutional Convention that drafted the document. As such, they were also 128129internal evaluators!) Their writings were influential at the time and are still used by jurists in the U.S. courts to interpret the meanings of the Constitution, illustrating the important actions that can come from reasoned judgments by experts about a product. Accreditation of institutions of higher education is the primary present-day application of expertise-oriented evaluations. The New England Association of Schools and Colleges, which granted the first accreditation and continues accreditations for colleges and universities in New England today, began in 1885 when a group of headmasters of preparatory secondary schools began meeting with presidents of colleges in New England to discuss what graduates should know to be prepared for college. Thus, more than 100 years ago, school and college leaders were talking about ways to align their curricula!

Formal Professional Review Systems: Accreditation

Historical Foundations.

To many, the most familiar formal professional review system is that of accreditation, the process whereby an organization grants approval of institutions such as schools, universities, and hospitals. Beginning in the late 1800s, regional accreditation agencies in the United States gradually supplanted the borrowed western European system of school inspections. These agencies became a potent force in accrediting institutions of higher education during the 1930s. Education was not alone in institutionalizing accreditation processes to determine and regulate the quality of its institutions. Parallel efforts were under way in other professions, including medicine and law, as concern over quality led to wide-scale acceptance of professionals judging the efforts of those educating fellow professionals. Perhaps the most memorable example is Flexner’s (1910) examination of medical schools in the United States and Canada in the early 1900s, which led to the closing of numerous schools he cited as inferior. As Floden (1983) has noted, Flexner’s study was not accreditation in the strict sense, because medical schools did not participate voluntarily, but it certainly qualified as accreditation in the broader sense: a classic example of private judgment evaluating educational institutions.

Accreditation in Higher Education Today.

Accreditation in the United States and in many other countries today meets our criteria for an expertise-oriented, formal review system. The systems make use of an existing structure (generally an independent regional or national accreditation organization in the United States or governmental agencies in other countries), standards published by the organization responsible for accreditation, a specified schedule (for example, reviews of institutions every 2, 5, or 10 years), and opinions of multiple experts, and the status of the institution, department, college, or school is affected by the results. Accreditation is an excellent example of expertise-oriented evaluation because it uses people with expertise in the subject matter of the program or institution to form a judgment regarding the quality of the entity to be evaluated. The accreditation of an institution or program provides consumers and other stakeholders with some indication of the quality of the institution, as judged by experts in the field, and may facilitate summative decisions. For example, many students use an institution’s or program’s accreditation status to aid their decisions about whether to apply to or attend an institution or program. Further, the feedback the accreditation process provides to the institution can be used for program and institutional improvement and decision making. Thus, the accreditation process serves a formative purpose as well.

130 131

Accreditation in the United States is most common for institutions of higher education.1 We will spend a little time describing this process because it has recently become quite political and controversial, and even for those readers not involved in accreditation, the arguments illustrate the types of political issues and choices that often arise in any evaluation. These include disagreements over the purpose of the evaluation (formative or summative); the neutrality and independence of the experts or evaluators; the criteria to be used to judge the product and, thus, the data to be collected or reviewed; and the transparency of the process (what should be available to the public or other stakeholders outside the organization). These controversies have emerged as the U.S. Department of Education, which has a stake in accreditation through provision of student loans to accredited institutions, has begun to take issue with the accreditation practices of the independent regional accrediting bodies that have traditionally reviewed colleges and universities for accreditation.

As noted earlier, in many countries, including Germany, the Netherlands, India, and the countries of the United Kingdom, institutions of higher education are required by law to be accredited. Government agencies, generally through a ministry or department of education, conduct the accreditation process. In some countries, such as Canada, there is no accreditation process for higher education, partly because most institutions of higher education are run by the provincial governments and that governance is considered to provide sufficient oversight. In the United States, accreditation evolved in a way that very much mirrors U.S. citizens’ distrust of government. With a desire to minimize government’s role nonprofit or voluntary associations carry out the accreditation tasks often fulfilled by government agencies in other countries.

As noted earlier, the New England Association of Schools and Colleges was the first accreditation organization in the United States. Originally established as a mechanism for dialogue between administrators of secondary schools and leaders of colleges in the region in 1885, it eventually evolved into the accrediting association for colleges and institutions in the region (Brittingham, 2009). Other regional associations followed, with each taking responsibility for accrediting institutions of higher education in their region. Today, there are six regional accrediting organizations in the United States, each pursuing similar activities within their region.2 These associations focus primarily on accrediting institutions of higher education, though often they are also involved in accrediting K–12 schools. Finally, there are many accrediting associations that review programs in particular disciplines rather than entire institutions. For example, the American Bar Association accredits law schools, the Association of American Medical Colleges accredits medical schools, and the National Council for Accreditation of Teacher Education (NCATE) accredits teacher education programs, with the Teacher Education Accreditation Council (TEAC) emerging as a recent competitor to NCATE.

1 Secondary institutions and school districts are occasionally accredited as well. Some states, for example, are moving to review school districts for accreditation and associations such as AdvancED have been formed out of the North Central and Southern accrediting associations for higher education to focus on accrediting K–12 schools. Further, many private schools are accredited. Our focus is on accreditation in higher education because it has been established for the longest period and its traditions, therefore, illustrate much about expertise-oriented evaluation and its controversies.

2 The major regional accrediting associations in the United States are the Middle States Association of Colleges and Schools, the New England Association of Schools and Colleges, the North Central Association of Colleges and Schools, the Northwest Association of Accredited Schools, the Southern Association of Colleges and Schools, and the Western Association of Schools and Colleges. Although other accrediting organizations exist (for example, for religious institutions), these regional accrediting associations are considered the primary accrediting bodies in the United States.

131 132

Accreditation of institutions of higher education by the six regional associations has followed a similar plan and approach, the mission-based approach, since the 1950s. With the mission-based approach, accreditors focus on the extent to which the institution is pursuing and achieving its stated mission. Although each association also has standards for higher education that it uses in the evaluation, the mission-based approach reflects the philosophy of the associations in its evaluations. Barbara Brittingham describes the mission-based approach and the accreditation process in the United States as “unusually focused on the future” to help the institution improve (2009, p. 18).

The Process of Accreditation.

In the first stage of accreditation, the institution prepares a self-study report describing its mission and its progress toward that mission, as well as how the institution meets the standards of the accrediting body. The second major stage is the core of the expertise-oriented approach: a team of peers, faculty, and administrators from other institutions in the region receives the report and conducts a site visit during which they interview faculty, administrators, staff, and students; review institutional records on admissions, course curricula, student satisfaction and outcomes; observe facilities and classrooms, and so forth. Based on their review of the report and their experiences during the site visit, the team, usually three or four experts, writes a report expressing their views regarding the institution, their recommendations concerning its accreditation status, and their suggestions for improvement. The site visit report is then reviewed by a standing commission at the accrediting association, which may amend the conclusions. The commission then presents the final conclusions to the institution.

The process is expertise-oriented in several ways: (a) the association has expertise concerning standards for higher education, the state and status of other institutions, and the practice of accreditation and review; (b) the faculty and administrators who form the site team have expertise in participating in the governance of their own universities and others where they have been employed and receive some training from the association to serve as site reviewers. Therefore, the expertise of the site visit team and the association allows those involved to make use of the standards of the association, their review of the report, and their site visit to form a final judgment of the quality of the institution. This process is a common one followed not only by the regional accrediting organizations but also by the organizations that accredit programs in individual disciplines in higher education and by organizations that accredit other educational institutions, including school districts, private schools, charter schools, secondary schools, vocational schools, and religious schools.

Accreditation Controversies: Accreditation Politicized.

