An Introduction to Context and Its Role in Evaluation Practice

Jody L. Fitzpatrick

Abstract

Evaluators have written about the need to consider context in conducting evaluations, but most such admonitions are broad. Context is not developed fully. This chapter reviews the evaluation literature on context and discusses the two areas in which context has been more carefully considered by evaluators: the culture of program participants when their culture is different from the predominant one and the cultural norms of program participants in countries outside the West. We have learned much about how the culture of participants or communities can affect evaluation and should continue our learning there. Evaluators also need to expand their consideration of context to consider the program itself and its setting and the political norms of audiences, decision makers, and other stakeholders of the program. © Wiley Periodicals, Inc., and the American Evaluation Association.

Context. What do we mean by context? Consider the context in which you are reading this chapter. First, of course, there is the setting—the room and building in which you are reading. But the context for reading this chapter also includes your own background, knowledge, and beliefs about evaluation and its role in society; about governments; about clients, communities, and programs. Context is an amorphous issue. Many evaluators have written of the importance of context to evaluation practice. This
chapter summarizes some of the relevant literature to acquaint the reader with different approaches to context in evaluation and the factors to consider in conceptualizing context.

The importance of context to evaluation has been cited by authors since the beginnings of program evaluation in the United States (Stake, 1974; Stufflebeam, 1971; Weiss, 1972). A search of the American Journal of Evaluation reveals 902 citations for the word context. The Program Evaluation Standards (Joint Committee on Standards for Educational Evaluation, 2011) and Guiding Principles for Evaluators [American Evaluation Association (AEA), 2004] both refer to the importance of considering context in conducting evaluations. The Sage Handbook of Evaluation (Shaw, Greene, & Mark, 2005) includes three chapters concerning different contextual issues in evaluation. Finally, the “Evaluation Roadmap for a More Effective Government,” developed by the AEA Evaluation Policy Task Force (2009) to guide evaluation policy and management at the federal level, emphasizes context. They note that one evaluation strategy or design does not meet the needs of all programs nor all stakeholders or decision makers. Evaluations must be adapted to the history and stage of the program and be conducted in a manner that “is appropriate for program stewardship and useful for decision making” (p. 3).

Yet the various contextual factors that influence evaluation are rarely considered in much depth in the evaluation literature. Nods and admonitions are given to the importance of considering context and to its impact on evaluation plans, methods, implementation, and use, but few develop the construct in depth. Many evaluators focus on particular aspects of context that are particularly relevant to their own experience and approach to evaluation (Greene, 2005). But Greene notes there is a need for “a more sophisticated conceptualization and study of just how context matters in evaluation” (p. 84). We hope to begin a dialogue that will bring about more sophisticated conceptualizations of context to spark evaluators to consider and study how context matters. The different facets of context brought together here teach us something about what we mean by context and the factors we should consider in planning and conducting an evaluation, and illustrate where we still need to go in defining and elaborating on context.

Let us begin by considering some definitions of context. The Oxford English Language Dictionary (Oxford University Press, 2010) defines context as “the circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood or assessed.” The first part of the definition uses the broad term circumstances to refer to the myriad factors that comprise context; the latter part of the definition, however, is particularly noteworthy for evaluation. This part of the definition stresses its purpose; that is, knowledge of context is necessary in order to fully understand or assess something. The dictionary further defines the common phrase out of context as meaning “without the surrounding words or circumstances and so not fully understandable” (Oxford University Press,
AN INTRODUCTION TO CONTEXT AND ITS ROLE

2010). Thus, in settings broader than evaluation knowledge of context refers to the purpose of understanding something fully.

More specific to the field of evaluation, Greene (2005) defines context as “the setting within which the evaluand (the program, policy, or product being evaluated) and thus the evaluation are situated. Context is the site, location, environment, or milieu for a given evaluand” (p. 83). She identifies five specific dimensions to context in evaluation: demographic characteristics of the setting and the people in it, material and economic features, institutional and organizational climate, interpersonal dimensions or typical means of interaction and norms for relationships in the setting, and political dynamics of the setting, including issues and interests. Her identification of five dimensions of the contexts of evaluations is an important contribution to our beginning to think about context in a comprehensive, systematic fashion.

