American Journal of Evaluation

http://aje.sagepub.com

Toward Accurate Measurement of Participation: Rethinking the Conceptualization and Operationalization of Participatory Evaluation

Pierre-Marc Daigneault and Steve Jacob

American Journal of Evaluation 2009; 30; 330 originally published online Jul 10, 2009;
DOI: 10.1177/1098214009340580

The online version of this article can be found at:
http://aje.sagepub.com/cgi/content/abstract/30/3/330

Published by:

http://www.sagepublications.com

On behalf of:

American Evaluation Association

American Journal of Evaluation
Volume 30 Number 3
September 2009 330-348

10.1177/1098214009340580
http://aje.sagepub.com
hosted at
http://online.sagepub.com

Toward Accurate Measurement of Participation

Rethinking the Conceptualization and Operationalization of Participatory Evaluation

Pierre-Marc Daigneault
Steve Jacob
Université Laval, Quebec, Canada

While participatory evaluation (PE) constitutes an important trend in the field of evaluation, its ontology has not been systematically analyzed. As a result, the concept of PE is ambiguous and inadequately theorized. Furthermore, no existing instrument accurately measures stakeholder participation. First, this article attempts to overcome these problems by using the works of G. Goertz (2006) and J. Gerring (1999) on concept formation and evaluation to assess current conceptualizations of PE. Second, an amended version of the framework developed by J. B. Cousins and E. Whitmore (1998) is proposed as an alternative to current conceptualizations. This amended framework is then operationalized and adapted in a participation measurement instrument. The proposed conceptualization and instrument have the potential to contribute to the production of sound empirical knowledge about evaluation and to reflections on PE practice.

Keywords: *participatory evaluation; collaborative evaluation; ontology; conceptual analysis; operationalization and measurement*

Involving stakeholders in the evaluation process is a principle that is now generally accepted within the evaluation community (see Mathison, 2005; Whitmore, 1998). Some authors even refer to this trend as the "participatory orthodoxy," underlining the wide consensus on participatory methods (i.e., Biggs, 1995 as cited in Gregory, 2000, p. 180). Even for those who do not endorse participation as an ideal, the saliency in the evaluation field of such themes as stakeholder participation, inclusion, and empowerment can hardly be disputed.

While the popularity of *participatory evaluation* (PE) is good news for the proponents of stakeholder involvement, it also raises serious concerns in terms of conceptual development

Authors' Note: Correspondence concerning this article should be addressed to Pierre-Marc Daigneault, Department of Political Science, Faculty of Social Sciences, Université Laval, Pavillon Charles-De Koninck, Quebec, Canada, G1K 7P4. E-mail: pierre-marc.daigneault.1@ulaval.ca. This is a revised version of a paper presented at the annual conference of the American Evaluation Association, Baltimore, Maryland, November 10, 2007. Financial support from the Canada School of Public Service and the *Fonds québécois de la recherche sur la société et la culture* (FQRSC) to conduct this work is gratefully acknowledged. The views expressed in this document do not necessarily reflect those of the Canada School of Public Service or of the Government of Canada. We wish to thank J. Bradley Cousins, Jennifer Greene, and Jean King for their thorough critique of an earlier draft of this article. Special thanks go to Laurence Ouvrard for her feedback on the framework's operationalization. We also wish to thank Jean-François Bélanger, Martin Cossette, Nouhoun Diallo, and Mbaïrewaye Mbaï-Hadji for their comments. Finally, we are grateful to Kristen Nicole Leppington for her help with the revision of the last versions of the manuscript.

and production of empirical knowledge. As Huberman (1995) wrote more than 10 years ago: "Participatory evaluation is a noble but elusive construct. It seems to recur every thirty years in a new rhetorical guise, but it presents the same tough conceptual and practical problems" (p. 104). Many have concurred with this diagnosis (Murray, 2002; Rebien, 1996; Ridde, 2006).

PE is plagued by insufficient and/or inadequate conceptualization. First of all, multiple labels (e.g., collaborative evaluation, stakeholder evaluation, empowerment evaluation, interactive evaluation, democratic evaluation, fourth-generation evaluation, etc.) are used to characterize the same phenomenon (i.e., PE) which leads to misunderstandings among scholars and practitioners. PE is also polysemic, that is, it stretches to cover very different realities (Cousins, 2003; Garaway, 1995; Jackson & Kassam, 1998; Murray, 2002; Whitmore, 1998). For instance, this term ranges from a type of evaluation that "seeks to include program personnel in the evaluation process" (Torres et al., 2000, p. 27) to an approach which aims to be "an educational process through which social groups produce action-oriented knowledge about their reality, clarify and articulate their norms and values, and reach a consensus about further action" (Brunner & Guzman, 1989, p. 11). The polysemic nature of PE is not surprising, given the various rationales for undertaking it. Weaver and Cousins (2004) have identified three main rationales or justifications for stakeholder participation: *pragmatic* (problem-solving orientation), *political* (social justice orientation), and *epistemological* (validity of knowledge orientation). In addition, PE is inadequately theorized from an ontological perspective. Many references to participatory approaches to evaluation are found in the literature, but most of these discussions take a normative or prescriptive perspective (i.e., advocating for stakeholder involvement) or limit themselves to a vague and informal definition of the meaning of the term. Systematic analyses of the constitutive dimensions or fundamental attributes of PE are indeed relatively rare. A last problem, which in part derives from the other shortcomings mentioned, is that few satisfactory operationalizations of PE exist in evaluation literature, thereby hindering adequate measurement. This should not be surprising as "operationalizability" rests on consistent conceptualization.

This situation of conceptual ambiguity and "unoperationalization" is especially problematic in the light of repeated calls urging evaluators to undertake further empirical research about evaluation in general (e.g., Christie, 2003; Mark, 2001; Smith, 1993) and on PE in particular (e.g., Cousins, 2001; Cousins & Earl, 1999; Mark, 2001). Case reports of practice are useful in that respect, but knowledge claims that are based on such reports are not as strong as those based on systematic and rigorous empirical research.

The purpose of this article, which is divided in three parts, is to overcome these problems. The first part presents our theoretical approach to conceptualization, which is essential to the comprehension of the rest of the article. The second part proposes an amended version of the framework developed by Cousins and Whitmore (1998) and an argument defending the usefulness of this conceptualization. In the third part, a theoretically guided operationalization of the amended framework and its adaptation in a measurement instrument are presented.

Theoretical Framework to Conceptualization

Concepts play a central role in the social sciences (Gerring, 1999; Goertz, 2006). They act as building blocks for hypotheses and theories, among other functions. Various approaches to conceptualization can be mobilized for that purpose. We rely here on the frameworks developed by Gerring (1999) and especially Goertz (2006) to guide our conceptualization of PE. Using a consistent and systematic framework in the conceptualization process can substantially enhance the results of this endeavor.

Eight Criteria for Evaluating the Concept of PE

What is a valuable concept? Gerring (1999) identified eight evaluation criteria related to the functions fulfilled by concepts, namely familiarity, resonance, parsimony, coherence, differentiation, depth, theoretical utility, and field utility. According to Gerring, conceptualization is a matter of prioritization and tradeoffs between the different functions fulfilled by concepts, not of applying a "cookbook." According to this "criterial" framework, some steps can be taken to ensure that conceptualization is *pareto-optimal* which means that, beyond a certain point, improving the performance of a concept on one dimension will imply losses on other dimensions. For instance, more parsimony (the shortness of a term and of the number of its attributes) could mean less differentiation as fewer attributes are mobilized to distinguish this concept from others. Concepts should therefore be formed in relation to their purpose in a specific research endeavor.

