CHAPTER 3

THROUGH A CULTURAL LENS

Reflections on Validity and Theory in Evaluation

Karen E. Kirkhart
Syracuse University

Cultural diversity has become a critical issue for program evaluation in the United States due to population demographics in general and to the nature of the target populations of many educational and social programs in particular. Marked differences among the personal characteristics, backgrounds, and belief systems of the *consumers* of human service programs, the *providers* of human service programs, and the *evaluators* of human service programs make understanding the impact of culture a priority concern. Culture impacts all aspects of evaluation—from the formation of evaluation questions to the selection of data sources; from data gathering methods and data analysis techniques to strategies for communicating findings. As with all knowledge, evaluative understandings and judgments are culturally contextualized. To establish the validity of such understandings and judgments, cultural diversity must be explicitly addressed. Appreciations of diverse cultural perspectives strengthen validity; they must be expanded and deepened. Biases embedded in cultural diversity threaten validity; they must be exposed and interrupted (Fine & Powell, 1997).

In evaluation as in measurement, validity is *the* most important construct (Linn, 1997). It references the accuracy and limits of understandings; it guides what can and cannot appropriately be concluded from evaluative inquiry. Validity addresses the fundamental correctness of evaluation. A key dimension of validity involves appreciating the culturally bound nature of understandings and judgments. Valid evaluation presumes an understanding of culture and culturally based discrimination as well as the ability to identify appropriate and inappropriate considerations of cultural context in evaluation's epistemological, methodological, and theoretical foundations, professional practices, and standards and guiding principles.

VALIDITY THROUGH A CULTURAL LENS

Though culture belongs at the center of any conversation about validity, in practice it has often been excluded. Kirkhart (1995b) introduced the term *multicultural validity* to move considerations of culture to the center of validity arguments. Multicultural validity refers to the correctness or authenticity of understandings across multiple, intersecting cultural contexts (Kirkhart, 1995b). It focuses attention on how well evaluation captures meaning across dimensions of cultural diversity, and it scrutinizes the accuracy or trustworthiness of the ensuing judgments of merit and worth. Like validity in general, it is a multifaceted construct, permitting one to explore the many ways in which culture impacts meaning and understanding. Multicultural validity may be argued and understood in terms of methodology, consequences, interpersonal relationships, lived experience, and theory (Kirkhart, 2004). Figure 3.1 summarizes the five justifications. Each justificatory perspective directs attention to a different type of evidence to support or challenge validity. Methodological justifications of multicultural validity direct attention to the validity of measurement and design elements; consequential justifications examine the impacts or sequelae of evaluation to reflect on validity; interpersonal justifications scrutinize relationships among the researcher(s) and the researched; experiential justifications examine validity in terms of the lived experience of program participants; theoretical justifications of multicultural validity scrutinize theoretical foundations. This chapter explores the relationships among culture, validity, and theory and examines how culture may support or threaten theoretical justifications of multicultural validity of program evaluation.
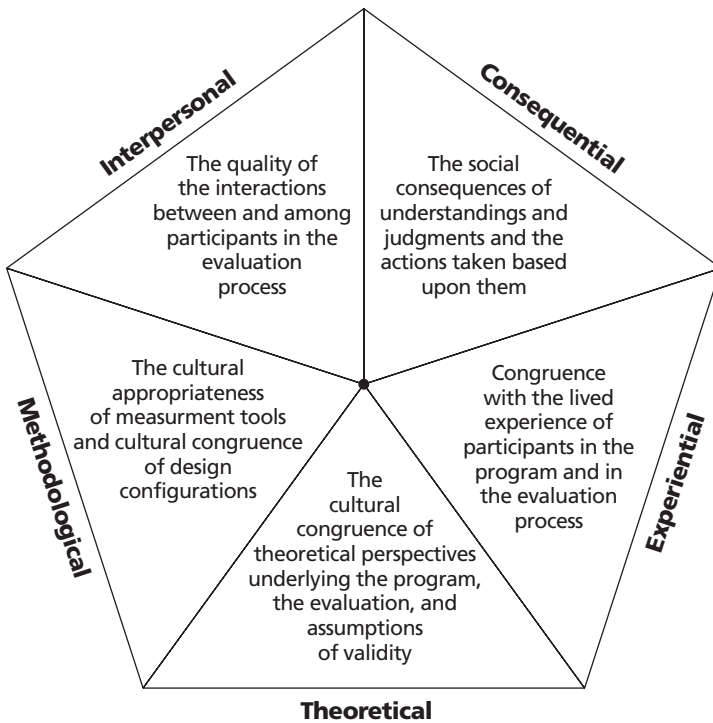
Figure 3.1.   Five justifications of multicultural validity.

THINKING MULTICULTURALLY ABOUT THEORY

Theory is a powerful source of influence on evaluation practice (Lipsey & Pollard, 1989; Weiss, 1995). Theory guides method. Before developing "culturally anchored methodology" (Hughes, Seidman, & Williams, 1993), one needs to pause to articulate and examine the foundational theory or theories in which methods are grounded. Similarly, to appreciate cultural influences on validity, one must scrutinize the theory that guides the practice. While absence of theory may weaken multicultural validity, the presence of theory is no assurance of multicultural validity, because the theory itself may be culturally biased. Bias in theory is less frequently examined than bias in method, yet both are tied explicitly to evaluation practice.