Others write about context with a focus on cultural contexts and the practice of evaluation that is responsive to different cultures, often ethnic or racial communities with less political power whose values have been ignored (Chouinard & Cousins, 2009; Frierson, Hood, & Hughes, 2002; SenGupta, Hopson, & Thompson-Robinson, 2004). In these writings, although the term culture is predominant, the terms context and culture are often used interchangeably. See, for example, Prado’s (2011) article entitled “‘Honor the Context’: Opening Lines for a Critical Multicultural Evaluative Practice.” Thomas (2004), however, differentiates culture and context. In describing a contextually-responsive evaluation framework and its use in urban schools in the United States, she sees context as broader than culture and defines context as

the combination of factors (including culture) accompanying the implementation and evaluation of a project that might influence its results, including geographical location, timing, political and social climate, economic conditions, and other things going on at the same time as the project. It includes the totality of the environment in which the project takes place. (p. 11)

Chouinard and Cousins (2009) also define context as subsuming culture, noting that context is “the site of confluence where program, culture, and community connect” (p. 461).

Historical Roots of Context in Evaluation

Evaluators writing in the early years of the field in the United States, the 1970s and into the 1980s, began to indirectly, and sometimes directly, define context for us. With a focus on use, though at different levels, they tended to focus on the context of those making the decisions. Stufflebeam’s new model moved evaluators from automatically concluding that their purpose was to determine whether objectives had been achieved to thinking about...
what decisions the managers were facing and how evaluation information might help them in making those decisions (Stufflebeam, 1968). (Unfortunately, today we seem to be moving back to automatically evaluating the achievement of objectives, today called outcomes, with a focus on accountability rather than use.) His development of the context–input–process–product (CIPP) evaluation model, is one of the early uses of the word context in evaluation (Stufflebeam, 1968, 1971). But Stufflebeam’s stages of evaluation were concerned with the stage of the program and the type of evaluation that might be appropriate. A context evaluation was conducted at the early stages of a program and its primary focus was needs assessment. Thus, the CIPP evaluation Web site indicates that context evaluation “assesses needs, assets, and problems within a defined environment” and is intended to answer the question “What needs to be done?” (http://www.wmich.edu/evalctr/archive_checklists/cippchecklist, 2007, p. 4). One element of the CIPP Evaluation Model Checklist (Stufflebeam, 2007) encourages the evaluator to “monitor and record data on the program’s environment, including related programs, area resources, area needs and problems, and political dynamics,” but the other elements of a context evaluation are concerned with assessing and determining the needs of program recipients to provide users with information on how to plan the intended program. Nevertheless, Stufflebeam brought evaluators’ attention in these early years to management and decision makers. As such, he awakened evaluators’ concern about Rog’s decision-making context (Rog, 2012). The evaluator was to learn about managers, their decisions, and the organizational context in which they were made in order to provide useful information.

Stake’s responsive model (1974) also dealt with context, but his focus was more on the program deliverers, in his case, teachers. The responsive model was developed in reaction to the dominant preordinate approach of the time, large studies designed to test big theories with fixed, quantitative methods. Greene and Abma (2001, p. 1) note that Stake reframed evaluation “from the application of sophisticated analytic techniques that address distant policymakers’ questions on program benefits and effectiveness ‘on the average,’ to an engagement with on-site practitioners about the quality and meanings of their practice.” Thus, Stake was calling upon evaluators to be responsive and, among other things, to focus on the local—local knowledge, local theories, and the events of an individual program. Like Stufflebeam, he wanted to shift evaluators from a focus entirely on objectives, but his solution was to become closer to the program, observing it, talking with program deliverers and drawing out their thoughts and perceptions, and developing case studies or in-depth descriptions of the program. Stake was moving evaluators toward seeing the program and, in so doing, to understand the program’s context. Thus, Stake (1980, p. 76) writes, “An evaluation will probably not be useful if the evaluator does not know the interests or the language of his audiences.” He recognized that stakeholders held different values and that it
was important for the evaluator to bring out their different value perspectives. By learning the interests, language, and values of audiences in a program, generally teachers and administrators, evaluators following Stake’s model consequently would become aware of different issues in the program’s context, not just information needs or decisions. But, primarily, Stake wanted to give voice to local users, to build on their expert knowledge, and to increase their control over their program (Stake, 1980) so that it could be responsive to local needs. He fought the uniformity he felt the federal government and outside experts were imposing on schools. Stake believed that local settings, local characteristics, and context make a difference in education.