We have already argued that PE is plagued by many problems. Translating these problems in the language of this "criterial" framework allows us to better grasp the areas needing improvement. First of all, greater *internal coherence* and *parsimony* are needed as PE is often defined with respect to an unduly long list of attributes. For instance, Jackson and Kassam (1998) have listed as many as nine defining characteristics whereas Burke (1998) has identified seven principles of PE and as many key elements of its process. Second, given the current state of the literature on PE, effectively distinguishing between participatory and non-PE (*differentiation*) is difficult. Where exactly does the border lie between PE and "conventional"[1] evaluation? Unfortunately, most authors have given only a vague definition of what they mean by PE and then do not bother to distinguish it explicitly from non-participatory approaches. Moreover, few of the current conceptualizations allow for comparing and ranking different evaluations or theoretical approaches according to their level of participation. When they do, as in the case of the framework developed by J. Bradley Cousins and his colleagues (Cousins, 2005; Cousins, Donohue, & Bloom, 1996; Cousins & Whitmore, 1998; Weaver & Cousins, 2004), the ranking of evaluation approaches is made with respect to certain process dimensions of collaboration (e.g., the diversity of participants in an evaluation) but not according to the overall degree of participation.

Emphasizing Ontology and Concept Structure

Gerring's (1999) framework allows us to make a diagnosis about PE's conceptualization, but it is less useful in pointing to possible remedies. This is why we supplement it with Goertz's (2006) framework which offers a consistent and practical guide to conceptualization. Goertz has adopted an ontological, causal, and realist perspective on conceptualization. It is ontological as it focuses on a phenomenon's essential attributes rather than on its secondary, accompanying or superficial characteristics: "Concepts are theories about ontology: they are theories about the fundamental constitutive elements of a phenomenon" (Goertz, 2006, p. 5). These essential characteristics of a concept play an explanatory role in theories and hypotheses as causal mechanisms. This approach stresses that to conceptualize is to reflect upon and analyze what a concept really *is*, that is, the phenomenon it refers to and its fundamental attributes—the disease—in contrast to the statistical and factor analytic approaches that emphasize its consequences—the symptoms. For instance, PE should be defined independently of its plausible consequences (e.g., evaluation use or empowerment) to avoid circularity.

This approach insists on concept structure and *concept-measure consistency*, which refers to "the degree to which the numeric measure reflects well the basic structure of the concept" (Goertz, 2006, p. 95). Important concepts have three levels, namely the

basic, secondary, and indicator/data levels. The *basic* level is the most general and characterizes concepts as they are used in theoretical propositions (e.g., PE increases evaluation use). The *secondary* level is made up of the constitutive dimensions or fundamental attributes of a concept. For instance, we argue that control of the evaluation process is one of PE's fundamental attributes. The most concrete level is the *indicator/data* level or operationalization level that guides the collection of empirical data about a phenomenon or, in other words, its measurement.

A sound concept has a consistent structure from its indicators to its basic level. Apart from the statistical approach, Goertz (2006) identifies two prototypical structures: (a) *necessary and sufficient condition* and (b) *family resemblance*. Mathematically, the necessary and sufficient condition concept structure is characterized by the classical logic operator "AND" or the intersection in set theory. This structure entails that all the relevant dimensions or attributes of a phenomenon are necessary (i.e., required) and jointly sufficient (i.e., no other dimension is required) for a phenomenon to fit into a concept. The second concept structure, family resemblance, is mathematically modeled by the classical logic operator "OR" or the union in set theory (Goertz, 2006). This structure allows the absence of one dimension to be compensated by the presence of another and is characterized by an "*m* of *n* rule" meaning that *m* dimensions out of *n* are needed to assume that we are in presence of the concept. Different structures can be used for different levels of conceptualization of a same concept. A common combination for concepts is the necessary and sufficient structure at the dimension level and family resemblance at the indicator level (Goertz, 2006). Indeed, we argue that PE is characterized by three necessary and jointly sufficient conditions at the secondary level (i.e., diversity of participants, extent of involvement, and control of the evaluation process) and by a family resemblance structure at the indicator level (i.e., no indicator is individually necessary; only a sufficient number of them).

In addition, Goertz (2006) has argued for considering all concepts as continuous (treating dichotomous concepts as special cases), as this reduces measurement error and allows theorists and researchers to better tackle the *negative pole* of a concept and the problem of borderline cases, which are related to the criteria of differentiation presented earlier. Goertz has contended that concept continuity can be usefully theorized using the tool of fuzzy logic and set theory (see also Ragin, 2000). In a nutshell, fuzzy logic rejects the "black or white logic" (i.e., a yes or no logic) of conventional sets and proposes that most phenomena do not fit perfectly in a clear-cut category. Partial membership is allowed and represented by intermediate scores between .00 (completely out the set) and 1.00 (completely in the set). Continuous dimensions and fuzzy logic allow for more precise and refined measurement. We will come back on the issue of indicators and dimensions aggregation using fuzzy logic when applying these tools to the conceptualization and operationalization of PE.

The frameworks and tools presented so far will assist us in proposing a conceptualization of PE that is more parsimonious, has a more consistent structure, and allows one to better differentiate it from neighboring concepts. Before going further, a quick note on our epistemological perspective is warranted. Although PE is often rooted in a constructivist, transformative, or emancipatory epistemology, it does not necessarily entail that all empirical research on this topic should be participatory or rooted in critical epistemologies. PE can indeed be studied from different perspectives. This study rests on a traditional, realist epistemology. Stated otherwise, we believe that there is a phenomenon "out there" called PE that can be measured. Furthermore, our focus is exclusively theoretical and empirical, not normative (i.e., we do not take sides with respect to the desirability of PE).

Toward the Development of an Amended Version of the Cousins and Whitmore (1998) Framework

A Useful Starting Point for Reconceptualizing PE

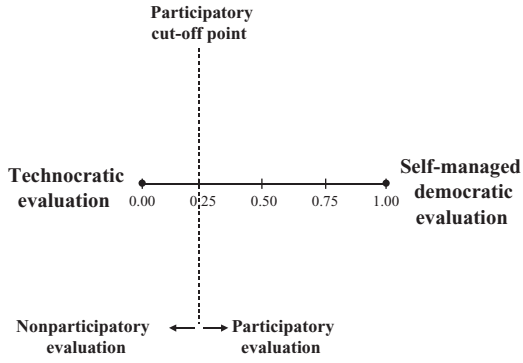
We have argued that the most serious shortcomings of current PE conceptualizations are their lack of internal coherence, parsimony, and external differentiation. We contend that the *necessary and sufficient condition structure* at the secondary level is the most appropriate aggregation procedure to overcome these challenges. Indeed, this structure emphasizes the constitutive elements that are required for an evaluation to fit into the conceptual category of PE and helps users distinguish it from non-PE. That being said, identifying the fundamental attributes of PE is not an easy matter. To maximize its usefulness in scientific research, the selection of PE's constitutive dimensions should rest as much as possible on the literature so that these dimensions are *familiar* to evaluators and stakeholders. Additionally, these dimensions must be congruent with our basic intuitions about instances of PE. When incongruence occurs, we must either revise our intuitions or modify the selection of dimensions (i.e., the method of reflective equilibrium; see Daniels, 2003).

