Journal of MultiDisciplinary Evaluation
Volume 10, Issue 22, 2014

JMDE
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180
http://www.jmde.com

Developing Criteria to Identify Transformative Participatory Evaluators

Michael A. Harnar*
Claremont Graduate University

Background: The evaluation discipline continues to evolve as more and more researchers study practice. The research described in this article further defines Transformative Participatory Evaluation (T-PE) by focusing on the key elements that practitioners and theorists agree define this evaluation practice.

Purpose: A multi-stage, mixed-method approach was used to develop and examine a set of statements that serve two purposes: First, they can help identify a subset of participatory practitioners from others and, second, they further theory development by showing how T-PE practitioners differ from other evaluation practitioners on key indicators.

Setting: In the first phase of this research, three prominent evaluation theorists comprised an expert panel to develop a set of statements that would identify T-PE practitioners. The American Evaluation Association membership was used to test the statements in the research's subsequent phases.

Intervention: NA

Research Design: A multi-stage, mixed-method approach was used to develop and test the statements.

Data Collection and Analysis: The panel was engaged in a web-based wiki to jointly edit the statements; an online questionnaire with mostly closed-ended items was used to gather AEA member input; a unique online modeling software and webinars were used to further understand findings. Analysis of variance was used to assess differences between groups and Rasch modeling and Wald tests were used to analyze the modeling data.

Findings: The eight core statements that emerged had acceptable internal reliability and limited construct validity. Though the statements' discrimination strength was tenuous, quantitative comparisons of preferred evaluation practice models showed congruence with the predicted underlying philosophies and therefore supports the statements' ability to discern T-PE evaluators from P-PE evaluators.

Keywords: *Transformative Participatory Evaluation; Participatory Evaluation; evaluation*

* Michael Harnar is now an independent evaluation consultant. Correspondence concerning this article should be addressed to Michael Harnar, 315 Adams Ave, Pomona, CA 91767. Contact: michaelharnar@gmail.com

Introduction

The evaluation discipline continues to evolve as more and more researchers study practice in attempts to further define its constituent theories. While much of this work has been met with enthusiastic support, recent attempts by the Collaborative, Participatory, and Empowerment Evaluation Topical Interest Group of the American Evaluation Association to refine the definitions of subgroups of collaborative evaluation theories have seen some criticism (Cousins, Whitmore, & Shulha, 2013). Cousins et al. contend that attempting to parse collaborative forms of evaluation by finer distinctions may be counterproductive and lack value in moving the discipline forward. The current study, while not intending to address this quandary directly, adds an informative example of research further defining one collaborative evaluation theory.

Cousins and Whitmore (1998) proffered two streams of participatory evaluation—transformative participatory evaluation (T-PE) and practical participatory evaluation (P-PE). Subsequent research has left the former untested while the latter has gained significant attention. It may be that, as they recently portended, these two streams are not all that dissimilar, and efforts to parse them in practice may lack value in moving the discipline forward. It is the position of this researcher that more understanding of our theories in practice can only aid in the evolution of our discipline. Thus, the research described in this article intends to bring more attention to T-PE by focusing on the key elements that its practitioners agree define their preferred evaluation practice.

A multi-stage, mixed-method approach (Greene, 2007) was used to develop and examine a set of statements to potentially describe transformative participatory evaluators and distinguish them from other participatory evaluators. In making this distinction, this tool may allow us to observe transformative participatory evaluation practice and compare it to other collaborative forms of evaluation in order to better understand its distinctiveness. Thus, the findings can further theory development by showing how these practitioners differ from others on key indicators.

Literature Review

Researchers in general focus on creating and understanding basic knowledge as it *might* be applied to real world problems. Applied researchers (e.g., evaluators) take this a step further and develop *new* knowledge in the direct pursuit of solving those problems. Evaluation is therefore a practitioner-based discipline and evaluation theory is derived from practice (Shadish, 1998). Our theories describe evaluation practice, but because few are empirically supported, they are prescriptive in nature (Alkin, 2004). That is, they provide guidelines for practicing evaluation in some manner depending on certain contextual issues, such as the questions guiding an evaluation or the potential for using the evaluation findings.