Stufflebeam and Stake both worked in educational settings and, as such, these contexts affected their approaches to evaluation. As Stake’s beliefs demonstrate, schools in the United States were, and in most cases still are, primarily local institutions intended to respond to local needs. Although federal funds had prompted the growth of educational evaluation, the focus was still local. Meanwhile, other early evaluators (Weiss, Datta, Wholey) were working for the federal government conducting evaluations on a variety of large, often costly, social programs. The evaluations they conducted were often the ones referred to by Greene and Abma when they cite evaluations applying “sophisticated analytic techniques that address distant policy-makers’ questions on program benefits ‘on the average.’” Their potential audiences included elected officials and political appointees who had the power to make decisions regarding continuation, expansion, or elimination of national programs or their pilots. Influenced by this context, Carol Weiss became disappointed with policy makers’ failure to make much use of program evaluation (Weiss, 1972, 1973.) Her chapter on “The Turbulent Setting of the Action Program” in her early and influential evaluation text (Weiss, 1972) described sources of friction between evaluators and program planners, including differences in goals, roles, values, interests, and frames of reference. Weiss provides more detail than previous writers on how these two groups differ. Her goal was to make evaluators aware of the strikingly different contexts of program planners and, in turn, help evaluators increase the use of their studies. She also addressed the importance of other contexts including the social context of the program, its setting, and the “social frameworks of neighborhood and community” and “national systems of values, laws, and sensitivities,” noting that “local mores even determine what can be studied and what cannot” (Weiss, 1972, p. 108). And, foreshadowing today’s discussions on culturally responsive evaluation, Weiss discussed the need for evaluators to explore the social context of program participants writing “He [the program participant] does not come to the program empty, unattached, or unanchored. He has beliefs and values, he has friends and relatives, habits, patterns of behavior, and ideas. . . . One implication for evaluation may be the value of exploring the supportive and inhibiting features of
the interpersonal context. It might investigate the attitudes and behaviors of key people in the participant’s environment (family, coworkers, teachers)” (Weiss, 1972, p. 108). She encourages evaluators to become aware of these many contextual factors, closing with “The lesson for the evaluator is: Be alert. The studies that are ultimately most practical and useful are often those that open our eyes to new elements on the scene” (Weiss, 1972, p. 109). Thus, Weiss made early evaluators aware of the importance of studying additional contextual elements in an effort to increase the use and impact of evaluations.

Wholey, also working with evaluation in the U.S. federal government, shared Weiss’s concerns about use (Wholey, 1979, 1987). To learn more about the problem, Wholey and his colleagues studied federal evaluations and identified three apparent causes for the failure of evaluations to be used and to improve programs: lack of definition, lack of clear logic, and lack of management (Horst, Nay, Scanlon, & Wholey, 1974). Wholey developed evaluability assessment to counter these problems. Today, we recall evaluability assessment primarily for its contribution of the use of logic models to help in defining the problem, establishing goals, and linking program actions to those goals. But evaluability assessments also required the evaluator to attend to some new critical contextual issues for the program and the evaluation: the managers and their ability and motivation to use evaluations to improve programs. Wholey’s attention to managers and the decision context in which they operated was a new step in examining context.

Michael Patton’s original utilization-focused model (Patton, 1976) followed Wholey’s focus on managers and encouraged use by identifying one key manager who has the position to do something with the evaluation and the disposition, or interest, to use it. Patton’s model moved evaluators to begin thinking more about the organizational contexts in which their evaluations were used and to explore those contexts in thinking about utility. He made different and more specific suggestions for evaluators to use in working with managers.