Fortunately for us, Cousins and Whitmore (1998) have developed a valuable framework to classify various forms of PE and collaborative inquiry in general. Their conceptual framework distinguishes between two streams of PE, namely *practical* (P-PE) and *transformative* (T-PE), that are characterized by different rationales, organizational decision making and problem solving on one hand and empowerment of disadvantaged or oppressed groups on the other. The framework also extends and formalizes three dimensions that have been more or less explicit in Cousins' earlier work (see Cousins & Earl, 1992; Cousins & Earl, 1995; Cousins et al., 1996). These process dimensions are conceived as analytical tools used to describe, classify, and rank various types of collaborative inquiry or PE. The first dimension is *control of the evaluation process* (or simply the *control* dimension) which refers to the control of the technical decisions related to the conduct of evaluation as opposed to decisions about whether and when to initiate an evaluation. It is conceived as a continuum ranging from total control by the evaluator to total control by other stakeholders. The second dimension, *stakeholder selection for participation* (or simply the *diversity* dimension), refers to the types of stakeholders involved in the evaluation and ranges from the inclusion of primary users to all legitimate groups. The third dimension is *depth of participation*; it ranges from consultation (implying no decision-making authority) to deep participation, namely involvement in all evaluation tasks (i.e., design, data collection, data analysis, reporting, and decisions about dissemination of findings and use).

We argue that these three dimensions also happen to be PE's fundamental attributes or constitutive dimensions. The fact that the dimensions identified by Cousins and Whitmore (1998) correspond to the basic intuitions that many evaluators have about the nature of PE supports that claim. Using selected references, we first show that the three dimensions are more or less implicit in many discussions of participatory approaches to evaluation and already are familiar to many evaluators. We then demonstrate how the framework satisfies the concept structure of necessary and sufficient conditions for PE. We also argue that this framework helps to overcome the limitations of current conceptualizations that have been identified above.

The three constitutive dimensions of PE (diversity of participating stakeholders, involvement in the evaluation process and control) are found in a more or less explicit form in the work of many authors in the field of evaluation. For instance, Mathie and Greene (1997) have put forward the following definition of PE: "A defining feature of PE is the active engagement

Figure 1
The Participation Index and Two Polar Constructs



of multiple stakeholders'' (p. 279). Other authors have offered very similar accounts of PE, for instance, Rossi, Lipsey, and Freeman (2004): ''The participating stakeholders are directly involved in planning, conducting, and analyzing the evaluation in collaboration with the evaluator whose function might range from team leader or consultant to that of a resource person called on only as needed'' (p. 51). Similarly, King (2005) has stated that ''*participatory evaluation* is an overarching term for any evaluation approach that involves program staff or participants actively in decision making and other activities related to the planning and implementation of evaluation activities'' (p. 291). More or less implicit in these definitions are the facts that people not usually involved in traditional evaluation (i.e., ''multiple stakeholders'', ''participating stakeholders'', ''program staff or participants'') are involved in the evaluation and that their involvement takes the form of meaningful participation (i.e., ''active engagement'', direct or active involvement in decision making and other tasks).

First, who are those people not usually involved in traditional evaluation but who are involved in PE? A fair, yet imprecise, answer is *stakeholders*—a concept defined as the ''people who have a stake or a vested interest in the program, policy, or product being evaluated ... and therefore also have a stake in the evaluation'' (Greene, 2005, p. 397). This answer lacks in precision because evaluators and evaluation sponsors are also stakeholders and have always been involved in evaluation (see Forss, 1993 as cited in Rebien, 1996; King, 2005): evaluation sponsors defined the evaluation terms of reference, provided payment for the evaluation, and used the findings; evaluators conducted the evaluation per se; and, in some instances, program staff and beneficiaries provided data for the evaluation. In that broad sense, every evaluation is participatory. Nevertheless, PE departs from this picture of traditional evaluation on a major issue. PE directly involves not only the evaluator but also various actors in the process of actually producing the evaluation. In contrast, most stakeholders are not involved in traditional evaluation and when they are, it is indirectly either by planning the evaluation process or by providing data. Traditionally, evaluators and sponsors (typically decision makers) plan the evaluation and evaluators carry it out. PE is thus characterized by the fact that nonevaluative stakeholders play a significant role in the evaluation process, that is, evaluators share their tasks with other stakeholders. One could even imagine an extreme case in which all evaluative tasks are carried out by nonevaluative stakeholders without the support of an individual trained in evaluation theory, methods, and practice (see in Figure 1, *self-managed democratic evaluation*). In this case, stakeholders become de facto evaluators.

Second, the involvement of nonevaluative stakeholder takes the form of meaningful participation, which is made up of two distinct dimensions. The first is involvement in a number of different evaluation tasks, which is a necessary, but not sufficient, condition for PE. ''Participation'' might be characterized by evaluator-directed interactions with stakeholders and passive involvement of the latter in the evaluation process. In that case, stakeholders are objects of the evaluation and act as data providers. Participation can also signify that stakeholders are real subjects of the evaluation and have a significant degree of control or decision-making power over a number of issues such as methodological design (King, 2005). Control of the process is therefore the second facet of meaningful participation and the third fundamental dimension of PE. It has a broad intuitive appeal within the evaluation community and is explicitly supported by evaluation literature (Burke, 1998; Greene, 1987, 2005; King, 2005; McAllister, 1999; Murray, 2002; Rebien, 1996; Weiss, 1986, 1998). Back to the Cousins and Whitmore (1998) framework, its wide influence within and outside the field also lends credence to this claim (e.g., see Butterfoss, Francisco, & Capwell, 2001; Themessl-Huber & Grutsch, 2003): their article is indeed one of the most cited chapters ever published in *New Directions for Evaluation* (King, 2007). The dimensions of control, diversity, and depth of participation are familiar to many evaluators and satisfy one of the criteria of good conceptualization.

We contend that each of the dimensions presented above—including control—is required to classify an evaluation as participatory. Therefore, stakeholders selected for participation, depth of participation and control of the evaluation process dimensions are all *necessary* conditions for PE (see above, i.e., the section presenting our theoretical framework). Moreover, we argue that these dimensions are *jointly sufficient* for membership in this category. In other words, other ''defining'' characteristics of PE that are found in the literature are *un*necessary or nonessential conditions. For example, Burke (1998) argued that the process of PE ''*should use multiple and varied approaches to codify data*'' and ''*should explicitly aim to build capacity*'' (p. 46). Indeed, PE *might be more likely* than conventional evaluation to use more than one method to codify data or to promote capacity-building, but these elements are not necessary conditions: they are only accompanying characteristics for membership in the PE category.