So what can be controversial here? As one author defending the system notes, “Who better, one might ask, to evaluate the quality of a college or university than those who work in the field?” (O’Brien, 2009, p. 2). O’Brien argues that the evaluation and the relationship between the accrediting organizations and the institution should not be adversarial, noting, “The evaluators are not inspectors coming in with their white gloves” (O’Brien, 2009, p. 2). But the history of the controversy traces back to the GI Bill passed by Congress after World War II to provide financial assistance to returning soldiers to attend colleges and universities. The government wanted to ensure that the financial assistance went for worthwhile post secondary educational activities, but did not want to get directly into the business of examining colleges and universities for quality. So, it decided to rely on the independent regional accrediting associations, which were already reviewing colleges and universities, to determine the institutions students could receive financial aid to attend. Today, with increasing costs of higher education and more and more students attending colleges and universities, U.S. loans to students are big business. The government continues to rely on regional accrediting associations to identify the institutions of higher education that are eligible for aid, but has an increasing stake in the quality of those processes given the large amounts of money distributed in student loans and other forms of aid. In addition, the institutions themselves have a large stake in the process, because many students would not attend an institution that is not accredited, for quality and financial aid reasons.

Through the Higher Education Act, originally passed in 1965, the U.S. government influences higher education in many areas, from student loans to access. In recent years, many in the U.S. Department of Education have become concerned that accreditations are not sufficiently rigorous in weeding out schools that are performing poorly. Even proponents of the system note that current regional accreditation in the United States carries a “light touch” compared with government evaluations of higher education conducted in other countries (Brittingham, 2009, p. 18).

In 2005, the U.S. Department of Education appointed the Commission on the Future of Higher Education to study four issues critical to higher education, one of which was accountability. In “The Need for Accreditation Reform,” a paper prepared for that report, Robert Dickeson called the current U.S. system of accreditation, “a crazy-quilt of activities, processes, and structures that is fragmented, arcane, more historical than logical, and has outlived its usefulness. More important, it is not meeting the expectations required for the future” (2006, p. 1). He concluded that “any serious analysis of accreditation as it is currently practiced results in the unmistakable conclusion that institutional purposes, rather than public purposes, predominate” (Dickeson, 2006, p. 3). He recommended that Congress create a National Accreditation Foundation to accredit institutions of higher education. The final report of the Commission, called the Spellings Commission for then Secretary of Education Margaret Spellings, was quite critical of current accreditation processes (U.S. Department of Education, 2006, http://www2.ed.gov/about/bdscomm/list/hiedfuture/reports/final-report.pdf). The report inspired much controversy and discussion in the higher education community, with 133134organizations such as Phi Beta Kappa and the Association of American Colleges and Universities issuing statements both of support and concern regarding the report. The final 2008 amendment of the Higher Education Act ultimately chose to ignore some of these recommendations, but the concerns raised by the Commission will continue (O’Brien, 2009) and, for our purposes, reflect some of the political concerns raised about evaluation today and, in particular, about expertise-oriented evaluation.

The regional accrediting associations see their purpose in evaluating institutions of higher education as primarily formative, helping these institutions improve. They see these goals as the best way to serve institutions, their students, and the public. By helping colleges and universities to improve and better achieve their stated mission, the accrediting associations believe they are helping students to receive a better education. In contrast, the U.S. Department of Education’s emphasis is summative. It is concerned with maintaining the U.S. position in higher education in the world and in providing educated and skilled graduates for the economy of the twenty-first century. The Department and other critics see the purpose of accreditation as providing parents, students, and other consumers with information to help them decide which institutions they should attend and where they should spend their tuition dollars. In other words, accreditation should help these consumers make summative decisions about which institutions to choose. Further, accreditation should help make summative decisions about which institutions should continue. One critic notes that in the 60 years since the GI Bill was passed, “a mere handful of schools have been shut down and those largely for financial reasons . . . Meanwhile, on the accreditors’ watch, the quality of higher education is slipping” (Neal, 2008, p. 26). So, the accrediting associations have developed a process that is most useful for formative evaluation when critics see the primary purpose as summative.

Increasing Emphasis on Outcomes.

Another area of disagreement concerns the factors that should be considered in accreditation. Today, the emphasis in education, and in much of evaluation around the world, is on outcomes and impacts. (See Chapter 2.) The Spellings Commission report notes the following:

 Too many decisions about higher education—from those made by policymakers to those made by students and families—rely heavily on reputation and rankings derived to a large extent from inputs such as financial resources rather than outcomes. Better data about real performance and lifelong learning ability is absolutely essential if we are to meet national needs and improve institutional performance. (U.S. Department of Education, 2006, p. 14)

Just as K–12 education has moved to measuring student learning by focusing almost entirely on the extent to which state standards are achieved, the Spellings Commission would like evaluations of institutions of higher education to rely much more heavily on measures of student outcomes.3 Although regional accrediting associations have begun to require institutions to provide measures of student outcomes and, for accreditations of professional programs, evidence concerning passage of licensing exams or job placements, the regional accreditation process also emphasizes the importance of input and process variables. Input variables include factors such as the quality of faculty, library holdings, IT capacity, classroom space and facilities, student admissions processes and decisions, and other elements that create the academic environment of the institution. Process variables articulated in standards, reviewed in self-reports, and examined by site visit teams include curricula, course requirements, and teaching quality; assistance to students through tutoring, advising, and other mechanisms; faculty-student interactions; internships; and other elements of the learning process. Regional accrediting associations also consider multiple outcomes, including graduation and drop-out rates, time to graduation, knowledge and skills of graduates, and job placements. Accrediting associations argue that they must examine the entire process of higher education to make a valid judgment of the quality of the institution and to provide advice for improvement. Examining only student outcomes does not give the experts in the accreditation process sufficient information to make useful recommendations for how to change the institution, and its inputs and processes, to achieve better outcomes (Murray, 2009).

3 One difference between standards for K–12 education and those for higher education is that the standards for higher education would be national ones, not developed at the state level as K-12 standards are.

134 135

Neutrality, Transparency, and Purpose in Accreditation.

Other criticisms of the current approach concern reviewers’ neutrality or objectivity and the transparency of the process. Evaluations are expected to be based on independent judgments. Such independence is intended to lead to more objective, and hence more valid, judgments of quality. Generally speaking, expertise-oriented evaluators should not be closely affiliated with the institution or product they are judging. For example, we are suspicious of an expert’s endorsement of a product when we know the expert has a financial relationship with the product’s manufacturer. Consider, for example, current discussions of the objectivity of medical research on the effectiveness of a drug when the research is funded by the pharmaceutical company that developed the drug. But accreditation processes make use of peer reviewers who are faculty and administrators from higher education institutions in the region. Accrediting organizations argue that these experts are in the best position to make the judgments and provide the advice institutions need, because they know what can be accomplished in the environment of such an institution—and how to accomplish it. They have worked in it themselves. Critics, however, are concerned that the closeness of the experts to those being judged and possible competition between institutions or departments present serious conflicts of interest that can lead to biased judgments. Judgments as blunt as Flexner’s evaluations of medical schools would not see the light of day, at least in written reports.

Concerns over objectivity are heightened by the lack of transparency in the process. The U.S. Department of Education would like data and reports to be far more open, meaning that they would be available to parents, students, and the public and would contain content that is readily understood by nonexperts. For example, the Spellings Commission advocated tables presenting data on the 135136knowledge and skills of graduates and other outcome measures for various colleges and universities. These tables would be available for the public to use in judging the quality of institutions, and for other colleges to use as benchmarks (U.S. Department of Education, 2006). Accreditors rely on the thick descriptions contained in self-study reports and the accreditation report. Defenders of the current system agree that the system relies heavily on confidentiality but argue that this confidentiality is one of the reasons for its success. Because of it, “institutions can be candid in their self-studies, and teams can be honest in their assessments” (O’Brien, 2009, p. 2). If reports were made public, those writing the self-report would be reluctant to discuss real problems, and accreditation teams would edit their wording for public consumption. Neither would facilitate learning about problems and making recommendations for change.