Evaluation theories do not say "in 95 out of 100 evaluations if you provide this type of data at this point in the evaluation to this particular set of stakeholders you will get this type of outcome." That requires extensive empirical study that takes into account all of the various issues and variables that interact within any given evaluation context so that prediction models might provide support for all potential outcomes of each choice. While this may be an ideal definition of descriptive or contingency theories, the discipline does, in fact, intend to move in that general direction so that it becomes a well-described and documented field. Shadish (1998) provided clear reasons why evaluation theory is important: It is what we talk about, it is what drives our conversations, it is the nomenclature that gives us a framework to guide practice, and it is what serves the researcher interests in many of us. Without evaluation theory, evaluation practice would be "too scattered, too ill-defined, and too vulnerable to poaching by the many other people who also claim that they can do evaluative work as well as we can" (Shadish, 1998, p. 13).

A deeper understanding of a full range of variables, issues, and contexts can aid in judging the merit of professional versus novice approaches in evaluation practice (Shadish, 1998). Those who are aware of the potential applications of various theories based on contextual variables will be more experienced and competent evaluators, comfortable with various applications. More novice evaluators may not have the necessary experience. When the evaluation discipline develops its theories to the point that all (or most) contingencies in the field of practice can be controlled for, then teaching new evaluators will be an easier feat, and the practice of evaluation more standardized and replicable (Miller, 2010).

In an effort to bring more definition to participatory evaluation, Cousins and Whitmore (1998) parsed it into two broad approaches—transformative participatory evaluation (T-PE) and practical participatory evaluation (P-PE). P-PE is seen mostly as a North American practice

(Brisolara, 1998; Cousins & Whitmore, 1998) focused on stakeholder involvement to foster greater relevance, ownership, and use (Cousins & Earl, 1992; Greene, 1988a, 1988b; Patton, 1997). T-PE uses many of the same processes as P-PE, but intends to produce social change by empowering the disempowered. This is similar to participatory action research's (PAR) focus on power redistribution (Brisolara, 1998; Estrella & Gaventa, 1997; Greenwood, Whyte, & Harkavy, 1993; Sabo, 1999; Suarez-Balcazar & Harper, 2003) and the notion of empowerment evaluation (Fetterman, 2005). While both T-PE and empowerment evaluation focus on empowering the disempowered, however, a T-PE evaluator maintains more technical control and is more engaged in managing and directing the evaluation than an empowerment evaluation evaluator. T-PE empowers participants through varied data collection strategies that encourage joint knowledge creation (Burke, 1998; Meyer, Park, Grenot-Scheyer, Schwartz, & Harry, 1998; Whitmore, 1988). Likewise, in the youth development field, evaluation is another opportunity to offer youth more ways to take ownership of their lives and development (Checkoway & Gutierrez, 2006; Checkoway & Richards-Schuster, 2003; Sabo Flores, 2008; Sabo, 1999).

The role of human agency in knowledge creation also helps define T-PE. Evaluation participants produce socially constructed knowledge through dialogue. As knowledge informs the evaluation, its creators are empowered by seeing their knowledge being used (e.g., Brisolara, 1998; Cousins & Whitmore, 1998; Sabo, 1999). As a result, participatory evaluators and evaluand stakeholders develop close relationships, mutual respect, and deep understandings (Gaventa, 1993; King, 1998). Also, participants gain a greater sense of control and agency when they see their knowledge put to use in a respectful, team-oriented manner.