The importance of theory stems from its ability to guide evaluators, to steer program developers and providers, and to set the standards of scientific credibility. Accordingly, this chapter addresses three domains of theory: evaluation theory, program theory, and validity theory. The

interactions of culture and validity within these three domains are presented as three propositions, each of which is justified and discussed. The propositions are grounded in the central argument that understanding how theory engages culture is a prerequisite to strengthening theoretical justifications of multicultural validity. The chapter closes with nine recommendations for scrutinizing theory through a cultural lens.

CULTURE, VALIDITY, AND EVALUATION THEORY

Evaluation theory guides epistemological, methodological, and practical choices; therefore, it is critical to question how well evaluation theory addresses dimensions of cultural context and conversely to reflect on how culturally biased assumptions may enter evaluation theory. Good theory suggests epistemological options. Good theory suggests the circumstances in which certain methods should be chosen over others, ways in which methods should be sequenced or combined, and costs and benefits of different methods, including the types of questions that are best addressed via a particular strategy. Good theory suggests who participates in evaluation via what roles. Simply put, it lays the foundation for evaluation practice. In evaluation theory, it is important to scrutinize how culture is framed and incorporated, if, in fact, it is addressed at all. Major sources in the field note the connection of theory to the background and worldviews of the various authors and proponents; however, none explicitly addresses cultural context (e.g., Alkin, 2004; Fitzpatrick, Sanders & Worthen, 2004; Shadish, Cook & Leviton, 1991; Stufflebeam, 2001). This is an omission that needs to be addressed.

> **Proposition 1:**   *Validity is enhanced to the extent that evaluation theory guides practitioners in selecting epistemologies, methods, and procedures appropriate to a well-informed consideration of cultural diversity and thoughtful reflection on potential culturally bound biases. Validity is threatened to the extent that culture is ignored or diversity variables are included as simplistic, atheoretical stereotypes.*

In their classic text on evaluation theory, Shadish and colleagues (1991) define *theory* as "a body of knowledge that organizes, categorizes, describes, predicts, explains, and otherwise aids in understanding and controlling a topic" (p. 30). Evaluation theory provides a logical framework for the evaluation enterprise, defining what is included and excluded as necessary and legitimate components, controlling what is left in or out of a definition of evaluation. The key issue here is how the boundaries of the topic are defined and legitimated and how the nature of the enterprise is understood. There is no single, universal evaluation theory; rather, multiple the-

oretical perspectives are available for evaluators' reflection and guidance. Each casts its own worldview on program evaluation. Shadish and colleagues see evaluation in terms of causal processes, Scriven (1991) in terms of logic and values, Schwandt (1997) in terms of practical hermeneutics, House (1980, 1993) in terms of social justice, Fetterman (1994a, 1994b; Fetterman, Kaftarian, & Wandersman, 1996) as empowerment, Mertens (1998) as emancipation, Smith (1987) in terms of justification of claims, Greene (1995) in terms of social advocacy, Preskill and Torres (2000) as transformational learning, Hood (2001) and Thomas and Stevens (2004) in terms of culture and context, and Patton (1997) in terms of managerial support, to illustrate from among current theorists. How the nature of evaluation is understood—the language, metaphors, symbols, images, and arguments—sets the context for defining and operationalizing constructs such as race, culture, and power and gender (Hopson, 2000). Some theories "presuppose fairness" (Thomas, 1997) and do not elaborate on it, while others explicitly address issues of power, privilege, and social justice. But these constructs themselves are not ones about which there is universal agreement.

The influence of evaluation theory and culture is bidirectional; culture both impacts and is impacted by evaluation theory. First, culture shapes theory in evaluation through the historical and cultural contexts of theory development, including the theorists themselves. Evaluators' personal characteristics, orientations and identifications, life histories, academic training, and cultural experiences are inescapably woven into the theoretical understandings they put forth for consideration. Issues of culture and evaluator voice are visible at the level of individual theorists (Alkin, 2004), and in the ways in which entire societies have been theorized (Smith, 1999). For example, Scheurich and Young (1997) trace racial bias affecting research from the individual level through a hierarchy that progresses through institutional racism, societal racism, and civilizational racism. They argue that *epistemological racism*, a construct relevant to thoughtful reflection on evaluation theory, is grounded in the broadest, civilizational level and must be understood in those terms to be productively addressed.[1] No evaluation theory is context-free, but among current evaluation theory the dominance of white social history must be acknowledged and understood as a source of cultural bias (Hood, 2001).