A few words on the addition of two dimensions to the original Cousins and Whitmore (1998) framework are warranted as it has been argued in Cousins' subsequent works that the diversity dimension was confounded and conceptually inadequate (Cousins, 2005; Weaver & Cousins, 2004). As a result, Weaver and Cousins (2004) have recast this original dimension as an almost identical dimension, namely *diversity among stakeholders selected for participation*, and two other dimensions, *power relations among participating stakeholders* and *manageability of evaluation implementation*. These new dimensions, respectively, address power differentials between participating groups (with power relations ranging from *conflicting* to *neutral*) and the feasibility of evaluation in relation to logistics, time, and resources in particular (from *unmanageable* to *manageable*). Although these dimensions add analytical power to the original framework and thus allow for a more fine-grained and precise description and classification of evaluation approaches, they do not constitute necessary attributes of PE. Manageability of evaluation implementation is a consequence of PE or a corollary of the three original dimensions. Indeed, participatory approaches are frequently seen as time-consuming and associated with higher costs (see Butterfoss et al., 2001), but this need not be the case. Some instances of non-PEs are harder to manage than some instances of PE, for instance, a multisite evaluation using randomized control trials. Moreover, stakeholder involvement sometimes *facilitates* the research process by reducing conflict between stakeholders relative to program decision making (see Weiss, 1983, p. 12). Thus, the manageability (or unmanageability) of evaluation implementation is not a fundamental attribute of participatory

approaches. Similarly, power relations between stakeholders cannot define PE because these relations exist even in the case of non-PE and can be conflicting. Thus, while the revised and expanded framework may be useful for descriptive and classificatory purposes, considering the new dimensions as fundamental attributes of PE is inappropriate. Consequently, we uphold the original framework for the conceptualization of PE.

To sum up our argument, the three process dimensions identified by Cousins and Whitmore (1998) can logically be described as necessary constitutive dimensions of PE. In addition to their familiarity to evaluators, four further advantages of this conceptualization should also be noted. First of all, with only three dimensions required to characterize PE and to distinguish it from nonparticipatory approaches, the framework is relatively parsimonious (i.e., the list of fundamental attributes is as short as it can be for that purpose). Second, the identified fundamental attributes display an impressive degree of internal coherence. The three dimensions are logically related through the necessary and sufficient condition structure and, as process dimensions, are located in the same unit of analysis. Third, these attributes really distinguish PE from more conventional forms of evaluation, that is if one of these attributes is lacking in a given evaluation, it does not feel participatory. Finally, the framework can accommodate both streams of PE (P-PE and T-PE) and various types of evaluation in which stakeholders participate meaningfully (fourth-generation evaluation, democratic evaluation, empowerment evaluation, etc.). As a result, it can be applied in a wide range of instances.

Using the Framework as a Measurement Instrument: Some Difficulties

Although the Cousins and Whitmore (1998) framework has been very useful for theorists and practitioners in its conceptual form since its development, it has limitations with respect to the goal of *measuring* participation. To be sure, the authors should not be faulted for such imperfections as they might not have intended to translate their framework into a fully specified operational instrument to measure participation. Yet, these shortcomings are real and should be tackled through careful operationalization of the secondary-level dimensions. Before turning to this operation, a few general issues related to operationalization need to be examined.

The first issue has to do with the substantive content of each dimension and of the basic concept. At the basic level, PE constitutes the positive pole of the continuum whereas conventional or non-PE is the negative pole. "Mechanically and numerically the negative pole can be operationalized as zero on all of the secondary-level dimensions that characterize the positive extremes" (Goertz, 2006, p. 32). Now, in its current state, the Cousins and Whitmore (1998) framework does not possess an explicit zero point for all dimensions. The diversity dimension ranges from the inclusion in the evaluation process of primary users to all legitimate groups. On the continuum of stakeholder selection, no other participants may exist except for the evaluation sponsor and the evaluator, as in the case of traditional evaluation. In other words, this should be the negative pole of the diversity dimension. The conceptualization of the depth of participation dimension is incoherent because the indicators used at each pole are not of the same type. At its positive end, the depth of participation dimension is indeed clearly conceptualized as the involvement in all tasks of an evaluation, namely design, data collection, analysis, reporting, dissemination of results, and use. The polar opposite of involvement in *all* evaluation tasks is *no* involvement *in any task* rather than *consultation*.[2] Because consultation refers more to low decision-making power than involvement in only a few tasks of the evaluation process, this dimension seems to be "contaminated" by the control dimension *at its lowest bound*. This is problematic because the control issue is already taken into account by the control dimension, thereby double counting the control attribute. Stakeholders might "only" be

consulted and yet be involved in all tasks of the evaluation process. Such a case would thus receive a high score on the depth of participation dimension but not on the control dimension. In short, a coherent conceptualization of each dimension is needed to set a clear cut-off point, a threshold, between an evaluation that is participatory and one that is nonparticipatory.

A second issue relates to the specification of indicators. The Cousins and Whitmore (1998) framework gives insufficient attention to the *indicator/data level* where a phenomenon is empirically measured to examine the match between this phenomenon and a given concept (Goertz, 2006). With the exception of the two extreme values of each continuum, the authors have not given any detailed indications as to how one should apply their framework. The justification as to why a given evaluation should get that specific rating—whether numeric or nominal—is entirely left to the appreciation of the rater. Even though the five-dimension version of the framework has been applied with apparent success, the precision and reliability of the ratings seemed to have been a concern (see Ridde, 2006; Weaver & Cousins, 2004). The conceptual framework in its actual form has, thus, a limited usefulness as a *measurement* device. In addition, indicator selection must be guided by theory and be consistent with the structure of the basic concept. An explicit and consistent aggregation procedure must thus link the indicators to the secondary-level dimensions. We argue that PE should be characterized by the necessary and sufficient condition structure at the secondary level and the family resemblance structure at the indicator level.

A third issue is that the Cousins and Whitmore (1998) framework only allows one to rank an evaluation on each dimension (e.g., control) but not according to its general level of participation. It makes no mention of the structure holding dimensions together. A participation index that would allow researchers to precisely measure participation and verify covariation with other constructs such as evaluation use is therefore needed.

Together with the framework for concept formation presented earlier, these issues will inform and guide the operationalization of secondary-level dimensions so that a significant level of concept-measure consistency is achieved. The operationalization of each dimension is examined in the next section.

From Conceptualization to Measurement: Operationalizing the PE Framework

Extent of Involvement

We have renamed the depth of participation dimension as *extent of involvement* because it better describes and resonates with the idea of stakeholder involvement in a number of evaluation tasks. Stakeholders can be involved throughout the whole process but only superficially (e.g., mere presence without decision-making authority, token participation). In that case, speaking of involvement is more accurate than participation, as the latter term conveys the idea of a certain level of control over the evaluation process. The term *extent* is more appropriate than *depth* to reflect the number of tasks in which stakeholders are involved, as the latter refers more to the quality and intensity of involvement than its "quantity." Because PE is about stakeholder involvement in the production of the evaluation, we restrict the measurement of this attribute to the *technical* tasks of the evaluation process such as evaluation design and interpretation of findings (see Weaver and Cousins, 2004). By using the term technical, we stress that these tasks are more technical in nature than the decision to initiate an evaluation and the use of evaluation findings. The technical tasks are those normally considered to be the responsibility of the evaluator. Contrary to the view held by Cousins and Whitmore (1998), the

use of evaluation findings is not considered here as an evaluation task per se. We do not consider it a task of evaluators to actually use the evaluation findings to enact change, however desirable the use of evaluation is.