Thus, accreditation is changing and is controversial. Like many evaluations in recent years, the accreditation of colleges and universities in the United States has moved to an increasing use of mixed methods and a greater focus on outcomes. Controversies concern the purpose of these expertise-oriented evaluations, the stakeholders they serve, the measures that should take priority, the neutrality and objectivity of the judgments of quality, the transparency of the process, and the availability of results to different stakeholders. Regional accrediting associations, which for many years had no competition, are being seriously challenged, not only by the federal government, but also by popular ratings of colleges and universities such as those published by U.S. News and World Report. As a result, accrediting associations are adapting and changing, but, with all their problems, they still remain a useful example of a formal review system using the expertise-oriented evaluation approach.

Other Formal Review Systems.

There are numerous examples of other formal review systems, particularly in education. For many years, the National Council for Accreditation of Teacher Education (NCATE) has been the primary body to accredit teacher education programs. In 2000, this organization began focusing more on outcomes of such programs by examining knowledge and skills of graduates of the program, scores on licensure tests, and evidence that graduates are able to transfer their knowledge and skills to the classroom. The Teacher Education Accreditation Council (TEAC) has emerged as a competitor to NCATE, but with a similar focus on outcomes (Gitomar, 2007; Murray, 2009).

Some states are beginning to develop systems to review and accredit school districts within their state. For example, the Colorado Department of Education began accrediting districts in 1999 and revised the procedures substantially in 2008. The focus is very much on student outcomes and growth, but includes standards concerning “safe and civil learning environments,” and budget and financial management. Reviewers conclude the process by assigning a district a rating at one of six different levels, from accreditation with distinction to probation and nonaccreditation. Like other formal review systems, the Colorado accreditation process for school districts includes published standards, specified schedules for review (annual for districts with lower ratings, 2 to 3 years for districts at higher levels of 136137accreditation), site visits by a team of external experts, and the districts’ status being affected by the results (http://www.cde.state.co.us/index\_accredit.htm).

Informal Review Systems

Many professional review systems have a structure and a set of procedural guidelines, and use multiple reviewers. Yet some lack the published standards or specified review schedule of a formal review system.

A graduate student’s supervisory committee for dissertations, theses, or capstone projects is typically composed of experts in the student’s chosen field and is an example of an informal system within expertise-oriented evaluation. Structures within the university, and/or faculty policies, exist for regulating such professional reviews of competence, but the committee members typically determine the standards for judging each student’s performance. Fitzpatrick and Miller-Stevens (2009) have described the development and use of a rubric to assess students’ performance on capstone projects to complete a master’s program in public administration. But, typically, such criteria do not exist. Instead, the multiple experts on the committee make judgments of the student’s performance, often without discussing their criteria explicitly. And, of course, the status of students is affected by the results.

The systems established for peer reviews of manuscripts submitted to professional periodicals might also be considered examples of informal review systems, though journals’ procedures vary. Many journals do use multiple reviewers chosen for their expertise in the content of the manuscript. Unlike site visit teams for accreditation or members of a dissertation committee, reviewers do not behave as a team, discussing their reviews and attempting to reach consensus. Instead, a structure exists in the form of an editor or associate editor who selects reviewers, provides a timeframe for their reviews, and makes a final judgment about the manuscript based on the individual reviewers’ comments. However, the schedule, like that for a graduate student’s defense of a dissertation or thesis, is based on the receipt of manuscripts, although reviewers are given a specified time period in which to conduct the review. Many journals, but not all, provide reviewers with some general standards. Of course, the status of the manuscript—whether it is published, revised, or rejected—is affected by the review process.

Ad Hoc Panel Reviews

Unlike the ongoing formal and informal review systems discussed previously, many professional reviews by expert panels occur only at irregular intervals when circumstances demand. Generally, these reviews are related to no institutionalized structure for evaluation and use no predetermined standards. Such professional reviews are usually one-shot evaluations prompted by a particular, time-bound need for evaluative information. Of course, a particular agency may, over time, commission many ad hoc panel reviews to perform similar functions without their collectively being viewed as an institutionalized review system.

137 138

Panels to Develop Standards.

Common examples of ad hoc review panels include panels organized in each state in the United States to develop or revise educational standards for a state or school district, funding agencies to judge proposals and make recommendations for funding, and blue-ribbon panels appointed to address particular issues. These ad hoc panel reviews have no routine schedule, but are organized by an agency or organization to receive input from experts on a particular issue. Thus, each of the 50 states has established standards that reflect that state’s expectations regarding what students will know in different subjects at different grades.4 There is considerable variation across the states in their standards, but the standards for each state were originally developed by a panel of experts. These experts typically consist of teachers, educational administrators, policymakers, and experts in the content area. The composition of the committee is intended to include experts with knowledge of the subject matter for which standards are being set and knowledge of the target population. Some sophisticated methods have been developed for the related task of expert committees identifying the cut scores, or scores that divide various test takers into groups based on their performance (Kane, 1995). (See Girard & Impara [2005] for a case study of the cut setting process by an expert panel in a public school district.)

Funding Agency Review Panels.

In the United States, most federal government agencies make use of funding panels—panels of experts in the research area to be funded—to read proposals, discuss them, and make recommendations. Generally, the funding agency has developed criteria for the reviewers and, often, members of the team meet in Washington, DC, or other locations to discuss their reactions and attempt to reach some consensus. But the standards for funding vary from discipline to discipline and with the particular funding emphasis. Nevertheless, in the model of expertise-oriented evaluation, experts are coming together to make a judgment about something. Some funding organizations compose committees whose members have different areas of expertise. Thus, committees to review proposals in education can consist of a mix of educational administrators or policymakers, teachers, and researchers. Likewise, committees that review proposals for community development or action can include research experts in the field as well as community members serving as experts on the particular community and its needs.

Blue-Ribbon Panels.

Blue-ribbon panels are typically appointed by a high-level government official and are intended to provide advice, not on funding, but on how government should address a particular issue. The Commission on the Future of Higher Education, which was discussed earlier in this chapter, was appointed by the U.S. Department of Education in 2005, at a time when the government was concerned with the long-term status of higher education in the United States and needed input from experts in the area. Members of such panels are appointed because of their experience and expertise in the field being studied. They typically are charged with reviewing a particular situation, documenting their observations, and making recommendations for action. Given the visibility of such panels, the acknowledged expertise of panel members is important if the panel’s findings are to be considered credible. At the local level, where ad hoc review panels are frequently used as an evaluative strategy for many endeavors ranging from economic development and environmental policies to school governance, expertise of panel members is no less an issue, even though the reviewers may be of local or regional repute rather than national renown. Although recommendations of ad hoc panels of experts may have major impact, they might also be ignored, since there is often no formalized body charged with following up on their advice.

4 These actions are somewhat in response to the federal legislation commonly known as No Child Left Behind, but many states had developed standards prior to the legislation.

138 139

Ad Hoc Individual Reviews

Another form of expertise-oriented evaluation is the individual, professional review of any entity by any individual selected for his or her expertise to judge the value of the entity and, in some cases, to make recommendations for change or improvement. Employment of a consultant to perform an individual review of some educational, social, or commercial program or activity is commonplace in many organizations.

Educational Connoisseurship and Criticism

In the previous section, we discussed applications of the expertise-oriented approach in which the experts are not necessarily evaluators. They are experts in something else—the content they are judging. Further, these applications are examples of the expertise-oriented approach, but they were formed and exist independent of the professional evaluation community. In other words, we can study these processes as examples of expertise-oriented evaluation approaches, but those in the evaluation community are generally not involved in establishing these activities or in conducting them, as is the case with the other approaches we will discuss. As noted, we have begun our discussion of approaches by focusing on the oldest evaluation approach, one used for centuries before formal program evaluation emerged, to make judgments about important issues.