Second, theory shapes cultural understanding. Evaluation theory offers a framework for addressing cultural context in the evaluation of social programs. With each theoretical perspective comes a potential window on cultural influence. For example, Shadish and colleagues' (1991) version of evaluation theory is primarily interested in describing causal processes that mediate a relationship. In their worldview,

the ideal (never achievable) evaluation theory would describe and justify why certain evaluation practices lead to particular kinds of results across situations that evaluators confront. It would (a) clarify the activities, processes, and goals of evaluation; (b) explicate relationships among evaluative activities and the processes and goals they facilitate; and (c) empirically test propositions to identify and address those that conflict with research or other critically appraised knowledge about evaluation. (pp. 30–31)

Consistent with their view of the fundamental purpose of program evaluation theory, they define the scope of evaluation theory in terms of five theoretical bases: social programming, knowledge construction, valuing, knowledge use, and evaluation practice. Each of these theory bases can then be scrutinized for how it may support valid treatment of cultural diversity in evaluation or how it may constrain understandings and contribute to bias and discrimination. For example, under the value component of evaluation theory, Shadish and colleagues address the need to "make value problems explicit, deal with them openly, and produce a sensitive analysis of the value implications of programs" (p. 47). They open a window to discussions of power and privilege, although the culturally bound nature of values is only touched upon, and no dimensions of diversity are illustrated or explored. In their discussion of values, they privilege descriptive theory (theory that describes values without advocating one as best) over prescriptive theory (theory that advocates the primacy of particular values), defending descriptive theories as "more consistent than prescriptive theories with the social and political organization of the United States, which is based upon fostering a pluralism of values that compete against each other in social and political arenas" (p. 47). One can then trace the implications of this position for the multicultural validity of their evaluation theory. If values are to be sorted out in the social and political arena, then those arenas themselves must be scrutinized with respect to cultural bias. The other components of their theoretical model may be similarly examined (Kirkhart, 1995a).

The central point of this argument is that within any theoretical framework, one must scrutinize implications for multicultural validity. If the theoretical basis of an evaluation were changed, then the parameters of the enterprise, the types of cultural bias perceived, the relationship of diversity to evaluative judgments of social programs, and the methods deemed appropriate would all shift. Evaluation theory provides powerful lenses, magnifying certain aspects of evaluation, minimizing or even filtering out others. Contrast the role of utilization of results, for example, in the theories of Patton (1997), who views it as integral to the enterprise, and Scriven (1991), who views it as optional to undesirable, depending on context.

In summary, evaluation theory is impacted by culturally bound bias ranging from individual to civilizational levels. Theory also defines frame-

works within which evaluation can address cultural diversity and reflect on issues of power and privilege. Evaluation theory is the first domain in which culture and validity interact. The second domain is program theory.

CULTURE, VALIDITY, AND PROGRAM THEORY

Program theory describes causal linkages presumed to exist between program activities and intended outcomes. It guides program design and implementation as well as program evaluation. Program theory can illuminate program operations and hidden assumptions, potentially including assumptions about culture. Two major sources of program theory are social science and practice wisdom, each of which is subject to cultural bias. Therefore, program theory raises two distinct validity issues.

> **Proposition 2A:**   *Validity is strengthened to the extent that program theory is guided by empirical social science research that itself addressed cultural diversity in a valid manner. Conversely, validity is undermined by the adoption of social science models as frameworks for practice or evaluation despite inappropriate or inadequate attention to culture.*

First, program theory may originate in social science (i.e., the logic underlying a social program or its intervention components may come from discipline-based, empirically tested social science theory). It is appropriate to strive to build upon prior knowledge; if relevant social science theory exists, it can form a solid groundwork for program intervention. However, three potential threats to multicultural validity emerge from this strategy. First, the original research may have been conducted from a majority perspective that at best severely limits and at worst severely distorts the conclusions. Here, the desire to ground program theory in "proven" social science knowledge bases opens up a wealth of challenges to the multicultural validity of those understandings. The uncritical use of social science theory proliferates majority viewpoints that may systematically exclude cultural standpoints. Evaluators must be trained to examine the assumptions underlying social science theory and the methods previously used to develop and test that theory before relying upon the social sciences as a firm foundation for program theory. The first potential threat to the validity of social-science-based program theory lies within the research itself.

The second potential validity threat lies in the translation of social science to social programs and the fact that social science theory does not lend itself to being used "off the shelf," as Chen and Rossi (1983, p. 285) acknowledge. This threat includes, but extends beyond, traditional con-

cerns with the generalizability of social science theory to persons different from the populations studied. It also addresses the adaptations that are inherent in the translation of theory to practice. The modifications may involve subtle shifts of emphasis or operationalization, or major recasting or selective use of theory components. The application contexts of concern may be the social problems themselves or the programs that are designed to address them. In the first instance, the translation of social science theory into program policy and procedures must be scrutinized with respect to assumptions underlying the definition of the social problem and plausible programmatic solutions. To what extent, for example, do middle-class Anglocentric understandings of a social problem lead to selective use of social science theory? How do medical models of aging relate to the life experiences of octogenarians? How does a plausible solution or intervention strategy make the leap from controlled environment to culturally contextualized service delivery setting, replete with resource constraints and implementation challenges? As social science theory is translated into program theory, what is lost or gained in the translation? This is the second point at which cultural diversity may influence social-science-based program theory.