In PE, like in traditional evaluation, involvement in the following key decision points is essential:

1. *Evaluation questions and issues definition/methodological design* is the moment when a decision is made about the framing of the evaluation including the selection of evaluation questions and issues, theoretical framework, methods, techniques, and instruments. Guiding questions: What is the rationale for conducting the evaluation (program improvement, accountability, or knowledge production)? What is the evaluation focus (needs, processes, outputs, or outcomes)? What is the evaluation type (formative, summative, internal, external, impact, implementation, etc.)? What are the informational needs to which the evaluation can and will answer? Which criteria (relevance, effectiveness, efficiency, equity, etc.) should guide normative judgments? What type of research design is chosen for the evaluation (experimental, quasi-experimental, qualitative, quantitative, meta-analysis, etc.)? What is the methodological logic underlying the evaluation (exploratory, confirmatory, etc.)? Which sources of data will be used?
2. *Data collection and analysis* is the moment where a decision about how to concretely collect, assemble, code, and analyze data (documents, interviews, quantitative data pertaining to treatment effect, etc.) is made and when these tasks are actually carried out. Guiding questions: Who will collect, assemble, code, and analyze data? How?
3. *Judgments and recommendations formulation* is the moment where a decision is made about determining the merit and worth of a program on one hand, and formulating suggestions for future action on the other. Guiding questions: With respect to the selected quality criteria, what standard of performance is considered adequate? What is the merit and worth of this program? Why? What will be done about it?
4. *Report and dissemination of evaluation findings* is the moment where a decision is made about the reporting and diffusion of evaluation findings and their implications. Guiding questions: What communication strategy will be used? Who will be targeted?

Each task is considered a dichotomous indicator of the type *involvement of nonevaluative stakeholders in the task* (presence of the indicator) or *no involvement of nonevaluative stakeholders in the task* (absence of the indicator) where no particular indicator (e.g., involvement in a specific evaluation task such as evaluation design) is necessary for membership in the secondary-level dimension. This is a substitutability logic: as long as the number of indicators is sufficient, the extent of involvement dimension is present. Thus, involvement is defined as the presence of stakeholders (excluding the evaluator) during evaluation key moments. The assumption underlying the coding scheme is that the more tasks nonevaluative stakeholders are involved in, the more participatory an evaluation is, all other things being equal. It does not matter how many types of stakeholders are involved in each task of the process for this dimension. The unit of analysis for this dimension is the evaluation process and its tasks, not the participants. To avoid double counting, the number of participating stakeholders is only taken into account by the diversity dimension (see below).

To effectively distinguish between participatory and nonparticipatory approaches to evaluation, a cut-off point needs to be established. We argue that the involvement of nonevaluative stakeholders in one evaluation task is the minimum required for an evaluation to be considered participatory on the extent of involvement (see Table 1). In terms of the *m* of *n* rule, it means that $m = 1$ and that $n = 4$. In Table 1, the coding values represent the level of membership in the logical set of this dimension. Each indicator has thus a weight of .25. This simple scheme improves ease of use and interpretation and facilitates aggregation with other dimensions.

Table 1
Coding Scheme for Extent of Involvement

Number of Tasks Nonevaluative Stakeholders are Involved in	Level of Membership	
	Intuitive label	Numerical
0	No involvement	.00
1	Limited/weak involvement	.25
2	Moderate involvement	.50
3	Substantial/strong involvement	.75
4	Full involvement	1.00

Table 2
A Typology of Nonevaluative Stakeholders

Types	Description	Examples Drawn From the Extension
Policy makers and decision makers	People politically, legally, and organizationally accountable for the program and its evaluation	Elected and appointed officials, high-ranking civil servants, chief executive officers of nonprofit private foundations, think tanks, etc.
Implementers and deliverers	People responsible for the midlevel management and implementation of the program and the delivery of the intervention and/or services	Lower level program managers; street-level civil servants, frontline staff, and professionals (psychologists, nurses, receptionists, international development volunteers, etc.)
Target populations and intended beneficiaries; indirect beneficiaries and injured parties	People toward which the program is directed to modify their behavior and/or improve their well-being; local people indirectly and/or potentially affected by the program, either positively or negatively.	Juvenile offenders, gays and lesbians, psychotic university students, large families with violence problems, K-12 girls, drunk drivers, HIV-infected farmers, tribal council, community members (neighbors, village elders, fellow believers, classmates, local storekeepers, etc.), family members, etc.
Civil society and citizens	People and organizations having a political interest in the program and its evaluation	Interest groups, unions, think tanks, NGOs, professional associations, private firms, intellectuals, political parties, scientists, etc.

A quick example might help to clarify the selected coding. Say the frontline staff of a program (i.e., implementers and deliverers) is involved in one way or other in two tasks: (a) data collection and analysis; (b) report and dissemination of evaluation findings. Program beneficiaries, another type of stakeholders, are involved in only one task, namely data collection and analysis. Nonevaluative stakeholders are thus involved in two different tasks inasmuch as the involvement of more than one type of stakeholders in the same task, that is data collection and analysis, should not be counted twice. According to the coding scheme proposed, the extent of involvement of this specific evaluation would thus be coded as .50 and considered moderate.

Some tasks carry more weight for the conduct of evaluation than others in terms of stakeholder influence on the content of evaluation. That being said, involvement in the evaluation process has a substantive importance that is distinct from control. For example, from a learning, empowerment or process use perspective, involving stakeholders in

Table 3
Coding Scheme for Diversity of Participants

Number of Nonevaluative Stakeholder Types Involved	Level of Membership	
	Intuitive	Numerical
0	No diversity	.00
1	Limited/weak diversity	.25
2	Moderate diversity	.50
3	Substantial/strong diversity	.75
4	Full diversity	1.00

reporting the findings may be more important than involving them in issues definition and methodological design. All tasks are thus considered of equal importance (i.e. indicators are equally weighted).

Diversity of Participants

We renamed the second dimension, namely stakeholder selection for participation, as *diversity of participants*. We assume that the more diverse the types of stakeholders involved, the more participatory an evaluation is, all other things being equal. As used here, the term *stakeholders* refer to nonevaluative stakeholders, as presented in Table 2.

Our typology builds on and extends the categories of stakeholders as devised by Greene (2005). Stakeholders are classified with respect to their distance from the program and its management, decision makers being the closest and civil society the farthest. Although nothing is necessary in this relationship (e.g., unions and first-line civil servants could be directly involved in program management in corporatist settings), we contend that it represents a fairly accurate picture of program structure for most settings. Moreover, a four-category typology is convenient for coding and measurement purposes because it easily allows discriminating different evaluations in terms of their participants. The typology is intended to apply to a wide range of contexts, including the evaluation of international development aid projects. According to the proposed conceptualization, the fact that two or more organizations (e.g., the Departments of Commerce and Health) collaborate to sponsor an evaluation is not sufficient to label it participatory. Nonevaluative stakeholders need to directly contribute to the evaluative process.