But, the expertise-oriented approach has also been part of the discussion of evaluation theories. In the early days of evaluation, Elliot Eisner was a key figure in discussing what evaluation should be, and his writings provide the theoretical foundation for the expertise-oriented approach and connect it to the evaluation literature (Eisner, 1976, 1985, 1991a, 1991b, 2004). Alkin and Christie (2004), in their evaluation tree depicting the origins and theories of evaluation, place Eisner, along with Michael Scriven, at the base of the valuing branch because their emphasis was on the valuing role of evaluation—determining the value, the merit or 139140worth, of the thing being evaluated. Eisner drew from the arts to describe his approach to evaluation. His perspective was a useful counterpoint to the emphasis in the 1970s on social science methods and program objectives. We will briefly discuss his concepts of connoisseurship and criticism, the fundamentals of his evaluation approach. These concepts fall within the expertise-oriented approach, because they require expertise in identifying and judging critical components or elements of the thing being evaluated.

The roles of the theater critic, art critic, and literary critic are well known and, in the eyes of many, useful roles. Critics are not without their faults. We may disagree with their views, but their reviews are good examples of direct and efficient application of expertise to that which is judged. Their criticism prompts us to think about the object being evaluated in different ways, even if we continue to disagree with their judgment. That is one goal of a written review or criticism: To prompt us to think about elements of the object that we, as nonexperts, might not have considered. Eisner (1991a) proposes that experts, like critics of the arts, bring their expertise to bear in evaluating the quality of programs in their areas of proficiency. Eisner does not propose a scientific paradigm but rather an artistic one, which he sees as an important qualitative, humanistic, nonscientific supplement to more traditional inquiry methods. He argues that we need to see the thing being evaluated from multiple perspectives and that the emphasis on quantitative, reductionist methods fails to convey many important qualities of the whole. He notes that numbers play a role in educational evaluation, his area of interest, but also limit what we see:

 [W]e should be recognizing the constraints and affordances of any form of representation we elect to use. Just as a way of seeing is also a way of not seeing, a way of describing is also a way of not describing. The tools we employ for noticing have an enormous impact on what it is that we become aware of. If we want a replete, fulsome, generous, complex picture of a classroom, a teacher, or a student, we need approaches to that perception of such phenomena and, in addition, a form of presentation that will make those features vivid. (Eisner, 2004, p. 200)

The key elements of Eisner’s approach are connoisseurship and criticism (Eisner, 1975, 1991b). Connoisseurship is the art of appreciation—not necessarily a liking or preference for that which is observed, but rather an ability to notice, “to recognize differences that are subtle but significant in a particular qualitative display” (Eisner, 2004, p. 200). The connoisseur has developed knowledge of the important qualities of the object and the ability to observe and notice them well and to study the relationships among them. The connoisseur, in Eisner’s view, is aware of the complexities that exist in observing something in real-world settings and possesses refined perceptual capabilities that make the appreciation of such complexity possible. The connoisseur’s perceptual acuity results largely from a knowledge of what to look for (advance organizers or critical guideposts) gained through extensive previous experience, education, and reflection on that experience.

The analogy of wine tasting is used by Eisner (1975) to show how one must have many experiences to be able to distinguish what is significant about a wine, using a set of techniques to discern qualities such as body, color, bite, bouquet, flavor, and aftertaste, to judge its overall quality. The connoisseur’s refined palate and gustatory memory of other wines tasted is what enables him or her to distinguish subtle qualities lost on an ordinary drinker of wine and to render judgments rather than mere preferences. Connoisseurs exist in all realms of life, not solely the gustatory or artistic. Eisner describes a good coach as a connoisseur of the game who, when watching others at the sport, can recognize subtleties that those with less experience would miss: “We see it displayed in blazing glory in watching a first-rate basketball coach analyze the strengths of the opponents, their weaknesses, as well as the strengths and weaknesses of the team that he or she is coaching” (2004, p. 198).

Connoisseurship does not, however, require a public description or judgment of that which is perceived. The public description is the second part of the Eisner approach. “Criticism,” Eisner states, “is the art of disclosing the qualities of events or objects that connoisseurship perceives” (1979a, p. 197), as when the wine connoisseur either returns the wine or leans back with satisfaction to declare it of acceptable, or better, quality. Or, more akin to public evaluation, criticism is when the wine critic writes a review of the wine. Evaluators are cast as critics whose connoisseurship enables them to give a public rendering of the quality and significance of that which is evaluated. Criticism is not a negative appraisal but rather an educational process intended to enable individuals to recognize qualities and characteristics that might otherwise have been unnoticed and unappreciated. Criticism, to be complete, requires description, interpretation, and evaluation of that which is observed. “Critics are people who talk in special ways about what they encounter. In educational settings, criticism is the public side of connoisseurship” (Eisner, 1975, p. 13). Program evaluation, then, becomes program criticism. The evaluator is the instrument, and the data collecting, analyzing, and judging are largely hidden within the evaluator’s mind, analogous to the evaluative processes of art criticism or wine tasting. As a consequence, the expertise—training, experience, and credentials—of the evaluator is crucial, because the validity of the evaluation depends on the evaluator’s perception. Yet different judgments from different critics are tolerable, and even desirable, since the purpose of criticism is to expand perceptions, not to consolidate all judgments into a single definitive statement.

Eisner’s educational criticism focuses on four dimensions that should be portrayed in a criticism: description, development of themes, interpretation, and evaluation. The focus is on expert, and sometimes, detailed description of the factors that are important in judging the quality of the product or program. Obviously, the approach would not be the most direct for clearly establishing cause-and-effect relationships, but it can be useful in helping us to understand the nature of the intervention and the manner in which it leads to different outcomes. As Eisner recently stated, “Educational connoisseurship and educational criticism represent an effort to employ what the arts and humanities as partners with the social sciences have to offer in advancing our understanding of the process and effect of education. In an age of high-stakes testing, it is a perspective we badly need” (Eisner, 2004, p. 202).

Influences of the Expertise-Oriented Approach: Uses, Strengths, and Limitations

Expertise-oriented approaches, generally referred to by other names, are used extensively in the United States and other countries today. Accreditation efforts are changing and expanding. Governments continue to appoint expert commissions to study issues and make recommendations. Often, such commissions help to protect government leaders from the ire of citizens when government needs to address a controversial issue. For example, closing military bases in the United States has been a controversial issue, in spite of the fact that too many bases exist. Congress and the president have resorted to appointing commissions of experts to provide “objective, non-partisan, and independent reviews” of recommendations for major base closures (http:www.brac.gov, homepage). The process has been used five times since the first commission was appointed in 1988, most recently in 2005. Like many blue-ribbon panels, the commissions have included experts in a variety of areas related to the issue. The commissions conduct site visits, seek input from the public and other experts, review information, and make recommendations to the President. The recommendations take effect unless Congress rejects the proposal within 45 days. These commissions have been able to take important actions to improve the efficiency and effectiveness of the placement of military bases.

Collectively, expertise-oriented approaches to evaluation have emphasized the central role of expert judgment, experience, and human wisdom in the evaluative process and have focused attention on such important issues as whose standards (and what degree of transparency) should be used in rendering judgments about programs. Conversely, critics of this approach suggest that it may permit evaluators to make judgments that reflect little more than personal biases. Others have noted that the presumed expertise of the experts is a potential weakness. Those using or contracting for expertise-oriented evaluations should consider carefully the various areas of expertise required for their team of expert judges. Too often the team contains only content experts, people who know various elements of the subject matter to be judged, but may lack experts in the evaluation process itself. The articulation of standards, whether by the contracting organization or by the team of experts, is also important to clarify the criteria and methods used to make the judgments requested. Of course, as Elliot Eisner would argue, experts should look beyond the standards and use their connoisseurship to describe, interpret, and judge the dimensions they know to be important to the quality of the product. But, articulated standards help to introduce some consistency across experts and to facilitate useful discussions among the experts when disagreements do occur.