While these descriptions may lead one to believe these two streams are well supported, there is actually limited empirical research supporting Cousins and Whitmore's (1998) definition of transformative participatory evaluation (for example, see Cousins & Chouinard, 2012). Relative to other forms of participatory evaluation (e.g., practical participatory, empowerment, collaborative), there is a paucity of empirical research on T-PE. The definition of T-PE is therefore dependent upon only Cousin and Whitmore's supposition (1998). And beyond the Cousins and Choinard (2012) book, there is limited independent research focused on understanding T-PE in practice. It is therefore a form of evaluation with a prescriptive theory that lacks substantial empirical support.

Because evaluation research is informed by practice, one way of further developing T-PE descriptions is to identify practitioners of T-PE and observe their practice. Identifying T-PE practitioners is a necessary step in that process and this research focuses on developing a set of statements as criteria to identify T-PE practitioners as distinct from other evaluators. These statements were tested on a population of evaluators, using convergent validity and internal reliability measures, to see if they identify what the extant research describes as T-PE practitioners.

Methodology

Procedures

Three prominent evaluation theorists were invited as a purposeful panel to develop a set of statements that would identify transformative participatory evaluators from other practitioners: Drs. Elizabeth Whitmore, J. Bradley Cousins, and Donna Mertens. These three were selected because they had well-established publication records focused on participatory and transformative evaluation. Eight statements were created through a deductive process guided by Cousins and Whitmore's (1998) framework and Mertens' (2009) description of transformative evaluation. Statements were created to identify practitioners who were relatively more participatory and philosophically transformative in their principles, activities, and outcomes. The statements were then honed through an iterative process with the panel of experts. In short, the researcher generated a preliminary set of statements and the panelists reduced it to a smaller set through online interactions. The statements were also edited for clarity during the survey's pilot phase. (See Harnar, 2012, for a full discription of the process.)

The final list of statements reflected four components of participatory evaluation: depth of participation, selection of stakeholders, control of the evaluation, and philosophical preference for individual and program transformation and social justice. In order to understand if these statements could actually identify T-PE practitioners, study participants were asked their agreement on the statements. Their responses were then used to categorize them as T-PE, P-PE, or PE, and other survey responses were used to more closely examine these categorizations. If a practitioner agreed with all of these statements, it was

determined that they would fit the literature's definition of a T-PE evaluator. The statements are provided in Table 1 and categorized by the dimensions they address.

Table 1
Eight Identifying Statements

Dimension	Statement
Depth	I always try to involve non-evaluator participants in my evaluations.
	I prefer not to take on an evaluation unless it has a strong participatory component.
Selection	Program beneficiaries should participate in carrying out evaluation.
	People representing all important perspectives should be involved in any evaluation.
Control	Evaluators should share technical decision-making with non-evaluator participants.
Transformative Philosophy	Evaluators should help train all legitimate groups to do evaluation.
	Evaluation should focus on bringing about individual empowerment, emancipation, and self-determination.
	Evaluation should focus on bringing about social justice.

Participants

The American Evaluation Association (AEA) is the largest association of evaluators in the world, with a membership of more than 7,000 individuals representing all 50 states and more than 60 countries. Its members comprise a broad spectrum of evaluators and evaluation-interested people working in a variety of contexts. AEA members were selected for participation in the research because they represent the largest sample of evaluators available in one place, and because AEA practitioners represent a broad range of approaches that was expected to provide a large sample of T-PE evaluators.

After a cleaning process, the AEA membership list included 6,615 potential subjects who were contacted by email. In response to the first invitation, 546 respondents opted out, 84 email addresses bounced, and 43 email addresses resulted in "out of office" replies for the remainder of the survey period. Another 208 opted out in response to a follow-up reminder. Combined, the opt-out rate was 11.4%. Overall, 1,323 individuals began the survey, but 59 dropped out before providing more than cursory data, and another 36 did not provide complete data. The final sample providing complete data was 1,228, yielding a response rate for the entire sample of 18.6%; after removing the bounced addresses, the response rate rose to 18.8%.