The third point of influence relates specifically to the use of social science theory in the practice of *evaluation*. The best use of social science in developing program theory is identified early and used to guide the program itself. However, it is sometimes the case that evaluators peruse social science literature to design a theory-based evaluation of an essentially atheoretical program. In addition to the validity threats identified above, this runs the additional risk of creating a "straw" program theory for purposes of evaluation that bears no connection to either the intention or the reality of the program providers. Potential bias inherent in this translation may either go unexamined and unrecognized or be recognized and superficially acknowledged yet fail to influence evaluation practice. Either way, cultural bias has entered the evaluation process via the use of social-science-based program theory, undermining the multicultural validity of the evaluation.

An alternate source of program theory is program providers and other stakeholders. Local theory has the advantage of intimate knowledge of program realities and operations and can potentially offer greater attention to the program's cultural context. These advantages notwithstanding, cultural bias may enter program microtheory through either the content of local knowledge or through the process of theory construction itself.

Proposition 2B: *To the extent that program theory is guided by practice wisdom grounded in the cultural traditions and beliefs of program participants, validity is supported. However, when practice wisdom itself is prejudicial, validity is threatened.*

There are two distinct paths through which culture shapes local theory. First, culture enters program microtheory through the *content* of what key players bring to the table. Second, culture enters theory development through the *process* of development itself. In the first case, local theory reflects the knowledge and understanding of persons close to the programs—designers, administrators, providers—regarding how the program operates or is intended to operate. Many models of evaluation in fact privilege this source of program theory over others (Fetterman, 1994a; Guba & Lincoln, 1989; Patton, 1997). There is an implicit—and sometimes explicit—assumption that grounding theory in the understandings of local stakeholders is necessarily more genuine, more multiculturally valid and less biased. But there is a danger here that often goes unaddressed. What happens to program theory when "common knowledge" is based on ignorance, misunderstanding, or lack of historical context of cultural standpoints? For example, validity of program theory is threatened when "local wisdom" is racist or otherwise prejudicial (Stanfield, 1993). How this is addressed (or not addressed) leads to a second way in which culture shapes local theory, and that is through the development process itself.

Examining how culture affects the process of developing local theory shifts attention to the broader arenas of participation and group dynamics. At least two issues can be highlighted here. First, one must consider who is invited or included in the conversations surrounding program theory. This includes personal characteristics, but it also includes issues of stakeholder role and privilege. Most discussions of program theory development recognize the need to include a diverse cross-section of providers (e.g., program designers, administrators, middle managers, and direct service providers); however, consumers are rarely represented in the dialog. Madison's (1992) advocacy for primary inclusion of consumers in the evaluation process applies equally to the development of local program theory. Second, one must reflect carefully on how the interaction is orchestrated or managed. The extent to which the interactions are structured, and by whom, are key parameters of the formal process. Mathie and Greene (1997) mark an important distinction between creating opportunities for meaningful engagement and treating diversity as token representation. But issues of power and privilege also influence the interactions and communications informally in terms of who is heard. For example, Bell (1997) asserts that Blacks are often listened to but not believed unless speaking White scripts.

In applying program theory to the development of comprehensive community initiatives for children and families, Weiss (1995) discusses the process through which local theory may be developed.

> An important step will be to discuss the theories that practitioners and residents engaged in community-building activities actually have in mind as they

go about their practice. Often their theories will be implicit rather than explicit and it may take time for them to think through their assumptions about how their work will lead to the effects they seek. Nevertheless, *the feasibility of theory-based evaluation rests on their ability to articulate their assumptions (or assent to someone else's formulation),* and it is important to see how well this phase of the task can be done. (p. 82, emphasis added)

Issues of power and privilege imbedded in the development of local theory are clearly visible in this passage. There is a vast difference between articulating one's own assumptions and assenting to someone else's formulation, and the validity of the resulting theory hangs in the balance.

In summary, program theory provides a window for viewing the logic and assumptions that undergrid human service programs. What it reveals must be scrutinized to identify and challenge cultural bias. This demands an attitude of vigilance that is equally critical of social-science-based theory and local theory. Cultural bias may be present in either program theory content or the process of theory development. These are complex, difficult issues that merit the evaluation profession's most thoughtful attention; inattention leaves program theory vulnerable to bias and distortion, undermining multicultural validity.

CULTURE, VALIDITY, AND VALIDITY THEORY

Finally, validity theory itself must be examined reflexively. Validity theory provides rules for judging the accuracy of knowledge claims and data-based inferences as well as judging the worth of their applications. Validity theory sets the parameters of what is to be included and excluded under the rubric of validity and how the multiple dimensions or facets of validity fit together. The scope of what should be included under the umbrella of validity or validation has been much debated. Consistent with the argument that follows, this chapter adopts an inclusive definition of validity, as *an overall judgment of the adequacy and appropriateness of evaluation-based inferences and actions and their respective consequences.*

> **Proposition 3:** *Validity is strengthened to the extent that validity theory provides a broad framework for examining the accuracy of understandings across cultural differences, and promotes reflection on bias, misunderstandings and assumptions born of majority privilege. The validity of evaluation is threatened to the extent that validity theory is narrowly construed, excluding dialogues about social consequences.*

Proposition 3 is supported by two related arguments. The first argument refers to the inclusiveness of the definition of validity, while the second concerns the culturally bound nature of validity itself as a social construc-

tion. First, consider culture in relation to the scope of the definition of validity itself. Validity occupies a privileged status as "the most fundamental consideration in developing and evaluating tests" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, p. 9). To retain that position, it must continue to evolve to incorporate the considerations essential to determining the correctness of our understandings and actions. Nowhere is the need for an inclusive definition of validity more apparent than in the practice of program evaluation. Consider three illustrations of such inclusion: inclusion of alternate epistemologies, of actions and their consequences, and of social justice.