All these early evaluators, Stufflebeam, Stake, Weiss, Wholey, and Patton, were encouraging evaluators to become more familiar with the context of the program and the managers and deliverers who were responsible for it. Weiss added more in noting the importance of contextual factors for participants and, more broadly, contextual elements of the community, state, or country. However, the arenas in which these evaluators worked differed, and these different arenas explained some of their different emphases concerning context. Wholey and Weiss evaluated federal programs, and their audiences were federal policy makers. As such, they were more distant from the programs they evaluated and considered political contexts in more depth. Stufflebeam, Stake, and Patton were working more closely with individual programs or schools, and thus could attend to the specific contextual circumstances of a school or a program setting.
More recently, Pawson and Tilley (1997) have proposed a model of realistic evaluation that makes explicit use of context in a quite different way. Rather than considering contextual elements to increase use or to give voice to local issues, their purpose is to use context to explain different program effects. They note that few programs fail completely or work perfectly. Instead, most programs succeed in some contexts with some clients. The evaluator’s job, then, is to identify those contextual elements that prompt a program to succeed or fail. Pawson and Tilley propose developing context–mechanism–outcome configurations to describe and explain program success.

The above are models of evaluation that their authors hope will influence evaluators in their practice. Some models require more allegiance, and less deviation, than others. More recent research on evaluators’ actual practice shows that evaluators do not follow one model (Christie, 2003), but adapt, though the results on why they are adapting are not quite clear. Fitzpatrick’s interviews of well-known evaluators shows that they do adapt to a variety of contextual elements, including the culture of the participants, what is known in the field of the program, views and interests of citizens and other stakeholders, the stage of the program, and the culture of the organization (Fitzpatrick, Christie, & Mark, 2009). These interviews highlight some of the different contextual elements that cause experienced evaluators to adapt and change their practices; yet they, too, fail to lead us to a comprehensive view of the potential contextual elements that might affect an evaluation.

Today, many evaluation models and approaches call for working closely with managers, program deliverers, and clients to learn more about their preferences and interests in both the program and in the evaluation, their values, and for clients and community, their culture. As such, most evaluators have moved from an early experimental, “hands off” tradition where they were concerned that their involvement might change the program or threaten the perceived neutrality of the evaluation to one in which evaluators are immersed in the program. These activities give evaluators the potential to consider and learn about context, but being immersed does not mean one notices the water. To illustrate the merits of considering context and its effect on evaluation more carefully, we review two particular areas of evaluation that have taught us more about context and its importance: cross-cultural or multicultural evaluation and international evaluations.

**Cross-Cultural Evaluation: Culture as Context**

As the United States has grown increasingly diverse in race and ethnicity, and as educational and social programs often serve people from minority or disadvantaged groups, evaluators have begun to realize that we need to know more about the context and culture of the clients served by the programs we
are evaluating. In the 1940s and 1950s, two African American educational evaluators, Aaron Brown and Leander Boykin, made others aware of the special needs and history of African Americans and argued that evaluators should seek their perspectives in conducting evaluations (Hood, 2001). Almost 40 years later, in her presidential address to members of the American Evaluation Association, Kirkhart (1995, p. 1) urged evaluators to turn their attention to culture, stating “multicultural influences shape and are shaped by our work.”

Since then, many have written about multicultural evaluation (Conner, 2005), cross-cultural evaluation (Chouinard & Cousins, 2009), and cultural competence (Thompson-Robinson, Hopson, & SenGupta, 2004). Frierson, Hood, and Hughes (2002) developed a model for culturally responsive evaluation (CRE) and have since expanded on the model (Hood, Hopson, & Frierson, 2005; Hopson, 2009). By 2009, the practice of cross-cultural evaluation had spread such that Chouinard and Cousins (2009) identified 52 evaluations making use of cross-cultural principles and used these articles to examine trends and propose a theoretical framework for future research. These writings help evaluators recognize their responsibilities to attend to cultures different from their own or different from the dominant culture, seek knowledge and understanding of different cultures, and involve stakeholders from participating cultures in the planning, conduct, interpretation, and reporting of the evaluation. U.S. evaluators, who are primarily from white, middle-class, educated backgrounds, have recognized that their own personal contexts and values influence how they see, or fail to see, other cultures. As such, our evaluations are invalid. They become invalid in many ways: by identifying the wrong questions to frame the evaluation, by ignoring key stakeholders who are potentially strong users of the evaluation, by misinterpreting stakeholders’ priorities or even program goals, by collecting data with the use of words or nonverbal cues that have different meanings to the audience, by failing to describe the program accurately or to understand its outcomes because the evaluator is unable to notice nuances or subtleties of the culture, by reporting results in means only accessible by the dominant culture or those in positions of power, and so on.