Sample

Almost all the survey respondents (Table 2) were active practitioners (94%), with about two-thirds (66%) conducting from one to six evaluations each year, and just over a quarter (28%) conducting seven or more. Almost half (48.4%) had been in the evaluation field for 10 years or fewer and they were evenly split on primary or secondary identity as evaluators (44.6% and 44.2%, respectively). Most (43%) saw themselves as having intermediate knowledge and experience and had either master's or doctorate degrees (43.8% and 44%, respectively). There was a great diversity of disciplines represented in the degrees held by this sample, with education and psychology standing above the rest.

Table 2 also includes an overview of the evaluation characteristics of the respondents. As practitioners, participants said they preferred a broad range of theoretical orientations. To determine their orientations, they were asked the following question:

> *Is your preferred theoretical orientation similar to any of these? I know that many evaluators say that they design evaluations that are context specific and none of these orientations covers every evaluation. But, I also know that you probably have a perspective you "prefer."*

Utilization-focused evaluation (UFE) was most endorsed (24.3%), and participatory evaluation gained about half as many (11.5%) selections. The respondents typically either did a mixture of internal and external evaluations (34.4%) or external evaluations only (32.1%). Most did program evaluations (88.4%) in a variety of contexts, with nonprofits (36.7%), health (35.9%), and K–12 education (34.2%) holding the top three spots. Their settings were mostly academia (38.6%), non-profits (25.6%) or private business/consulting (22.7%), and they general worked in North America (75.1%).

Table 2
Survey Participants' Characteristics

Variable		*n*	%
Evaluations Per Year			
1–3		528	39.9%
4–6		345	26.1%
7 or more		370	28.0%
None		75	5.7%
Blank		5	0.4%
	Total	1,323	100%
Years in Evaluation			
Less than two		87	6.6%
2–5		280	21.2%
6–10		273	20.6%
11–15		227	17.2%
16–20		113	8.5%
More than 20		185	14.0%
Blank		158	11.9%
	Total	1,323	100%
Evaluation Identity			
Primary		590	44.6%
Secondary		585	44.2%
Not my professional identity		62	4.7%
Blank		86	6.5%
	Total	1,323	100%
Evaluation Knowledge and Experience			
A relative beginner		190	14.4%
At an intermediate level		569	43.0%
At an advanced level		403	30.5%
Blank		161	12.2%
	Total	1,323	100%
Highest Education Level Completed			
High school/some college		1	0.1%
Associate's degree		2	0.2%
Bachelor's degree		70	5.3%
Master's degree		580	43.8%
Doctoral degree		582	44.0%
Blank		88	6.7%
	Total	1,323	100%
Field of Your Highest Degree			
Education		229	17.3%
Psychology		179	13.5%
Evaluation/Research methods		118	8.9%
Public health		103	7.8%
Public policy		76	5.7%
Sociology		62	4.7%
Business		38	2.9%

Economics		29	2.2%
Social welfare		29	2.2%
Anthropology		20	1.5%
Advanced quantitative methods		9	0.7%
Nursing/Medicine		9	0.7%
School administration		6	0.5%
Advanced qualitative methods		1	0.1%
Art/Music		1	0.1%
Other			
Applied social science		156	48.0%
Social science		78	24.0%
Natural science		30	9.2%
Humanities		27	8.3%
Applied science		14	4.3%
Formal science		9	2.6%
Interdisciplinary		6	1.7%
(blank)		5	0.3%
	Subtotal	325	100%
Decline to answer		2	0.2%
(blank)		87	6.6%
	Total	1,323	100%
Preferred Theoretical Orientation			
Utilization-focused		321	24.3%
Participatory evaluation		152	11.5%
Evaluation research		97	7.3%
Theory-driven		95	7.2%
Developmental evaluation		57	4.3%
Empowerment evaluation		37	2.8%
CIPP Model		30	2.3%
Stakeholder evaluation		26	2.0%
Social justice-driven		22	1.7%
Fourth generation evaluation		5	0.4%
Connoisseurship evaluation		1	0.1%
My theoretical orientation is not listed here		41	3.1%
I do not have a preferred theoretical orientation		166	12.5%
I do not know enough about these to select one		112	8.5%
(blank)		161	12.2%
	Total	1,323	100%
Role as an Evaluator			
External		424	32.1%
Internal		277	20.9%
Mix of internal & external		455	34.4%
(blank)		167	12.6%
	Total	1,323	100%
Primary Type(s) of Evaluations Performed			
Program evaluations		1,169	88.4%
Performance auditing/monitoring/reviewing		458	34.6%
Policy evaluations		330	24.9%
Curricula evaluations		266	20.1%
Evaluation of research		261	19.7%