Eisner’s writings influenced evaluators to think more about the nature of evaluation judgments and the role that experience and connoisseurship can play in helping them to notice important elements of the program or product to be evaluated. However, Eisner did not remain active in the evaluation field, and the approach was used infrequently, generally by his immediate students. Still, we continue to study his writings because of the influences he has had on evaluation practice today. Donmoyer (2005) notes that Eisner’s contributions prompted 142143evaluators to consider different approaches to evaluation and the implications of each. Eisner also provided an important rationale for qualitative methods at a time when quantitative methods dominated the field. His work was useful in prompting us to consider what we notice in an object. Connoisseurs know the important elements of a particular thing and learn how to form educated opinions about those elements. The connoisseurship-criticism approach also has its critics. Following Eisner’s initial proposals, House (1980) issued strong reservations, cautioning that the analogy of art criticism is not applicable to at least one aspect of evaluation:

 It is not unusual for an art critic to advance controversial views—the reader can choose to ignore them. In fact, the reader can choose to read only critics with whom he agrees. A public evaluation of a program cannot be so easily dismissed, however. Some justification—whether of the critic, the critic’s principles, or the criticism—is necessary. The demands for fairness and justice are more rigorous in the evaluation of public programs. (p. 237)

However, more recently, Stake and Schwandt emphasize the importance to evaluation not only of measuring quality but also of conveying quality as it is experienced. Reminiscent of Eisner’s recognition of connoisseurship, they observe that “we do not have good enough standards for recognizing an evaluator’s practical knowledge that arises from a combination of observational skill, breadth of view, and control of bias” (2006, p. 409). They conclude that “as with connoisseurs and the best blue ribbon panels, some of the best examples of synthesizing values across diverse criteria are those that rely on the personal, practical judgment of fair and informed individuals” (2006, p. 409).

The Consumer-Oriented Evaluation Approach

Like the expertise-oriented approach, consumer-oriented evaluation has existed in the practice of individuals making decisions about what to purchase, or trade, for centuries. The approaches are similar in other ways: Their primary purpose is to judge the quality of something, to establish the value, the merit or worth, of a product, program, or policy. Although all evaluations are concerned with determining merit or worth, valuing is the key component of these two approaches.5 Their principal audience is the public. Unlike approaches that will be discussed in other chapters in this section, evaluations relying on these approaches often do not have another audience—a foundation, manager, policymaker, or citizens’ group—who has hired the evaluator to provide them with useful information to make a decision or judgment. Instead, the audience for consumer-oriented and expertise-oriented approaches is a broader one—the purchasing or interested public—and is not directly known to the evaluator. Therefore, the evaluator is the major, often the only, decision maker in the study because he or she does not have other important, direct audiences to serve. But the consumer-oriented approach and the expertise-oriented approach differ dramatically in their methodologies, with the latter relying on the judgments of experts and the arts as a model. On the other hand, consumer-oriented evaluation relies on more transparent and quantitative methods, with the judgment typically being made by an evaluator, a person with expertise in judging things, but not with the particular content expertise of expertise-oriented or connoisseur evaluations.

5 Other evaluation approaches focus on various types of use, such as stakeholder involvement or organizational change, and methodology, such as establishing causality or providing thick descriptions as the central component. These evaluations, too, may ultimately make a judgment of merit or worth, but that judgment, the valuing of the program or product, is not so central to the evaluation approach as it is in expertise-oriented or consumer-oriented evaluation. (See Alkin [2004], Shadish et al. [1991].)

143 144

Popular examples of consumer-oriented evaluations that the reader will know include Consumer Reports and the U.S. News and World Report ratings of colleges and universities, but examples exist around the world. Which? is a magazine and web site in the United Kingdom that serves a mission similar to that of the Consumers’ Union, the sponsor of Consumer Reports and its web site, in the United States. Both organizations act as consumer advocates and test products to provide information to consumers on the effectiveness of various products.

The Developer of the Consumer-Oriented Evaluation Approach

Consumer-oriented evaluations first became important in educational evaluations in the mid to late 1960s as new educational products flooded the market with the influx of funds from the federal government for product development. Michael Scriven is the evaluator best known for prompting professional evaluators to think more carefully about consumer-oriented or product evaluations (1974b, 1991c). Scriven, of course, is known for many things in evaluation, and consumer-oriented or product-oriented evaluations represent only one of his contributions. His most important contributions include making evaluators aware of the meaning and importance of valuing in evaluation (Shadish et al. 1991; Alkin, 2004). He often uses examples of product evaluation in his writing to illustrate the nature of valuing and the process of deriving a value in evaluation. For many years, he considered Consumer Reports to be “an almost flawless paradigm” in product evaluation. However, he has expressed disappointment with their reluctance to discuss and improve their methodology and has recognized PC Magazine and Software Digest as developing more methodologically sound procedures (Scriven, 1991a, p. 281).

Scriven’s approach to determining the value of a product, however, is quite different from Eisner’s connoisseur approach. In fact, Scriven’s critical view of Eisner’s approach illustrates his own priorities. He states that evaluations using the connoisseurship model “may generate a valuable perspective, but it abandons much of the requirement of validity. In particular it is vulnerable to the fallacy of irrelevant expertise, because connoisseurs are at best a bad guide to merit for the novice—and are also affected by the swing of fashion’s pendulum” (Scriven, 1991a, p. 92). 144145So, while Eisner’s model rests on the noticing abilities attained by the connoisseur, Scriven’s methods for product evaluation are not concerned with expertise in the content of the product, but with the evaluator’s expertise in testing and judging key components of the product. Further, although Eisner emphasizes interpreting and evaluating the product, he believes that the value added of his approach is in the description—in helping others perceive, and experience, key elements they may have overlooked. Scriven’s concern is in answering the question, “How good is this product?” To do so, he collects information to judge the product’s performance and that of its competitors on explicit, critical criteria and works to remove subjectivity from the approach. Thus, he notes the procedures used by two consumer-oriented magazines he admires represent a “‘pure testing’ approach, that is, one which minimizes the amount of subjective judgment in a particular case” (Scriven, 1991a, p. 281).

Stake and Schwandt (2006), in a discussion of the importance of evaluators discerning quality, shed some light on the differences in Eisner’s and Scriven’s approaches. They identify two approaches to conceptualizing quality: quality as measured and quality as experienced. Quality as experienced is derived from practical knowledge and personal experience, and is significant, they argue, because it is the means by which many people determine quality. Eisner’s connoisseurship model would appear to be an example of evaluation that builds on such quality, through the eyes and experience of a connoisseur. In contrast, quality as measured is illustrated in Scriven’s logic of evaluation and his method for evaluating products. These include determining the important criteria to consider in evaluating the product, establishing standards for the criteria, examining or measuring the performance of the products and its competitors against the criteria using the standards, and synthesizing the results to determine the quality of the key product. Both views of quality have a role. We have discussed Eisner’s approach. Let us now describe more of Scriven’s model for judging the quality of a product.

Applying the Consumer-Oriented Approach

A key step in judging a product is determining the criteria to be used. In the consumer-oriented model, these criteria are explicit and are presumably ones valued by the consumer. Although Scriven writes about the possibility of conducting needs assessments to identify criteria, his needs assessments are not formal surveys of consumers to determine what they would like. Instead, his needs assessments focus on a “functional analysis” that he writes is “often a surrogate for needs assessments in the case of product evaluation” (Scriven, 1983, p. 235). By functional analysis, Scriven means becoming familiar with the product and considering what dimensions are important to its quality:

 Once one understands the nature of the evaluand, . . . one will often understand rather fully what it takes to be a better and a worse instance of that type of evaluand. Understanding what a watch is leads automatically to understanding what the dimensions of merit for one are—time-keeping, accuracy, legibility, sturdiness, etc. (1980, pp. 90–91)

145 146

Thus, his criteria are identified by studying the product to be evaluated, not by previous, extended experience with the product. Standards, developed next, are levels of the criteria to be used in the measurement and judgment process. They are often created or recognized when comparing the object of the evaluation with its competitors. Since the goal is to differentiate one product from another to inform the consumer about quality, standards might be relatively close together when competitors’ performances on a criterion are similar. In contrast, standards might be quite far apart when competitors differ widely. Standards, of course, can be influenced by factors other than competitors, such as safety issues, regulatory requirements, and efficiency factors that provide common benchmarks.