First, validity theory must be inclusive in its attention to alternate epistemologies. Epistemological racism and other bias imbedded within traditional methods of inquiry has led to calls for alternate epistemologies (Gordon, Miller, & Rollock, 1990; Scheurich & Young, 1997; Stanfield, 1993). These calls have been answered both within and outside of the evaluation community by the creation of alternative epistemologies, including but not limited to epistemology grounded in race and ethnic standpoints, feminist theory and queer theory (see, e.g., Alcoff & Potter, 1993; Fetterman, et al., 1996; Guba & Lincoln, 1989; Harding, 1998; Jagose, 1996; Kershaw, 1992; Ladner, 1973; LaFrance, 2003; Parker, 2004; Seigart & Brisolara, 2002; Smith, 1999; Sullivan, 2003). The validity of these new epistemologies must then be examined in its own right. To do so requires an expanded definition of validity, however, since the alternative epistemologies are based on arguments that fall outside of (and in some cases explicitly oppose) traditional definitions. In this sense, considerations of culture open up the construct of validity. Addressing culture within validity requires a construct capacious enough to honor and build upon traditional concepts plus embrace alternative paradigms, including those previously marginalized (Rosaldo, 1993).

Second, validity theory must be inclusive in terms of attention to actions and the consequences of those actions. Invalid, prejudicial inferences drawn from data are certainly a concern, but one must also consider the actions taken based upon inferences and the consequences of those actions, both of which may have profound implications for underrepresented populations. Cronbach (1971) explicitly included within the scope of validity decisions and actions based on test scores as well as descriptive interpretations, noting that a decision is a choice between courses of action. Similarly, Messick (1981) explicitly included actions as the object of validation, defining test validity as "an overall evaluative judgment of both the adequacy and the appropriateness of both inferences and actions derived from test scores," a definition reiterated in subsequent work (Messick, 1988, 1989a, 1989b, 1995). Applied to validity of program evaluation,

this means that in considering how cultural diversity interacts with validity, one must examine not only the interpretations made and the evaluative conclusions reached, but also the actions taken based on evaluation results.

The second half of the argument for inclusion addressed by point two refers to the consequences of the actions taken based upon evaluation. As Shepard (1997) points out, these two arguments are related, since once the definition of validity expands beyond descriptions or interpretations without actions to include the soundness of test-based decisions, one must necessarily think about the consequences of those decisions. Messick and his supporters argue that the consequences of drawing certain test-based inferences or taking certain test-based actions must be scrutinized as part of the validation process.[2] In parallel, judging the validity of evaluation requires attention to the social consequences of the judgments made and the evaluation-based actions taken. The focus on consequences is especially necessary in examining issues of fairness and equity across individual, institutional, and societal levels. Shifting focus to societal consequences leads to the third aspect of the argument for an inclusive definition of validity, social justice.

Inclusion of social justice comes from broad concerns about the ways in which power and privilege are distributed and controlled in our society and the role of evaluation within that system. House (1980) first cast social justice as a validity concern for evaluators. His broad definition of validity as "worthiness of being recognized" (p. 249) includes but extends beyond the truth claims of measurement validity and experimental method to include credibility (of the evaluator and of the evaluation) and normative correctness or justice.[3] The following quote captures the argument well:

> Evaluation is by nature a political activity. It serves decision-makers, results in reallocations of resources, and legitimizes who gets what. It is intimately implicated in the distribution of basic goods in society. It is more than a statement of ideas; it is a social mechanism for distribution, one which aspires to institutional status. Evaluation should not only be true; it should also be just. Current evaluation schemes, independently of their truth value, reflect justice in quite varying degrees. And justice provides an important standard by which evaluation should be judged. (House, 1980, p. 121)

In effect, validity becomes a point of intersection of justice with social program evaluation because of the focal issue of power. Support for the connection between validity and issues of power is contributed from perspectives as diverse as psychometrics (Cronbach, 1988) and dialectical materialism (Enerstvedt, 1989). To evaluate social programs with validity means expressly addressing issues of power and privilege—whose interests are served and not served—and how those interests are registered and understood. In matters of validity, as in matters of social justice, evaluators

are not value-neutral (Cronbach, 1980; House, 1993). An ethic of public responsibility underlies and connects these arguments.

In sum, three arguments have been advanced to support the need for an inclusive definition of validity as a construct. First, validity needs to be defined broadly enough to address alternate epistemologies, many of which have been developed to give greater voice to historically underrepresented groups. Second, validity must include more than evidence-based inference; it must address decisions made and actions taken based on those understandings as well as the consequences of those actions. Third, a broad definition is needed to address the social justice implications of understandings, actions, and consequences. Granted the necessary scope, validity has the power to interrupt dynamics of prejudice, while a narrow, tightly constrained definition of validity works to restrict dialog and maintain the hegemony of the status quo.