The proposed coding scheme considers each category of stakeholders a dichotomous indicator of the type involvement/no involvement in a given evaluation. To be considered participatory, an evaluation must involve at least one type of nonevaluative stakeholder, thus attributing a weight of .25 to each indicator forming this secondary-level dimension (see Table 3). No particular type of nonevaluative stakeholders (e.g., target populations and intended beneficiaries) is necessary for an evaluation to be participatory. In terms of the m of n rule, it means that $m = 1$ and $n = 4$. Involvement of stakeholders is not necessarily progressive in terms of types. For instance, nothing prevents representatives from unions and lobbies (i.e., civil society groups) from being the sole nonevaluative stakeholders participating in an evaluation.

We mentioned earlier the need for a framework that adequately captures the dynamic nature of the evaluation process. For instance, frontline staff may be involved in most steps of the evaluation process, whereas the involvement of program beneficiaries is limited to defining the evaluation questions and issues. The unit of analysis for the diversity dimension is the

evaluation itself, not the different steps of its process. Whether two types of stakeholders are involved together at one step of the process or separately in two different tasks, the diversity of participants is the same. Furthermore, an evaluation that involves two types of stakeholders at one step in the process is not fundamentally different from another where involvement is sustained throughout the entire process *with respect to the diversity dimension*. Thus, diversity should measure the total number of nonevaluative stakeholder types involved in a given evaluation, not the diversity at different steps. Scoring this dimension is thus relatively straightforward: one just has to add .25 for each different type of stakeholder involved at one point or the other (excluding evaluators). If, for example, frontline staff and direct beneficiaries of a program participate in a given evaluation (disregarding the particular tasks in which they are involved), this particular evaluation would get a diversity score of .50.

Control of the Evaluation Process

PE is characterized by the fact that nonevaluative stakeholders partially or totally control the evaluation process. The assumption underlying PE's third constitutive dimension, namely control, is that the more control nonevaluative stakeholders have over the various tasks of the evaluation process in which they are involved, the more participatory an evaluation is, all other things being equal. This dimension is theorized and measured in relative terms. Thus, one has to compare the control that participants (taken as a whole) have over the process to the control the evaluator and sponsors have. An important precision has to be made here. Evaluation sponsors can also be participants in this evaluation. For instance, the board of trustees of a private foundation can sponsor an evaluation and be involved directly in the actual production of this evaluation (collecting data, formulating conclusions, etc.). In that case, sponsors are considered participants in the evaluation. Control is thus measured in terms of the share of control nonevaluative stakeholders have (in that case, decision makers that sponsor the evaluation) compared to the evaluator.

The operationalization of this dimension is somewhat less straightforward than the other two dimensions because control varies substantively during the process (Themessl-Huber & Grutsch, 2003). In contrast to diversity of participants, which is measured for the whole evaluation, control is inseparable from the different tasks of the process. Thus, this score must reflect the way control is exercised at different moments of the process even though control is measured for the whole evaluation.

We suggest two main indicators to measure control of the evaluation process. The first indicator is *authority* to make decisions, which refers to the legal and organizational power and legitimacy to decide what to do with respect to a given evaluation task. For instance, a law might force evaluators to involve teachers and parents in decisions related to the methodological design of an educational program evaluation. Whereas the first indicator refers to what could be labelled official or formal power, the second indicator captures a less formal influence of the evaluation process. Thus, *other resources of influence* refer to other resources besides authority that stakeholders can mobilize to influence the evaluation process, such as substantive and methodological expertise; money and other material resources; mobilization power; values, norms and principles; and persuasiveness and social skills. These sources of influence are indirect but, if mobilized, can lead stakeholders to have real control over the evaluation. To pursue with the preceding example, teachers could exert substantial influence over another evaluation task for which the law does not grant them official recognition (e.g., the formulation of recommendations) because they possess expertise on education and can mobilize effectively to voice this expertise. These resources are used to influence decision making directly

Table 4
Coding Scheme for Control of the Evaluation Process

Level of Membership	
Intuitive	Numerical
Exclusive control by evaluator and/or nonparticipating evaluation sponsor	.00
Limited/weak control by participants	.25
Shared control between participants and evaluator and/or nonparticipating evaluation sponsor	.50
Substantial/strong control by participants	.75
Exclusive control by participants	1.00

Note: "Participants" refer to nonevaluative stakeholders participating in the evaluation. Evaluation sponsors are considered "participants" if they are directly involved in the actual production of the evaluation.

(i.e., what is done) but also wield influence indirectly through agenda setting (i.e., what evaluation options are discussed at each stage of the process) and control of the operational setting in which the evaluation takes place (i.e., who participates, when and how, etc.).

The multifaceted nature of control makes the process of coding the control indicators and aggregating their scores not a clear-cut operation. We, therefore, propose a more intuitive and flexible coding scheme than for PE's other dimensions. Whereas the secondary-level dimension of control should still use the same scale of measurement as the other dimensions for aggregation purpose (see Table 4), we do not propose any set-in-stone procedure to derive this score from indicators. The upper limit of the control scale (see Table 4) represents the possibility that nonevaluative stakeholders become both evaluators and sponsors of the evaluation. Theoretically, good reasons exist for adhering to the family resemblance structure of aggregation which implies that no particular indicator is required for the secondary-level dimension of control to be present. Indeed, many routes may be used to influence the evaluation process. A scale ranging from .00 to 1.00 for *indicators* would allow for fine-grained measurement, but would also make it cumbersome (especially as the second indicator takes into account different types of resources: financial, expertise, persuasive, and for last normative). The question of whether one should give different weight to the indicators also makes measurement a tricky issue. Furthermore, measurement should take into account variations in control during the different steps of the evaluation process. Some kind of averaging scheme must then be devised to aggregate the different scores. Consequently, we believe coding the control dimension should be a matter of informed judgment, rather than mechanistic rule following. We do not argue that coding the control dimension should always remain a matter of informed judgment, only that it seems the most sensible option for now. We are confident that subsequent studies involving the application of this framework to real evaluations will contribute to clarifying these issues.

Combining Dimensions to Measure Participation

Now that each dimension has been operationalized, a broader understanding of participation in the evaluation is needed. How do dimensions combine to form the basic concept of PE? Once aggregated, what does PE look like? The presence of *all* three fundamental attributes of participation is required for membership in the category of PE. To be qualified as participatory, an evaluation must therefore have a score equal to or greater than .25 on each dimension and, consequently, at the basic level. This is the set-membership threshold that we

have proposed. Other thresholds are possible but this one has the advantage of being sensitive to low levels of participation while still allowing differentiation with non-PEs. Score calculation at this level is straightforward: the *minimum* score of the secondary-level dimension determines the overall PE score (Goertz, 2006). Suppose an evaluation includes all types of nonevaluative stakeholders (thereby having a score of 1.00 for the diversity dimension). This evaluation would get an overall participation score of 1.00 *only if* stakeholders are involved in all evaluation tasks (extent of involvement = 1.00) *and* if they are in total control of decision making all along the way because of the authority and informal resources they possess (control = 1.00). If, however, the score for control is .50, the same evaluation would instead get a score of .50 at the basic level. If the score on the control dimension is .00, the evaluation would not be considered participatory, even though the evaluation gets a perfect score on the other two dimensions. This is the necessary and sufficient condition logic.