Student/Trainee evaluations		174	13.2%
Personnel evaluations		162	12.2%
Consumer evaluations		84	6.4%
Product evaluations		57	4.3%
I do not do evaluations		4	0.3%
	Total	2,965	
Primary Context(s) of Evaluations Performed			
Nonprofits		486	36.7%
Health/Public health		475	35.9%
K–12 education		452	34.2%
Higher education		399	30.2%
Youth development		332	25.1%
Adult education		318	24.0%
Government		288	21.8%
Child care/Early childhood education		281	21.2%
Advocacy and policy change		258	19.5%
Human services		256	19.4%
Public policy/Public administration		251	19.0%
Evaluation methods		241	18.2%
STEM		227	17.2%
Educational technologies		212	16.0%
Special needs populations		196	14.8%
Organizational behavior		193	14.6%
Workforce/Economic development		188	14.2%
Alcohol or drug abuse		178	13.5%
Foundations		170	12.9%
Social justice		156	11.8%
International/Cross-cultural		143	10.8%
Environmental programs		141	10.7%
Social work		139	10.5%
Human development		129	9.8%
Indigenous peoples		105	7.9%
Business and industry		93	7.0%
Law or criminal justice		80	6.1%
Medicine		78	5.9%
Disaster/Emergency management		76	5.7%
Gender rights		71	5.4%
Human resources		71	5.4%
Information systems		67	5.7%
Media		47	3.6%
LGBT		41	3.1%
	Total	6,838	
Primary Setting(s) of Evaluations Performed			
College or university		511	38.6%
Nonprofit foundation/organization		339	25.6%
Private business or consulting		300	22.7%
Federal government agency		108	8.2%
State/Provincial government agency		104	7.9%
School system		91	6.9%

Local government agency		40	3.0%
	Total	1,493	
Primary Location(s) of Evaluations Performed			
North America		994	75.1%
Asia		104	7.9%
Africa		102	7.7%
Europe		56	4.2%
South America		48	3.6%
Central America		35	2.7%
Australia/New Zealand		31	2.3%
	Total	1,370	

Results

Categorization of Respondents

Responses to the full set of eight T-PE statements were used to place each participant into one of four groups. Specifically, survey respondents were asked to indicate how strongly they agreed or disagreed with the eight statements. They were provided with the following instructions and working definitions:

> *In answering these questions, please think about how you prefer to practice evaluation. I know that answers to these questions are almost always context dependent, and "it depends" might be your answer choice. But, I would like you to think of your ideal evaluation situation.*
>
> *The term "stakeholder" is used here to mean anyone, other than the evaluator, with a vested interest in the entity (evaluand) being evaluated.*
>
> *"Participants" are those stakeholders who take an active role in the evaluation.*
>
> *"Participation" is any active role and may vary widely in breadth and depth.*

As shown in Table 3, 332 responded positively (agreed or strongly agreed) to all eight T-PE statements and were therefore labeled T-PE evaluators. The small group who disagreed or strongly disagreed with the key principles of social justice and empowerment but agreed or strongly agreed with the other six statements comprised a comparison group and were labeled as practical participatory evaluators (P-PE) (n = 54). Those who agreed or strongly agreed with only three statements related to participatory evaluation were labeled as participatory evaluators (PE) (n = 198). The others were labeled Non-PE (n = 739).