The second argument of Proposition 3 examines the cultural implications of the construct of validity itself. This reflection on the validity of the validity question itself has been well articulated by Kvale (1995, p. 36) who asks, "Is the question of validity in social science a valid and legitimate question?" What are the cultural boundaries of validity as a social construction? Neither the term itself nor the concept of seeking validation carries the same meaning across cultural contexts. It follows that scrutinizing validity may not always be a productive avenue for improving practice. What if attention to validity in fact hinders or impairs practice or distracts us from more important ways of considering culture? For example, concern for control and legitimation might potentially overshadow and suppress creativity and production of new insights (Kvale, 1995).

Whether validity and creativity work to complement or to oppose one another is a matter of empirical examination, but Kvale's point concerning issues of control surrounding definitions of validity in the scientific community is well worth noting. As argued above, validity acts in important ways as gatekeeper to what is seen as legitimate contributions to the knowledge base of practice. To the extent that this gatekeeping function suppresses nonmajority viewpoints, supports the status quo, and restricts conversations that challenge power distributions, validity may itself become oppressive or discriminatory in its consequences. This interacts with the scope of the definition of validity accepted as legitimate, the first argument advanced under Proposition 3.

This section has advanced two related arguments to support Proposition 3. The first is the need for an inclusive definition of validity to create a construct of sufficient breadth to scrutinize the many ramifications of culture and cultural bias. The second is a need to be critically reflective about the validity of validity itself and the ways in which it dispenses legitimacy and

influence within the academy in general and the evaluation profession in particular.

CONCLUDING THOUGHTS AND DIRECTIONS

This chapter has argued that culture is relevant to all aspects of evaluation and that it must be appropriately addressed to establish the fundamental validity of evaluative inferences and actions. Multicultural validity marks the centrality of culture in reflections on validity. Theory can support or undermine multicultural validity of evaluation, an argument advanced by considering three dimensions of theory: evaluation theory, program theory, and validity theory itself.

Multicultural validity of evaluation can be strengthened by theory that approaches culture in knowledgeable, thoughtful, and respectful ways; yet no theory ensures perfect understanding across cultural differences and none can be above scrutiny. Evaluators must continue to remain vigilant. To support necessary critical reflection, the chapter closes with nine recommendations for examining theory through a cultural lens.

1. Notice who the authors are and explore their background, location, training, experience, and personal characteristics. Theories are not "value-free" and understanding the persons behind the theory will give greater insight into the cultural context of the authors' ideas. Notice also how the conventions of format and style of professional publications shape the author information available for inspection.

2. Consider the time period in which the theory was developed and/or came to prominence. What is the historical context of civil rights, economic trends, and political issues of that period? Pay particular attention to how the public good was defined and understood.

3. Notice whether the theory assumes an implicit strengths or deficit model of the phenomenon of interest. Theories are not neutral in their relationship to their subject matter.

4. Notice when multiculturally valid practice requires a shift in theory base. Absolute commitment to a single theory of practice may hamper cultural responsiveness.

5. Consider how the theory positions the evaluator in relation to those who are evaluated—either as program providers or consumers. Look past the rhetoric of intent to notice the power dynamics implicit or explicit in the parameters of theory. Theories named with language of empowerment or emancipation should not be exempt from scrutiny for colonizing or paternalistic stances (Smith, 1999).

6. Notice the theorist's use of language and metaphor. Language is a manifestation of power (Patton, 2000). The language of evaluation shapes wider evaluation agendas, defining the issues we consider and the answers that we seek (Kaminsky, 2000). Metaphors may be used to enrich communication and generate fresh ways of thinking about evaluation or to frame arguments, justify inclusion or exclusion of key elements, and restrict options (Smith, 1981).

7. Consider both heterogeneity within cultural domains and similarities across domains. To what extent are these taken up in this theory? Is the theory sufficiently nuanced to capture the complexity of similarities and differences, disrupting simplistic, dichotomous representation?

8. Work both within and across cultural standpoints to explore the relevance of diverse theoretical perspectives. Evaluators must listen carefully, reflect deeply, and challenge articulately the prejudice that infuses theory viewed from a particular standpoint. However, one must also avoid reducing the complexities of culture and power to a simplistic univariate model in which any single standpoint is the only consideration.

9. Consider both what's included in a particular theory and what is *not* addressed. Gordon and colleagues (1990) express this in the words of Burke (1935), who noted, "a way of seeing is also a way of not seeing." Theory must be examined for what it reveals and what it conceals.

As we evaluators hone our ability to scrutinize theory, we will gain a fuller understanding of the many ways in which culture intersects theory. Deeper appreciation of the influence of culture will in turn support theory development as well as more thoughtful selection of culturally congruent theories in the practice of evaluation. These advances in theory and practice stand to improve the multicultural validity of evaluation throughout the profession.