Understanding different cultures is not an easy task. Stafford Hood, one of the developers of culturally-responsive evaluation, has written, “It is difficult, if not impossible for evaluators to see, hear, and understand cultural nuances that differ from their lived experience” (2005, p. 97). He and others have proposed practical, specific methods for beginning to understand those cultures. These methods include careful examination of one’s own values, assumptions, and cultural context; inclusion of community members and program participants in evaluation planning and other phases; careful observation and respectful interactions and reflection on what has been learned; and training the evaluation team to be culturally responsive (Frierson et al., 2002; Madison, 2007; SenGupta et al., 2004). Most recently,
McBride (2011) has used sociocultural theory to suggest new strategies for conducting culturally responsive evaluations, including collecting data on local history from existing documents and through interviews with local leaders.

Most pertinent to this chapter is what this literature has taught us about the dimension of context that concerns culture. The writings on culture in evaluation typically focus on racial or ethnic cultures in the United States, although evaluators in New Zealand have also written about their work with indigenous New Zealanders and the need to gain cultural competence (Cram, 2012; Cram, Smith, & Johnstone, 2003). Generally, the cultures discussed are those of groups of individuals, for example, African Americans, American Indians, or Māori in New Zealand, and are considered because they are typically program recipients. These authors have alerted us to elements of context that we have not previously considered carefully: the norms and history of racial or ethnic groups in a particular community, school, or geographic area; their verbal and nonverbal means of communication; their norms concerning interpersonal interactions and dealing with people from other classes and cultures (like the evaluator); and, finally, their values and beliefs regarding the program, the needs it is addressing, its outcomes and activities, the evaluation, and the evaluation team and individuals on it. We must consider these cultural differences in selecting measures, in involving stakeholders, in reporting results, and in considering scalability and transferability of a program or policy to another setting (Cram, 2012).

From this literature, evaluators learn that we cannot list all the elements of culture any more than we can totally know a culture. But, as with our initial discussions of context, we can learn to learn more, to observe, to listen, and to be aware of these dimensions in order to conduct better evaluations.

One observation from this literature on culture and its role in evaluation: The authors tend to be from the educational evaluation arena. Their organizational context is schools. The cultural diversity in U.S. schools today and the increasing pressure on schools to close the achievement gap and to help all students achieve high educational standards may have made these evaluators more sensitive than those in other fields to the importance of learning more about the racial and ethnic culture of students. But other elements of context have also influenced educational evaluators to attend to culture. Schools are local, identifiable, and valued by the community, generally well known by many who live there and considered important to the community’s quality of life. Schools may bring together groups who are often apart and who have different cultures and values, different student and parent groups, teachers, school administrators, and coaches. Their work in schools has raised the awareness of evaluators in other arenas—social services, community development, employment, the environment—to the need to consider the culture of those served by the program. Finally, culturally responsive evaluations have been primarily a U.S. phenomenon, perhaps
because the United States is more diverse than many countries. The large majority (79%) of the cross-cultural studies found by Chouinard and Cousins (2009) were conducted in the United States and the most common organizational contexts were education and health.

Cross-cultural evaluation models and practice have increased our awareness of contextual issues pertinent to the culture of clients and community stakeholders, especially those from minority racial and ethnic groups, but much of the writing is primarily concerned with the context of the United States and schools. Will these issues and needs continue to emerge in other policy settings, in other countries, and with other racial and ethnic cultures? Certainly so, but we need to develop more knowledge of cultural norms that may affect programs and evaluations, power and dialogue, and ultimately understanding. We also need to be aware that cultural competence is needed in every evaluation, not just those where the cultural norms “hit us in the face” as being different from our own. Every group of participants has, or develops, a culture and that culture influences the program and the evaluation. (Consider, for example, a cohort of students completing a program together. Each develops its own somewhat different culture that is important for the program administrators, instructors, and evaluators to consider.)