Table 3
Participants' Participatory Evaluation Categories

Grouping		***n***	**%**
T-PE		332	25.1%
P-PE		54	4.1%
PE		198	15.0%
Non-PE		739	55.8%
	Total	240	100%

Levels of Agreement with Statements

As expected, the eight statements drew generally positive responses from participants. More than two-thirds (78.7%) agreed or strongly agreed with the whole set, compared to just over one-fifth (21.2%) who disagreed or strongly disagreed with all eight (see Table 4). This is somewhat reflective of previous research with regards to participation. More specifically, the broad support (95.3%) for the statement, "I always try to involve stakeholders in my evaluations" echoed findings in prior studies (e.g., Cousins & Earl, 1992; Fleischer & Christie, 2009; Preskill & Caracelli, 1997).

Table 4
Participants' Responses to T-PE Statements

Statement	Dimension	Strongly Agree		Somewhat Agree		Somewhat Disagree		Strongly Disagree		Total
		n	%	*n*	%	*n*	%	*n*	%	*N*
Intended program beneficiaries should participate in carrying out evaluation.	Selection	407	34.3%	533	44.9%	195	16.4%	53	4.5%	1,188
People representing all important perspectives should be involved in any evaluation.	Selection	747	62.9%	359	30.2%	67	5.6%	14	1.2%	1,187
I always try to involve stakeholders in my evaluations.	Depth	795	67.0%	336	28.3%	43	3.6%	7	0.6%	1,181
I prefer not to take on an evaluation unless it has a strong participatory component.	Depth	197	16.6%	441	37.2%	431	36.3%	112	9.4%	1,181
Evaluators should share technical decision-making with stakeholders.	Control	575	48.4%	482	40.6%	115	9.7%	16	1.3%	1,188
Evaluators should help train all legitimate groups to do evaluation.	Control	416	35.0%	532	44.8%	196	16.5%	43	3.6%	1,187
Evaluation should focus on bringing about individual empowerment emancipation or self-determination.	Social Justice	265	22.3%	567	47.8%	284	23.9%	71	6.0%	1,187
Evaluation should focus on bringing about social justice.	Social Justice	272	22.9%	547	46.1%	275	23.2%	93	7.8%	1,187
	Total	3,674	38.7%	3,797	40.0%	1,606	16.9%	409	4.3%	9,486

Validity and Reliability

Convergent construct validity of these questions was explored using responses to three additional statements that were also included in the survey because they were expected to negatively correlate with the eight T-PE statements. All three negatively correlated with a computed mean of the T-PE items (Table 5).

Table 5
T-PE Items' Correlation to Convergent Items (*n* = 1,157)

Statement	*r*	*p*
Only key decision-makers should participate in carrying out evaluations.	-.240	< .001
Evaluators should maintain technical decision-making of evaluation projects.	-.259	< .001
I prefer to involve stakeholders in very limited ways.	-.432	< .001

Responses to these statements were expected to negatively correlate because: a) the choice of engaging only key decision-makers in carrying out an evaluation is more aligned with the definition of practical participatory evaluation and reflects a more utilization-focused evaluation stance (Cousins & Whitmore, 1998); b) it was made clear during the statement development process that T-PE evaluators negotiate the divestment of decision-making control as capacity is built; and c)

T-PE evaluators generally have a very broad definition of stakeholder involvement and do not limit that scope *a priori.*

The internal reliability of the eight items in Table 4, measured by the coefficient alpha, was moderately strong (Cronbach's α=.736) (DeVellis, 2003). When any of the items was removed, the internal consistency was reduced to unacceptable levels (Table 6). In particular, when either of the two items that address the core philosophical strength of T-PE of involving beneficiaries and bringing about empowerment and emancipation for the beneficiaries was removed, the alpha dropped below .7. Given that this is a relatively short set of items, and that their purpose was to identify different groups of practitioners, this was sufficient to consider these internally consistent (Crano & Brewer, 2002; Streiner & Norman, 2008).