NOTES

1. Scheurich and Young (1997, p. 8) offer the following definition of their term: "Epistemological racism means that our current range of research epistemologies—positivism to postmodernisms/poststructuralisms—arise out of the social history and culture of the dominant race, that these epistemologies logically reflect and reinforce that social history and that racial group (while excluding the epistemologies of other races/cultures), and that this has negative results for people of color in general and scholars of color in particular."

2. Though consequential justifications of validity are most frequently identified with the work of Messick (1989b), the roots of the idea also appear in

earlier measurement literature. Moss (1992) and, most recently, Shepard (1993, 1997) have traced the historical role of social consequences of assessment use as a component of validity, noting useful comparisons among the perspectives.

3. House (1980, p. 249) grounds his broad definition of the construct in the *Oxford English Dictionary* definition of validity as "the quality of being well-founded on fact, or established on sound principles, and thoroughly applicable to the case or circumstances; soundness and strength (of argument, proof, authority, etc.)."

4. Portions of this chapter were presented as part of the panel entitled "Racism, Validity, and Program Evaluation" (Karen E. Kirkhart, Chair), Public Interest Directorate Mini-convention on Psychology and Racism: The Colors of Privilege and Power, 105th Annual Convention of the American Psychological Association, August 15–19, 1997, Chicago.

REFERENCES

Alcoff, L., & Potter, E. (Ed.) (1993). *Feminist epistemologies*. New York: Routledge.

Alkin, M. C. (Ed.) (2004). *Evaluation Roots: Tracing theorists' views and influences*. Thousand Oaks, CA: Sage.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bell, D. (1997, August). *What President Clinton should tell whites about racism and how blacks should respond*. Paper presented at the annual meeting of the American Psychological Association, Chicago.

Burke, K. (1935). *Permanence and change*. New York: New Republic.

Chen, H-T., & Rossi, P. H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review, 7*, 283–302.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443–507). Washington, DC: American Council on Education.

Cronbach, L. J. (1980). Validity on parole: How can we go straight? In W. B. Schrader (Ed.), *Measuring achievement: Progress over a decade. New Directions for Testing and Measurement*, No. 5, 99–108.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 3–17). Hillsdale, NJ: Erlbaum.

Enerstvedt, R. T. (1989). The problem of validity in social science. In S. Kvale (Ed.), *Issues of validity in qualitative research* (pp. 135–173). Lund, Sweden: Studentlitteratur.

Fetterman, D. M. (1994a). Empowerment evaluation. *Evaluation Practice, 15*(1), 1–15.

Fetterman, D. M. (1994b). Steps of empowerment evaluation: From California to Cape Town. *Evaluation and Program Planning, 17*, 305–313.

Fetterman, D. M., Kaftarian, S., & Wandersman, A. (1996). *Empowerment evaluation: Knowledge and tools for self-assessment and accountability*. Thousand Oaks, CA: Sage.

Fine, M., & Powell, L. (1997, August). *White-out: Reflections on the study of racism in psychology*. Paper presented at the annual meeting of the American Psychological Association, Chicago.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines* (3rd ed.). Boston: Allyn & Bacon.

Gordon, E. W., Miller, F., & Rollock, D. (1990). Coping with communicentric bias in knowledge production in the social sciences. *Educational Researcher, 19*(3), 14–19.

Greene, J. (1995, November). *Evaluators as advocates*. Paper presented at the annual meeting of the American Evaluation Association, Vancouver, BC, Canada.

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

Harding, S. (1998). *Is science multicultural? Postcolonialisms, feminisms, and epistemologies*. Bloomington, IN: Indiana University Press.

Hood, S. (2001). Nobody knows my name: In praise of African American evaluators who were responsive. In J. C. Greene & T. A. Abma (Eds.), *Responsive evaluation, New Directions for Evaluation*, No. 92 (pp. 31–43). San Francisco: Jossey-Bass.

Hopson, R. K. (Ed.). (2000). *How and why language matters in evaluation, New Directions for Evaluation*, No. 86. San Francisco: Jossey-Bass.

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

House, E. R. (1993). *Professional evaluation: Social impact and political consequences*. Newbury Park, CA: Sage.

Hughes, D., Seidman, E., & Williams, N. (1993). Cultural phenomena and the research enterprise: Toward a culturally anchored methodology. *American Journal of Community Psychology, 21*, 687–703.

Jagose, A. (1996). *Queer theory: An introduction*. New York: New York University Press.

Kaminsky, A. (2000). Beyond the literal: Metaphors and why they matter. In R. K. Hopson (Ed.) *How and why language matters in evaluation. New Directions for Evaluation*, No. 86 (pp. 69–80). San Francisco: Jossey-Bass.

Kershaw, T. (1992). Afrocentrism and the afrocentric method. *Western Journal of Black Studies, 16*(3), 160–168.

Kirkhart, K. E. (1995a). Multiculturalism, validity and evaluation theory. *Evaluation Theories, 2*(2), 1–3, 5–6.

Kirkhart, K. E. (1995b). Seeking multicultural validity: A postcard from the road. *Evaluation Practice, 16*(1), 1–12.

Kirkhart, K. E. (2004, February). *Multicultural competence: A question of validity in evaluation*. Keynote address presented at the conference, RACE 2004: Relevance of Assessment and Culture in Evaluation, Arizona State University, College of Education, Tempe, AZ.