Scriven’s work in product evaluation focused on describing this process and, in part because identifying criteria can be difficult, in developing checklists of criteria for others to use in evaluating products. His product checklist published in 1974 reflects the potential breadth of criteria that he recommends using in evaluating educational products (Scriven, 1974b). This product checklist, which remains useful today, was the result of reviews commissioned by the federal government, focusing on educational products developed by federally sponsored research and development centers, and regional educational laboratories. It was used in the examination of more than 90 educational products, most of which underwent many revisions during the review. Scriven stressed that the items in this checklist were necessitata, not desiderata. They included the following:

 1. Need: Number affected, social significance, absence of substitutes, multiplicative effects, evidence of need

 2. Market: Dissemination plan, size, and importance of potential markets

 3. Performance—True field trials: Evidence of effectiveness of final version with typical users, with typical aid, in typical settings, within a typical time frame

 4. Performance—True consumer: Tests run with all relevant consumers, such as students, teachers, principals, school district staff, state and federal officials, Congress, and taxpayers

 5. Performance—Critical comparisons: Comparative data provided on important competitors such as no-treatment groups, existing competitors, projected competitors, created competitors, and hypothesized competitors

 6. Performance—Long-term: Evidence of effects reported at pertinent times, such as a week to a month after use of the product, a month to a year later, a year to a few years later, and over critical career stages

 7. Performance—Side effects: Evidence of independent study or search for unintended outcomes during, immediately following, and over the long-term use of the product

 8. Performance—Process: Evidence of product use provided to verify product descriptions, causal claims, and the morality of product use

 9. Performance—Causation: Evidence of product effectiveness provided through randomized experimental study or through defensible quasi-experimental, expost facto, or correlational studies

 146 147

 10. Performance—Statistical significance: Statistical evidence of product effectiveness to make use of appropriate analysis techniques, significance levels, and interpretations

 11. Performance—Educational significance: Educational significance demonstrated through independent judgments, expert judgments, judgments based on item analysis and raw scores of tests, side effects, long-term effects and comparative gains, and educationally sound use

 12. Cost-effectiveness: A comprehensive cost analysis made, including expert judgment of costs, independent judgment of costs, and comparison to competitors’ costs

 13. Extended Support: Plans made for post-marketing data collection and improvement, in-service training, updating of aids, and study of new uses and user data

These criteria are comprehensive, addressing areas from need to process to outcomes to cost. Scriven also developed a checklist to use as a guide for evaluating program evaluations, the Key Evaluation Checklist (KEC) (Scriven, 1991c, 2007). It can be found at <http://www.wmich.edu/evalctr/checklists/kec_feb07.pdf>.

Other Applications of the Consumer-Oriented Approach

Product evaluation is also used by organizations and industries to evaluate products at many different stages. Successful high-technology companies such as Apple have watched and studied consumers’ reactions to iPhones and Apple stores and used these data to make changes in their products, thus using consumer-oriented evaluations for formative purposes to revise their products. Amazon.com undertook a similar process with its electronic book, Kindle. Jonathan Morrell, an evaluator who has worked with industries to conduct many product evaluations, recently described the present-day use of product evaluations in industry. Although Scriven focused on product evaluations for summative, purchasing decisions by consumers, Morrell notes that most product evaluations in industries are formative in nature, as with the examples of Apple and Amazon.com. Evaluations take place through the product’s life cycle from initial design and the production process to marketing and circulation. The stakeholders for the evaluation include not only the managers of the organization and the consumers, but others associated with the product process as well. Morrell gives the example of pilots as a stakeholder for airplanes. Their opinions on human factors issues are important in creating a product that will permit them to perform optimally in flying the plane (Morell, 2005).

Influences of the Consumer-Oriented Approach: Uses, Strengths, and Limitations

As mentioned previously, the consumer-oriented approach to evaluation has been used extensively by government agencies and independent consumer advocates to make information available on hundreds of products. One of the best known 147148examples in education today is the What Works Clearinghouse (WWC), begun in 2002 by the U.S. Department of Education’s Institute for Education Sciences (IES). (See http://ies.ed.gov/ncee/wwc.) WWC is a source for consumer-oriented evaluation information on the outcomes of educational programs and products. Its intent, like the consumer-oriented approach reviewed here, is to help consumers—teachers, school psychologists, and educational administrators—make choices about which educational products to use.

WWC differs dramatically, however, from Scriven’s more comprehensive evaluation process because its criteria for determining program success are confined to program outcomes, and its standards are concerned with research confidence in those outcomes. The stated mission of WWC is “to assess the strength of evidence regarding the effectiveness of the program.”6 Products studied using randomized control trials (RCTs) or regression discontinuity designs, which are viewed by IES as superior for establishing a causal link between the product or program and the outcome, receive the highest ratings. Studies using quasi-experimental designs may be endorsed with reservations. Scriven’s checklists and writings argued for using several different criteria to reflect the elements of the product or program that were critical to successful performance. Although many of Scriven’s criteria concerned outcomes or performance (see his criteria for judging educational products listed previously), his process emphasized a comprehensive appraisal of the product, including need, side effects, process, support for users, and cost, as well as several criteria concerning outcomes or performance. WWC’s standards concern the extent to which the research establishes a causal effect, through preferred designs, between the program or product and the intended outcome. Although we bemoan the narrowing of the range of criteria and the standards to assess those criteria, WWC’s efforts do prompt the potential user to consider the effectiveness of the program in achieving its outcomes and to provide a central location for accessing comparable information on educational programs and products. Educators are currently under much pressure to increase achievement, and products can mislead in their marketing. However, WWC’s efforts to inform the consumer about the demonstrated success of programs and products is today’s most successful application of the consumer-oriented approach in education in terms of visibility and number of users. Consumers can search the web site by area of interest, with topics including Early Childhood Education, Beginning Reading, Middle School Math, Dropout Prevention, and English Language Learners. Many products are judged to have insufficient research evidence for a causal relationship between the product and the outcome. The only information provided on these products is the designation “no studies meeting eligibility standards.” However, for products with studies meeting the eligibility standards, reports provide a brief description of the program or product, the research conducted on it, and a final judgment of its effectiveness at achieving the intended outcome.

6 In an ironic combination of consumer-oriented and expertise-oriented approaches, a blue-ribbon panel was convened in 2008 to determine whether WWC’s review process and reports were “scientifically valid” and “provide accurate information about the strength of evidence of meaningful effects in important educational outcomes.” See http://ies.ed.gov/director/board/pdf/panelreport.pdf. Commenting that their charge was not to review the mission but to determine if the information was valid, the panel concluded that the information provided was valid.

148 149

Another prominent example of the consumer-oriented approach that illustrates the overlap between it and the expertise-oriented approach are the test reviews of the Buros Institute of Mental Measurements. The Institute was founded in 1938 and has been conducting well-respected reviews of educational and psychological tests since that time. It currently produces two series: The Mental Measurements Yearbooks, now in its 17th edition, and Test Reviews Online (see www.unl.edu/buros). The Institute is consumer oriented in that it is “dedicated to monitoring the quality of commercially-published tests . . . promoting appropriate test selection, use, and practice” (http://www.unl.edu/buros/bimm/html/catalog.html, paragraph 1). It is designed to provide consumers with information on the quality of tests used in education and psychology. Each test review provides a brief description of the test and a discussion of its development and technical features, including reliability and validity information, a commentary, a summary, and references. However, the reviews contain elements of the expertise-oriented approach because they are conducted by experts in psychometrics and, although the reviews make use of a prescribed format, the criteria and standards for reviewing each test and its competitors are not explicitly identified as would be done in Scriven’s approach. The Institute encourages its reviewers to use The Standards for Educational and Psychological Testing (1999), developed jointly by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME), as a guide, but the Institute’s primary criterion for providing information on quality are in the selection of its expert reviewers.