Contextual Factors Arising From Literature on International Evaluations

Examinations of the practice of evaluation in different countries have made evaluators aware of other important contextual factors, in particular the influence of governments, public administration, and citizens’ and policy makers’ expectations for evaluation. Recognizing how international evaluations have informed us, Michael Patton, in an interview with King and Greenseid (Patton, King, & Greenseid, 2007, p. 12) observed “international diversity [in evaluation practice] is challenging our thinking about what constitutes good evaluation work and what it means for evaluation to be used in different cultural and political contexts.” Nevo first called evaluators’ attention to contextual differences across countries in his 1982 article on international evaluation. Nevo, an educational evaluator in Israel, was concerned that American models of evaluation were not likely to be valid in other countries. He noted that the United States is a democratic country where citizens’ “right to know” prompted the start of evaluation and influenced evaluation practices and models. Today in the United States, there is widespread distrust of government. These elements of the political context influence evaluation choices and approaches. And, these contextual elements differ across countries.

Almost 30 years after the publication of Nevo’s article, evaluation has grown rapidly in other countries, but the nature of evaluation in those
countries differs dramatically (Fitzpatrick, Sanders, & Worthen, 2011; Rist, 1990; Stern, 2005). A study by Rist and his colleagues (Rist, 1990) of eight Western countries found historic governmental policies influenced the nature of present-day evaluation in those countries. First-wave countries, so called because evaluation started earlier in these countries, began evaluation with an emphasis on social experimentation and program improvement, for example, the United States and the Great Society in the 1960s. In these countries, evaluation was often decentralized, with evaluators working closely with the programs they were evaluating, reporting to them on how to improve. Second-wave countries, including the United Kingdom countries, the Netherlands, Denmark, and France, began evaluation during the 1980s, during the Margaret Thatcher era. There, evaluation was intended to help control national budgets and reduce government spending. Evaluation in these countries was typically centralized, reporting to high-level government agencies and officials concerned with summative decisions about programs and budgets. Although concern with accountability has become a norm today in many first-wave countries, early traditions remain. In the United States, evaluations are often conducted by independent consultants. Internal evaluations are sometimes viewed with suspicion (Love, 1991). And most evaluations concern educational or social programs. As such, American evaluators are more likely to have been educated in psychology or education with a focus on individuals as the units of analysis than European evaluators (http://www.eval.org/Scan/aea08.scan.report.pdf). (The AEA environmental scan referenced here is of AEA members. Many evaluators who conduct evaluation work in the United States are not members of AEA and have backgrounds in other areas.) In Europe, evaluators are more likely to be schooled in economics and political science than American evaluators. They study programs ranging from agriculture to transportation; having arisen from a concern over government spending, evaluations focus on many different issues (Fitzpatrick, Sanders, & Worthen, 2011). By examining differences in evaluation across countries, we become aware of the influence that government policies and directions have on evaluation. These policies and views about evaluation and its role are critical elements of the context of evaluation.

Stern (2005) identifies many “contextual challenges” for evaluation today, drawing on his experience as a practitioner in Europe and the United Kingdom and as president of their respective evaluation societies. Like many evaluators outside the United States, he sees political issues at the national and international level as central to the context of evaluation. Some of these contextual challenges, he writes, are “driven by global shifts in the nature of the State and public policy that are reshaping evaluation practice” (Stern, 2005, p. 293). New public management, decentralization and privatization of some government services, and greater attention to transparency and accountability have influenced evaluation in many Western countries. Stern observes that
evaluators in Europe are more concerned with policy, and decision makers expect evaluators to have knowledge of policy content. In fact, Fitzpatrick observed that evaluators in France and Spain generally use the term policy evaluation rather than program evaluation and there was an intended, meaningful difference. Policies are generally broader than programs, and thus the object of the evaluation differs. Stern (2005, p. 299) concurs, noting that European policy makers believe that “piecemeal evaluation of individual programmes are not simply additive; many good programmes do not make a good policy.” Again, we see government having an important influence on the nature of our work, influencing the nature and scope of what is evaluated.