Kvale, S. (1995). The social construction of validity. *Qualitative Inquiry, 1*, 19–40.

Ladner, J. A. (Ed.). (1973). *The death of white sociology*. New York: Random House.

LaFrance, J. (2003, November). *Creating an indigenous evaluation framework: Methodological implications for evaluation*. Paper presented at the annual meeting of the American Evaluation Association, Reno, NV.

Linn, R. L. (1997). Evaluating the validity of assessments: The consequences of use. *Educational Measurement, 16*(2), 14–16.

Lipsey, M. W., & Pollard, J. A. (1989). Driving toward theory in program evaluation: More models to choose from. *Evaluation and Program Planning, 12,* 317–328.

Madison, A. (Ed.). (1992). *Minority issues in program evaluation. New Directions for Program Evaluation,* No. 53. San Francisco: Jossey-Bass.

Mathie, A., & Greene, J. C. (1997). Stakeholder participation in evaluation: How important is diversity? *Evaluation and Program Planning, 20,* 279–285.

Mertens, D. M. (1998). *Research methods in education and psychology: Integrating diversity with quantitative and qualitative approaches.* Thousand Oaks, CA: Sage.

Messick, S. (1981). Evidence and ethics in the evaluation of tests. *Educational Researcher, 10*(9), 9–20.

Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 33–45). Hillsdale, NJ: Erlbaum.

Messick, S. (1989a). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5–11.

Messick, S. (1989b). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist, 50,* 741–749.

Moss, P. A. (1992). Shifting concepts of validity in educational measurement: Implications for performance assessment. *Review of Educational Research, 62*(3), 229–258.

Parker, L. (2004). Can critical theories of or on race be used in evaluation research in education? In V. G. Thomas & F. I. Stevens (Eds.), *Co-constructing a contextually responsive evaluation framework: The Talent Development Model of School Reform. New Directions for Evaluation,* No. 101 (pp. 85–93). San Francisco: Jossey-Bass.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.) Thousand Oaks, CA: Sage.

Patton, M. Q. (2000). Overview: Language matters. In R. K. Hopson (Ed.) *How and why language matters in evaluation. New Directions for Evaluation,* No. 86 (pp. 5–16). San Francisco: Jossey-Bass.

Preskill, H., & Torres, R. T. (2000). The learning dimension of evaluation use. In V. Caracelli & H. Preskill (Eds.), *The expanding scope of evaluation use. New Directions for evaluation,* No. 88 (pp. 25–37). San Francisco: Jossey-Bass.

Rosaldo, R. (1993). *Culture and truth: The remaking of social analysis.* Boston: Beacon Press.

Scheurich, J. J., & Young, M. D. (1997). Coloring epistemologies: Are our research epistemologies racially biased? *Educational Research, 26*(4), 4–16.

Schwandt, T. A. (1997). Evaluation as practical hermeneutics. *Evaluation, 3*(1), 69–83.

Scriven, M. (1991). *The evaluation thesaurus* (4th ed.). Newbury Park: Sage.

Seigart, D., & Brisolara, S. (Eds.) (2002). *Feminist evaluation: Explorations and experiences. New Directions for Evaluation,* No. 96. San Francisco: Jossey-Bass.

Shadish, W. R., Jr., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice.* Newbury Park, CA: Sage.

Shepard, L. A. (1993). Evaluating test validity. *Review of Research in Education, 19,* 405–450.

Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Education Measurement, 16*(2), 5–8, 13, 24.

Smith, L. T. (1999). *Decolonizing methodologies: Research and indigenous peoples.* New York: Zed Books, Ltd.

Smith, N. L. (Ed.) (1981). *Metaphors for evaluation: Sources of new methods: New perspectives in evaluation* (Vol. 1). Beverly Hills, CA: Sage.

Smith, N. L. (1987). Toward the justification of claims in evaluation research. *Evaluation and Program Planning, 10,* 309–314.

Stanfield, J. H., II (1993). Epistemological considerations. In J. H. Stanfield II & R. M. Dennis (Eds.), *Race and ethnicity in research methods* (pp. 16–36). Newbury Park: Sage.

Stufflebeam, D. L. (2001). *Evaluation models. New Directions for Evaluation,* No. 89. San Francisco: Jossey-Bass.

Sullivan, N. (2003). *A critical introduction to queer theory.* New York: New York University Press.

Thomas, C. L. (1997, August). Social justice and evaluative research: Implications for validity, design, and method. In K. E. Kirkhart (Chair), *Racism, validity, and program evaluation.* Symposium conducted at the annual meeting of the American Psychological Association, Chicago.

Thomas, V. G., & Stevens, F. I. (Eds.) (2004). *Co-constructing a contextually responsive evaluation framework: The Talent Development Model of School Reform. New Directions for Evaluation,* No. 101. San Francisco: Jossey-Bass.

Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. P. Connell, A. C. Kubisch, L. B. Schorr, & C. H. Weiss (Eds.), *New approaches to evaluating community initiatives: Concepts, methods, and contexts* (pp. 65–92). Washington, DC: The Aspen Institute.