Interestingly, the basic level coding scheme is also a participation index that can be used to rank different evaluations with respect to their level of participation and to distinguish participatory from non-PEs (see Figure 1). Scores reflect membership in the set of PE and their interpretation is similar to what has been done for secondary-level dimensions: .00 stands for no participation in the evaluation, .25 stands for weak/limited participation, .50 for moderate participation, and so on.

Let us now turn to the continuum formed by the basic level concept. The positive end, that is, where all dimensions are scored 1.00, can be conceived of as an *ideal type* (Goertz, 2006). It is not an ideal in the sense of a normative ideal that all evaluators should strive to attain, but rather in the sense that there are no (or very few) empirical cases that fit the type. Still, the ideal type is useful as a standard against which other types of PE could be more or less explicitly compared. We propose the label *self-managed democratic evaluation* for such an ideal type as it conveys the idea that all types or categories of nonevaluative stakeholders are totally in charge of the evaluation from beginning to end. In that case, stakeholders not only act as subjects (as opposed to objects) of the inquiry but also as both evaluation sponsors and evaluators. To our knowledge, no empirical instance of self-managed democratic evaluation exists. In regard to the polar opposite of this ideal type, which is also the negative end of the concept, many labels would adequately capture the idea of nonparticipation. Although the terms *conventional* and *traditional* evaluation are appropriate, the label *technocratic evaluation* (Murray, 2002; see also Jacob, 2005) seems to better convey the idea of nonparticipation in which an evaluation is exclusively realized by evaluators and specialists of various methodological tools and techniques for decision makers. To the extent that evaluators interact with nonevaluative stakeholders, the latter are only considered sources of data, not participants. While no or very few empirical instances of self-managed democratic evaluation exist, technocratic evaluation seems more common.

Discussion: Promises and Limitations of the Conceptualization and Measurement Instrument

Building on the Cousins and Whitmore (1998) conceptual framework, this article has proposed an expanded and clearer conceptualization of PE. We argued that *diversity of participants, extent of involvement*, and *control of the evaluation process* are PE's constitutive dimensions and thus capture the essence of PE. Conceptually, we contend that these attributes are familiar to evaluators, display a high level of internal coherence, are parsimonious, and efficiently differentiate between participatory and nonparticipatory approaches to evaluation. Moreover, the negative and positive poles of each dimension, as well as the continuum uniting

them, have been coherently theorized. Operationally, a set of explicit indicators for each dimension and aggregation rules have been have been proposed, although clear measurement procedures of the control dimension still need to be developed. Although the measurement framework still has to be fully validated, we contend that the proposed empirical measures of PE seem to display a high level of concept-measure consistency, as well as a priori validity. On one hand, the proposed indicators seem to have *face validity* inasmuch as they measure what they are intended to measure (see Singleton, Straits, & Straits, 1993). On the other hand, the selected empirical measures seem to have *content validity* in that they appear to adequately cover all facets or dimensions of PE (see Babbie & Benaquisto, 2002; Singleton et al., 1993). The measurement device has indeed a strong theoretical base to support its dimensions.

A number of limitations also need to be acknowledged, however, with respect to the proposed framework. First of all, the operationalization of PE is still incomplete. Operationalization is not limited to the specification of indicators; it also entails identifying specific procedures for scoring each indicator (Singleton et al., 1993). This weakness is particularly striking for the control indicators as the specific procedures for scoring each indicator, for averaging the scores across the steps of the evaluation process, and for aggregating those scores into the second-level dimension have not yet been formalized. Consequently, the measurement instrument is not yet totally ready to use. Further studies on PE will be needed to derive a fully operational device. Second, despite what has been said above on content validity, two facets of participation have deliberately been excluded from the framework. With respect to the extent of involvement, the proposed conceptualization does not take into account the *intensity* of participation and, as such, does not exhaust the meaning of participation. Some stakeholders are indeed more enthusiastic, proactive, and affectively and intellectually engaged than others. In addition, we have focused on diversity of *types* of stakeholders as opposed to diversity *within* and *across* particular types of stakeholder groups in terms of values, opinions, socioeconomic characteristics, gender, ethnicity, and language. A last limitation pertains to the neglect of some sites of power. The extent of involvement dimension focuses on the technical tasks of the evaluation process but not on the political task of actually initiating (or not) an evaluation, writing the terms of reference or using the results. These sites of power are excluded from our analyses on participation. At the same time, however, tradeoffs are inescapable in any research endeavor. We have thus strived for a balance between simplicity and precision.

We believe that the proposed conceptualization and instrument are going to be useful to both evaluation practitioners and researchers. From a practice perspective, the proposed framework may help evaluators to be more aware of their practices with respect to participation and foster reflection on them. For instance, the framework can be used in self-evaluation to compare the subjective beliefs of the evaluator (e.g., ''I am deeply committed to PE'') with a more objective measure of participation (e.g., ''According to the framework, the last evaluation I have conducted was weakly participatory''). This use of our framework is consistent with point 4 of the competence component of the American Evaluation Association (2004) *Guiding Principles for Evaluators* which states that:

> Evaluators should continually seek to maintain and improve their competencies, in order to provide the highest level of performance in their evaluations. This continuing professional development might include formal coursework and workshops, self-study, evaluations of one's own practice, and working with other evaluators to learn from their skills and expertise.

The framework's utility as a self-evaluation tool is in no way limited to individual evaluators. Because many governments, international organizations, and departments claim that

participation is an important ideal in evaluation, assessing whether this reflects the reality of practice would be illuminating.

The potential contribution of the framework for research is also important. A promising area of research relates to predicting and explaining the consequences of participatory approaches. Many studies on the relationship between PE and the use of evaluation and empowerment display a low level of precision in terms of measurement. Low comparability and generalizability of studies are the rule. We contend that the proposed device could foster quality research in this area by offering a more objective and systematic way to measure participation. Many research questions could benefit from the use of the proposed framework: Is the relationship between PE and evaluation use linear (i.e., does more participation always lead to greater use)? What is the effect size of this relationship? Do we observe differences for participatory effects between different streams of PE (P-PE vs. T-PE)? Are the *power relations among participating stakeholders* and *manageability of evaluation implementation* dimensions identified by the mediating factors for evaluation use of Weaver and Cousins (2004)? At a micro level, is evaluation influence a function of the specific types of nonevaluative stakeholders and the specific steps in which they are involved?

Many calls urging evaluators to undertake further empirical research on evaluation in general (e.g., Christie, 2003; Mark, 2001; Smith, 1993) and on PE in particular (e.g., Cousins, 2001; Cousins & Earl, 1999; Mark, 2001) have been made in the past. The proposed conceptualization and instrument is a tool that has the potential to help evaluators meet this challenge.

Notes

1. Various terms have been used by different authors to refer to nonparticipatory evaluation, for instance *conventional evaluation* (Rebien, 1996), *distanced evaluation* (O'Sullivan & D'Agostino, 2002), *independent evaluation* (Rossi et al., 2004), *old-style evaluation* (Weiss, 1986), *technocratic evaluation* (Murray, 2002), and *traditional evaluation* (VanderPlaat, Samson, & Raven, 2001). Even though these terms are ambiguous and need further specification, we will use the terms conventional or traditional throughout the text to qualify evaluations that are not participatory.