Although the consumer-oriented evaluation approach continues to be used by magazines and web sites that review products, the approach is not one that continues to be discussed extensively in the professional evaluator literature. However, Scriven’s writings on product evaluation in the 1970s, as well as Eisner’s writings on connoisseurship and criticism, were important in influencing evaluation in its early stages to consider its role in valuing a program, policy, or product and in considering methods other than traditional social science research methods, for doing so. Each approach influenced evaluation practice today.

Major Concepts and Theories

1. The hallmark of the expertise-oriented evaluation approach is its direct reliance on professional judgment in the area of the program being evaluated. 2. Variations in the types of expertise-oriented evaluations include formal and informal review systems and ad hoc panels or individual reviews. These evaluations vary as to whether they are housed under an existing structure or organization, have 149150published standards that are used to evaluate the program or product, use a predetermined schedule for review, employ single or multiple experts, and directly affect the status of the program. 3. Accreditation systems in higher education, extending to K–12 schools, are a prominent example of the expertise-oriented evaluation approach in the United States and are currently in a process of discussion and change. Differences between the regional accrediting associations in the United States and the federal government concerning the purposes of these evaluations, the nature of the data collected or reviewed (outcomes, process, and inputs), the independence or neutrality of the expert evaluators, and the transparency of the process illustrate many of the controversies and political issues that can arise in expertise-oriented and other evaluations. 4. Elliot Eisner’s educational connoisseurship and criticism model made evaluators more aware of the skills of an expert, or connoisseur, in noticing critical dimensions of a product or program and in using methods outside of traditional social science measurement, especially qualitative methods of observation and description, to provide a complete picture of the program or product. 5. The consumer-oriented evaluation approach differs from the expertise-oriented approach in that it does not rely on content experts, or connoisseurs of the product, but rather on experts in evaluation. The approach is also based more centrally on evaluation logic and quantitative methods. 6. Michael Scriven, who wrote extensively about such evaluations, described the key steps as identifying the important criteria for judging the product or program, developing standards to judge those criteria, collecting information or data, and synthesizing the information to make a final judgment that permits the consumer to compare the product with likely alternatives. 7. Both expertise-oriented and consumer-oriented approaches made evaluators aware of the importance of valuing in their work. It helped them recognize that the central task of evaluation is to make a judgment about the value of a program, product, or policy. The approaches advocate quite different methods for making that judgment and, therefore, each added separately to evaluators’ consideration of qualitative methods and of criteria, standards, and checklists as potential methods for collecting data. 8. Both approaches continue to be used commonly by public, nonprofit, and private organizations and industries, but are not the subject of much writing in professional evaluation today. The absence of evaluation literature on the subject is unfortunate. We hope evaluators will return their attention to these approaches commonly used by others to bring evaluative ways of thinking to the application of the approaches today.

A Case Study

For this chapter, we recommend an interview with Gary Henry on the development of the Georgia school report card in Evaluation in Action, Chapter 7. Although our interviews do not contain any evaluations that explicitly use an expertise-oriented or consumer-oriented approach, this interview illustrates the development of a school report card to be used by consumers, parents, and citizens of Georgia. Some of Dr. Henry’s work is concerned with identifying and developing the multiple criteria to be used on the report card, using research studies and input from surveys of the citizens of Georgia and the advisory council to the evaluation. He discusses this process of identifying criteria in his interview and the means for formatting the information in an accessible, easy-to-use manner, and then disseminating it widely. The journal source is Fitzpatrick, J. L., & Henry, G. (2000). The Georgia Council for School Performance and its performance monitoring system: A dialogue with Gary Henry. American Journal of Evaluation, 21, 105–117.

Suggested Readings

Eisner, E. W. (1991a). Taking a second look: Educational connoisseurship revisited. In M. W. McLaughlin and D. C. Philips (Eds.), Evaluation and education: At quarter century, Ninetieth Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press.

Eisner, E. W. (1991b). The enlightened eye: Qualitative inquiry and the enhancement of educational practice. New York: Macmillan.

Floden, R. E. (1980). Flexner, accreditation, and evaluation. Educational Evaluation and Policy Analysis, 20, 35–46.

O’Brien, P. M. (Ed.). Accreditation: Assuring and enhancing quality. New Directions in Higher Education, No. 145, pp. 1–6. San Francisco: Jossey-Bass.

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

Scriven, M. (2007). Key evaluation checklist. http://www.wmich.edu/evalctr/checklists/kec\_feb07.pdf

U.S. Department of Education. (2006). A test of leadership: Charting the future of U.S. higher education. Washington, DC. <http://www.ed.gov/about/bdscomm/list/hiedfuture/reports/final-report.pdf>

**8 Participant-Oriented Evaluation Approaches**

**Orienting Questions**

1.

Who were the historic leaders in establishing participant-oriented approaches and what did they contribute? What prompted their work?

2.

How might we differentiate among the many different contemporary participative approaches?

3.

How do the practical participative approaches differ from the transformative ones?

4.

Contrast the Practical Participatory Evaluation, Empowerment Evaluation, Development Evaluation, and Deliberative Democratic Evaluation approaches. How do their purposes and methods differ? Under what conditions would you use each?

5.

How are participatory approaches used in practice?

6.

What are the major strengths and limitations of participant-oriented evaluation approaches?

Participant-oriented approaches to evaluation currently include many different models, but their commonality is that they all use stakeholders—people with an interest or “stake” in the program—to assist in conducting the evaluation. They may use them to achieve different goals: stakeholders’ greater understanding and ownership of the evaluation leading to greater use of the results or to empowering stakeholders, building evaluation capacity in the organization, and increasing organizational learning and data-based decision making.

The approaches use stakeholders in quite different ways. Some approaches use them primarily at the beginning and ending stages of the evaluation to help 189190define the evaluation questions and, later, to interpret the findings and make recommendations. Others use stakeholders intensively throughout the process, sometimes having the stakeholders act as the major decision makers while the evaluator serves in the role of a technical consultant, as needed. In this chapter, we will describe how the earlier participant-oriented approaches emerged and the characteristics that continue to influence evaluation approaches today. Then, we will describe some of today’s contemporary participatory approaches including their purposes, principles, and methods. Finally, we will discuss their strengths and weaknesses and how they might be used.

Evolution of Participatory Approaches

In the early years of evaluation in the United States, after it was first mandated by Congress and governmental agencies, most evaluation practitioners relied on traditional social science research methods to determine whether the goals and objectives of programs had been achieved and then provided that information to government policymakers. As early as 1967, though, several evaluation theorists began to react to what they considered to be the dominance of mechanistic and insensitive approaches to evaluation in the field of education. These theorists expressed concerns that evaluators were largely preoccupied with stating and classifying objectives, designing elaborate evaluation systems, developing technically defensible objective instrumentation, and preparing long technical reports, with the result that evaluators were distracted from what was really happening in the programs they were evaluating. Critics of traditional evaluation approaches noted that many large-scale evaluations were conducted without the evaluators ever once setting foot on the participating program site(s). What began as a trickle of isolated comments grew to a deluge that flooded evaluation literature in education and the social sciences. More and more practitioners began publicly to question whether many evaluators really understood the phenomena that their numbers, figures, charts, and tables were intended to portray. An increasing segment of the education and human services communities argued that the human element, reflected in the complexities of the everyday reality of the delivery of programs and the different perspectives of those engaged in providing services, was missing from most evaluations.