But geography and culture influence evaluation differently in other countries as well. Dahler-Larsen and Schwandt, in this issue, describe how elements of the Danish political culture affect what is evaluated and the questions that are asked (Dahler-Larsen & Schwandt, 2012). For example, the European Union and OECD publish many comparisons of European countries. The comparisons have a growing influence on Danish citizens' and policymakers' views of Denmark and policy priorities. In contrast, the United States is not part of a multi-country organization like the EU. Further, its size and geographical isolation from other countries leads to a reliance on comparisons among the 50 states, rather than across countries. Americans and Danes also differ in their views of the roles and responsibilities of government. Many Americans want to minimize the role of government and distrust its actions and intents. These views lead evaluation to focus on accountability, outcomes, and the achievement of standards (Fitzpatrick et al., 2011). In contrast, Danish citizens are more trusting of their government and see government as playing an important role in income redistribution and maintaining the welfare of citizens (Dahler-Larsen & Schwandt, 2012).

But, our review so far has concerned only political contexts in Western countries. The context for evaluation in developing countries is quite different and varied. In many of these countries, evaluation was introduced by multinational development agencies like the World Bank and UN organizations. Nongovernmental organizations (NGOs) like World Vision, OXFAM, and CARE also play major roles in funding and conducting evaluations in such countries (Bamberger, 1999). Capacity building is a major focus in some countries (http://afrea.org/home/index.cfm). But the political and cultural context for evaluation in these countries is quite different. Bhola (2003) notes that evaluation is sometimes viewed as a Western notion imposed by former colonial powers.

In many of those countries, citizens’ relationships and their concerns with government differ sharply from those in the West. Bamberger (1999) observes that evaluators who seek participation from local stakeholders or collect data in ways that conflict with community norms may unknowingly create problems, loss of services, rejection, or violence for those who
participate. Both cultural norms and nondemocratic governments can provoke these challenges and others. Men may object to their wives or children participating in data collection, such as interviews outside of their presence. Village leaders may perceive the evaluation, participatory efforts, or data collection as threatening their authority. Government officials may limit services of those believed to be critical of government actions and restrict the choices and actions of the evaluator or dissemination of information.

Perhaps because the context of these evaluations is so different, evaluators are more likely to recognize the importance of learning about local contexts and cultural norms than when they are working in their own country. Stevenson and Thomas (2005, p. 215) state that “external evaluators [in developing countries] cannot design effective evaluations without incorporating ‘local knowledge.’” This need to incorporate local knowledge and to be observant of different political and social customs and expectations is a central focus of such evaluations. In contrast, culturally responsive evaluation has had to awaken American evaluators to the need to become culturally competent. In the context of public health evaluations in developing countries, Stevenson and Thomas see the environment contributing to “the development of a pragmatic, culture-sensitized approach to evaluation traditions” (2005, p. 218). Paradigm wars as seen in the United States in the 1990s have not occurred in developing countries; instead, these evaluators are comfortable with a wide range of qualitative and quantitative methods. Evaluations measuring cost effectiveness is more often a major focus of evaluation in these countries with scarce resources. In Africa, evaluations of development projects and consideration of human rights and gender issues are major topics for the African Evaluation Association (http://afrea.org). Again, we see the context of the country—the nature of government and its policies, expectations of NGOs and multinational development agencies, cultural norms of individuals and communities—all issues beyond the specific program to be evaluated, affecting the nature of the evaluation, the questions to be addressed, the methods of data collection, and stakeholder participation.