2. In all justice to the designers of the conceptual framework, it must be stressed that the "true" negative pole of PE has received some attention in an earlier article. The lower bound of the depth of participation dimension has indeed been labeled "No Participation/Consultation Only" (Cousins et al., 1996, p. 211, see also pp. 209-210). Even in that case, however, the use of "consultation" is still problematic.

References

American Evaluation Association. (2004). *Guiding principles for evaluators.* Fairhaven, MA: AEA.

Babbie, E. R., & Benaquisto, L. (2002). *Fundamentals of social research* (1st Canadian ed.). Scarborough, Ontario: Nelson Thomson Learning.

Brunner, I., & Guzman, A. (1989). Participatory evaluation: A tool to assess projects and empower people. *New Directions for Program Evaluation*, *42*, 9-18.

Burke, B. (1998). Evaluating for a change: Reflections on participatory methodology. *New Directions for Evaluation*, *80*, 43-56.

Butterfoss, F. D., Francisco, V., & Capwell, E. M. (2001). Stakeholder participation in evaluation. *Health Promotion Practice*, *2*, 114-119.

Christie, C. A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Directions for Evaluation*, *97*, 7-35.

Cousins, J. B. (2001). Do evaluator and program practitioner perspective converge in collaborative evaluation? *Canadian Journal of Program Evaluation*, *16*, 113-133.

Cousins, J. B. (2003). Utilization effects of participatory evaluation. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation* (Vol. 9, pp. 245-265). Dordrecht, The Netherlands: Kluwer Academic.

Cousins, J. B. (2005). Will the real empowerment evaluation please stand up? A critical friend perspective. In D. M. Fetterman & A. Wandersman (Eds.), *Empowerment evaluation in practice* (pp. 183-208). New York: Guilford Press.

Cousins, J. B., Donohue, J. J., & Bloom, G. A. (1996). Collaborative evaluation in North America: Evaluators' self-reported opinions, practices and consequences. *Evaluation Practice*, *17*, 207-226.

Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis*, *14*, 397-418.

Cousins, J. B., & Earl, L. M. (1995). Participatory evaluation in education: What do we know? Where do we go? In J. B. Cousins & L. M. Earl (Eds.), *Participatory evaluation in education: Studies in evaluation use and organizational learning* (pp. 159-180). London: Falmer Press.

Cousins, J. B., & Earl, L. M. (1999). When the boat gets missed: Response to M. F. Smith. *American Journal of Evaluation*, *20*, 309-317.

Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, *80*, 5-23.

Daniels, N. (2003). Reflective equilibrium. *The Stanford encyclopedia of philosophy* (Summer 2003 ed.). Retrieved May 23, 2008, from http://plato.stanford.edu/entries/reflective-equilibrium/

Garaway, G. B. (1995). Participatory evaluation. *Studies in Educational Evaluation*, *21*, 85-102.

Gerring, J. (1999). What makes a concept good? A criterial framework for understanding concept formation in the social sciences. *Polity*, *31*, 357-393.

Goertz, G. (2006). *Social science concepts: A user's guide.* New Jersey: Princeton University Press.

Greene, J. C. (1987). Stakeholder participation in evaluation design: Is it worth the effort? *Evaluation and Program Planning*, *10*, 379-394.

Greene, J. C. (2005). Stakeholders. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 397-398). Thousand Oaks, CA: Sage.

Gregory, A. (2000). Problematizing participation: A critical review of approaches to participation in evaluation theory. *Evaluation*, *6*, 179-199.

Huberman, M. (1995). The many modes of participatory evaluation. In J. B. Cousins & L. M. Earl (Eds.), *Participatory evaluation in education: Studies in evaluation use and organizational learning* (pp. 103-111). London: Falmer Press.

Jackson, E. T., & Kassam, Y. (1998). Introduction. In E. T. Jackson & Y. Kassam (Eds.), *Knowledge shared: Participatory evaluation in development cooperation* (pp. 1-20). Ottawa, ON/West Hartford, CT: International Development Research Centre/Kumarian Press.

Jacob, S. (2005). Réflexions autour d'une typologie des dispositifs institutionnels d'évaluation [Reflections around a typology of evaluation institutional mechanisms]. *Canadian Journal of Program Evaluation*, *20*, 49-68.

King, J. A. (2005). Participatory evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 291-294). Thousand Oaks, CA: Sage.

King, J. A. (2007). Making sense of participatory evaluation. *New Directions for Evaluation*, *114*, 83-86.

Mark, M. M. (2001). Evaluation's future: Furor, futile, or fertile? *American Journal of Evaluation*, *22*, 457-479.

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

Mathison, S. (Ed.). (2005). *Encyclopedia of evaluation.* Thousand Oaks, CA: Sage.

McAllister, K. (1999). *Understanding participation: Monitoring and evaluating process, outputs and outcomes.* Ottawa, ON: International Development Research Centre.

Murray, R. (2002). Citizens' control of evaluations: Formulating and assessing alternatives. *Evaluation*, *8*, 81-100.

O'Sullivan, R. G., & D'Agostino, A. (2002). Promoting evaluation through collaboration: Findings from community-based programs for young children and their families. *Evaluation*, *8*, 372-387.

Ragin, C. C. (2000). *Fuzzy-set social science.* University of Chicago Press.

Rebien, C. C. (1996). Participatory evaluation of development assistance: Dealing with power and facilitative learning. *Evaluation*, *2*, 151-171.

Ridde, V. (2006). Suggestions d'améliorations d'un cadre conceptuel de l'évaluation participative [Suggestions for improving a conceptual framework for participatory evaluation]. *Canadian Journal of Program Evaluation*, *21*, 1-23.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.

Singleton, R., Straits, B. C., & Straits, M. M. (1993). *Approaches to social research* (2nd ed.). New York: Oxford University Press.

Smith, N. L. (1993). Improving evaluation theory through the empirical study of evaluation practice. *American Journal of Evaluation*, *14*, 237-242.

Themessl-Huber, M. T., & Grutsch, M. A. (2003). The shifting locus of control in participatory evaluations. *Evaluation*, *9*, 92-111.

Torres, R. T., Stone, S. P., Butkus, D. L., Hook, B. B., Casey, J., & Arens, S. A. (2000). Dialogue and reflection in a collaborative evaluation: Stakeholder and evaluator voices. *New Directions for Evaluation*, *85*, 27-38.

VanderPlaat, M., Samson, Y., & Raven, P. (2001). The politics and practice of empowerment evaluation and social interventions: Lessons from the Atlantic Community Action Program for Children regional evaluation. *Canadian Journal of Program Evaluation*, *16*, 79-98.

Weaver, L., & Cousins, J. B. (2004). Unpacking the participatory process [Electronic Version]. *Journal of MultiDisciplinary Evaluation (JMDE), 1*, 19-40. Retrieved July 15, 2008, from http://survey.ate.wmich.edu/jmde/index.php/jmde_1/article/view/144/159

Weiss, C. H. (1983). The stakeholder approach to evaluation: Origins and promise. *New Directions for Program Evaluation*, *17*, 3-14.

Weiss, C. H. (1986). The stakeholder approach to evaluation: Origins and promise. In E. R. House (Ed.), *New directions in educational evaluation* (pp. 145-157). London: Falmer Press.

Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Whitmore, E. (1998). Editor's notes. *New Directions for Evaluation*, *80*, 1-3.