Consequently, a new orientation to evaluation arose, one that stressed first hand experience with program activities and settings and involvement of program participants, staff, and managers in evaluation. This general approach, which grew quickly after the early 1970s, was aimed at observing and identifying all (or as many as possible) of the concerns, issues, and consequences the different stakeholders had about the program being evaluated.

Due in large part to a reaction to perceived deficits in other evaluation approaches, this orientation now encompasses a wide variety of more specific 190191approaches that might be generally tied together by their acceptance of the constructivist paradigm, recognizing that there are many perspectives to knowledge and truth and, therefore, to a program and its evaluation (see Chapter 4.) Many of those who contributed to the development and use of participant-oriented approaches to program evaluation prefer naturalistic inquiry methods, as described later in this chapter. Moreover, most advocates of this approach see involving participants in the program, managers, staff, and other key stakeholders as a key principle of good evaluation—hence the descriptor “participant-oriented” as a label for this approach.

Developers of Participant-Oriented Evaluation Approaches and Their Contributions

Robert Stake and His Responsive Approach

Robert Stake (1967) was the first evaluation theorist to provide significant impetus to this orientation in the field of education. His paper, “The Countenance of Educational Evaluation,” with its focus on portraying and processing the judgments of participants, dramatically altered the thinking of evaluators in the next decade. Along with his later development of responsive evaluation (Stake, 1973, 1975a, 1975b), he provided conceptions and principles that guided the evolution of this approach. Stake’s early writings evidenced his growing concern over the dominance of program evaluation by what he saw as parochial, objectivist, and mechanistic conceptions and methods. Guba’s (1969) discussion of the “failure of educational evaluation” then provided further impetus to the search for an alternative to the rationalistic approach to evaluation.

The Countenance of Evaluation.

Stake’s first departure from traditional evaluation was his development of the Countenance Framework (1967). In it, he asserted that the two basic acts of evaluation are description and judgment, the two countenances of evaluation. Evaluations should provide a full description of the program and entity being evaluated and then make a judgment of its merit or worth. To aid the evaluator in organizing data collection and interpretation, Stake created the evaluation framework shown in Figure 8.1.

Using the framework shown in Figure 8.1, the evaluator first determines the rationale for the program, which includes the need the program is intended to serve and relevant features of the program’s development. (See Rationale box at far left.) The descriptive part of evaluation then focuses on first determining program intents (column 1) in regard to program antecedents (inputs, resources, and existing conditions), transactions (program activities and processes), and program outcomes. Through observations at each of these levels, the evaluator begins to thoroughly describe the program and to compare the program intents with the actual observations of the program in action. In the judgment stage, the evaluator explicitly identifies or develops standards (criteria, expectations, performance of 191192comparable programs) for judging program antecedents, transactions, and outcomes and, finally, records judgments made about the antecedent conditions, transactions, and outcomes. The evaluator analyzes information in the description matrix by looking at the congruence between intents and observations, and by looking at the dependencies (contingencies) of outcomes on transactions and antecedents, and of transactions on antecedents. Judgments are made by applying standards to the descriptive data.

FIGURE 8.1 Stake’s Layout of Statements and Data to Be Collected by the Evaluator of an Educational Program

Congruence Congruence Congruence Rationale Antecedents Transactions Outcomes

Source: From “The Countenance of Educational Evaluation” by R. E. Stake, 1967, Teachers College Record, 68, p. 529. Reprinted with permission.

Thus, the countenance structure gave evaluators a conceptual framework for thinking through the data needs of a complete evaluation. In reviewing his countenance paper years later, Stake (1991) noted that it underemphasized the process of describing the evaluation, a shortcoming that he addressed later in his responsive evaluation approach. In fact, it was his descriptive emphasis that was new to evaluation approaches of the times. Stake wanted evaluators to become familiar with the particulars of the programs they were studying and to gain a thorough understanding of them before examining outcomes. The evaluator’s understanding of the antecedents and the transactions would help him or her better interpret the successes or failures in achieving desired outcomes.

Responsive Evaluation.

Stake’s responsive evaluation, introduced in 1973, was more radical. In it, he truly tackled his concerns with the directions evaluation was taking at the time. Greene and Abma noted this in their foreword to a 2001 issue 192193of New Directions for Evaluation that focused on responsive evaluation, its influences, and its current applications and adaptations:

 Stake offered a new vision and rationale for educational and social program evaluation to the then-fledgling evaluation communities. In this vision, evaluation was reframed—from the application of sophisticated analytic techniques that address distant policymakers’ questions of program benefits and effectiveness “on the average,” to an engagement with on-site practitioners about the quality and meanings of their practice. These innovative ideas helped accelerate a transformation of the evaluation enterprise into its current pluralistic character, within which remain multiple and varied legacies of key responsive evaluation principles (2001, p. 1).

Although the seeds of this explication lie in his earlier work, Stake’s subsequent conceptions of responsive evaluation (1973, 1975b, 1978, 1980) are implicitly less formal and explicitly more pluralistic and process focused than his earlier countenance model. The responsive evaluation approach departed from past evaluation approaches foremost in its flexibility and in its responsiveness to the particulars of the evaluation setting and its naturalness. Stake noted that he was not proposing a new approach to evaluation, for “responsive evaluation is what people do naturally in evaluating things. They observe and react” (1973, p. 1). Rather, Stake saw the responsive approach as a means to improve upon and focus this natural behavior of the evaluator. Stake stressed the importance of being responsive to realities in the program and to the reactions, concerns, and issues of participants, rather than being preordinate1 with evaluation plans, relying on preconceptions, and setting formal plans and objectives of the program before gaining a full understanding of the program.

Stake defined responsive evaluation as follows:

 An educational evaluation is responsive evaluation if it orients more directly to program activities than to program intents; responds to audience requirements for information; and if the different value perspectives present are referred to in reporting the success and failure of the program. (1975a, p. 14)

Responsive evaluation differed from existing evaluation approaches in many ways that foreshadowed today’s participatory approaches. These included:

 (a) Flexible, changing methods and approaches; adapting to new knowledge as the evaluation proceeds; using an iterative, open-ended model.

 (b) Recognition of multiple realities and the value of pluralism. Programs are seen by others in many different ways and the evaluator is responsible for portraying those many different pictures.

 (c) Local knowledge, local theories, and the particulars of an individual program, its nuances and sensitivities, are more important to convey than testing any big theory or generalizing to other settings.

1“Preordinate” evaluation refers to evaluation studies that rely on prespecification, when inquiry tends to follow a prescribed plan and does not go beyond or vary from the predetermined issues and predefined problems.

193 194

 (d) Case study and qualitative methods are important, essential methods to understand the particulars of a case and to correspond to the natural ways in which people come to understand something.

 (e) Evaluations should strive to be holistic, to convey the full complexity of a program, not to reduce or simplify.

 (f) Evaluation reports should follow this natural approach, presenting a rich set of information in full narratives emphasizing description and understanding.

 (g) The evaluator may make a judgment, but his or her individual judgment may differ from those of others presented with the information; thus, the evaluator’s role is also that of a learner and a teacher—a facilitator—to help others reach their own judgments.

The responsiveness and flexibility of the model are reflected in the clock Stake (1975b) developed to reflect the prominent, recurring events in a responsive evaluation. (See Figure 8.2.) Although the evaluator would typically begin the evaluation at twelve o’clock and proceed clockwise, Stake emphasized that any event can follow any other event, and at any point the evaluator may need to move counterclockwise or cross-clockwise, if events and increased understanding 194195warrant such changes. Further, many events may occur simultaneously, and others will occur several times during an evaluation.

One revealing comparison of responsive and preordinate evaluation approaches was provided by Stake’s (1975b) analysis of what percentage of their time evaluators of each persuasion would spend on several different evaluation tasks (p. 20):