International evaluations also make us aware of widely different cultural norms—norms concerning communication, consensus, and conflict, openness to new information, styles of decision making—and their effects on evaluation practice. Nevo (1982) identifies these and other related issues as a major factor in considering the validity of American evaluation models in other countries:

Without discussing the validity of such concepts as national character or national personality, one might suspect some differences in personal characteristics among decision-makers and evaluators in various societies. Rationality, punctuality, intellectual curiosity, need for structure or tolerance for ambiguity might be of special relevance to the conduct of evaluation. (p. 74)
Examples of such differences and their effects on evaluation have emerged since in discussions of evaluation in different countries. Patton (2007) cites cultural differences between Japan and the United States that have strong implications for the communication of evaluation information. He comments that he has found “perspectives for handling feedback and engaging in learning to be fundamentally different in Asian contexts” (p.112), observing that Americans tend to have a “very competitive and blaming culture.” As such, pointing out problems with or failures in individual programs is more or less expected in American evaluations. In contrast, in Japan social harmony is a much stronger cultural value and “for the sake of group harmony, they’re not into blaming and embarrassing people and pointing out their faults” (Patton et al., 2007, p. 112). Thus, communicating results of an evaluation in these different cultural contexts must be handled in a quite different way. Fitzpatrick’s American norms for communication created difficulties in Spain. Communication that would be seen as forthright and constructive, albeit negative, in the United States was avoided in Spain and, as such, obtaining input or reactions from stakeholders and colleagues was more difficult and took much more time.

Every country has different cultural and political norms that affect many aspects of evaluation from selection of evaluation questions, stakeholders and potential users, methods of data collection, and means for reporting. Evaluators who work in countries different from their own recognize the need to include evaluators native to that country on their evaluation team and to be sensitive to the quite different context in which their evaluation is conducted. Wallis, Dukay, and Fitzpatrick (2008) describe how they worked with a multidisciplinary evaluation team composed of evaluators from the United States and from Tanzania to evaluate an orphanage for children whose parents had died of AIDS. Within Tanzania, as in any country, the evaluators from the university in Dar es Salaam differed in many ways from the people who lived in the rural town of the orphanage. Thus, they also sought extensive participation in planning, data collection, and interpretation from Tanzanians from that village and adjacent villages.

Finally, international work helps us to recognize some of the cultural and political contexts in our own countries and variations within our country, for example, across states and cities. Knowledge of international evaluations and the contextual elements that affect them can improve our own evaluations even if we never practice evaluation in another country. We become more sensitive to the larger contextual issues affecting our evaluations, for example, roles of government, evaluation policies and expectations for evaluation, styles of communicating, views concerning what constitutes credible evidence, and a variety of other contextual factors that affect our choices and the use of our findings. As Conner (2011) has observed, American evaluators could learn from treating our evaluations in the United States as if we were in a foreign country.
Closing

This chapter presents only a sampling of the evaluation literature that concerns context, but does delve into two areas, cross-cultural evaluation in the United States and international evaluation practice, which provide more depth on contextual factors than other areas. Other literature that alerts us to contextual elements we might otherwise overlook includes the literature on organizational learning (Preskill & Torres, 1998), the contexts of internal and external evaluators (Love, 1991; Sonnichsen, 1999), and differences across government, nonprofit, and private-sector programs (Davies, Newcomer, & Soydan, 2005). Our work here attempts to motivate readers to consider these and other elements of context, ones we might have overlooked. American evaluators (Frierson et al., 2002) recognize contextual differences across the cultures of racial and ethnic groups in the United States. The models of culturally responsive evaluation and standards on cultural competence come from their own contexts, evaluating youth at risk or underserved groups in a United States that has long been relatively heterogeneous. As such, they remind evaluators in other countries, countries with growing immigrant populations, to consider the cultural contexts of those served by the programs they are evaluating. Conversely, European evaluators are more likely to evaluate policies in many areas. Their knowledge of the contexts in which these programs and policies are developed, characteristics of government and policy making, and expectations of citizens regarding government can prompt American evaluators to consider these political contextual elements in their own work. We expand our awareness and improve our practice by moving beyond our own disciplinary and domestic boundaries to learn about how others consider and frame context.

References


Conner, R. (2011, November). The effect of political values and expectations on evaluation: Perspectives from different countries. Remarks as discussant for this session at the meeting of the American Evaluation Association, Anaheim, CA.


**JODY L. FITZPATRICK** is associate professor in the school of public affairs at the University of Colorado, Denver, and will be the president of the American Evaluation Association in 2013.