Table 6
T-PE Items' Coefficient Alphas

Statement	α if item removed
Intended program beneficiaries should participate in carrying out evaluation.	.687
People representing all important perspectives should be involved in any evaluation.	.707
I always try to involve stakeholders in my evaluations.	.715
I prefer not to take on an evaluation unless it has a strong participatory component.	.702
Evaluators should share technical decision-making with stakeholders.	.697
Evaluators should help train all legitimate groups to do evaluation.	.719
Evaluation should focus on bringing about individual empowerment emancipation or self-determination.	.727
Evaluation should focus on bringing about social justice.	.714

Group Comparisons

Comparing groups aligned with different theoretical orientations (known groups method) may also provide some construct validity by showing how well they identify a unique group of evaluators. As described earlier, if participants agreed with all eight statements, they were labeled T-PE. If they disagreed with two statements regarding social justice and empowerment but agreed with the remaining six, they were labeled P-PE. If they agreed with the three participatory statements and only some of the other five, they were labeled PE. If they did not agree with any one of the three PE statements, they were labeled non-PE.

Those who were categorized as P-PE should have chosen a utilization-focused evaluation theoretical approach more often than those in the T-PE group. To test this hypothesis, the theoretical orientation selections were dummy coded so as to compare those who selected a particular orientation across groups. Those who selected utilization-focused evaluation were coded "1" and those who selected another were coded "0." The same process was undertaken for those who selected participatory evaluation, "I do not have a preferred theoretical orientation," and "I do not know enough about these to select one."

There were no significant differences between the four groups (T-PE, P-PE, PE, non-PE) in terms of the number who selected utilization-focused evaluation over any other, or among those who selected "I do not know enough about these." There was a significant difference in the distribution across groupings for people who selected participatory evaluation (F = 34.801, df = 3, MS = 3.496, p < .001) and those who selected "I do not have a theoretical orientation" (F = 3.104, df = 3, MS = .361, p = .026) (Table 7).

Table 7
Differences in Theoretical Preference by PE Groupings

	Sum of Squares	*df*	Mean Square	*F*	*p*
Utilization-focused	.908	3	.303	1.575	.194
Participatory	10.408	3	3.469	34.801	.000
I do not have one	1.084	3	.361	3.104	.026
I do not know enough	.334	3	.111	1.355	.255

Post hoc analyses on these two significant findings (Table 8), controlling for familywise error rate using Bonferroni methods (Howell, 2002), showed significant differences for only a few comparisons. Of those who selected participatory as their preferred theoretical orientation, significantly more were categorized as T-PE than non-PE (mean diff = .2167, $p < .001$) or PE (mean diff = .1203, $p < .001$). Significantly more of those who selected "I do not have a theoretical orientation" were categorized as non-PE than T-PE (mean diff = .0662, $p = .026$). No differences surfaced between those categorized as P-PE and T-PE.

Table 8

Comparisons Between PE Groupings on Two Theoretical Preferences

Dependent Variable	(I)	(J)	Mean Difference (I-J)	Std. Error	*p*	95% CI	
						Lower Bound	Upper Bound
PE	Non-PE	PE	-.09638*	.02565	.001	-.1642	-.0286
		P-PE	-.12019	.05023	.101	-.2529	.0125
		T-PE	-.21668*	.02140	.000	-.2732	-.1601
	PE	Non-PE	.09638*	.02565	.001	.0286	.1642
		P-PE	-.02381	.05369	1.000	-.1657	.1181
		T-PE	-.12030*	.02859	.000	-.1958	-.0448
	P-PE	Non-PE	.12019	.05023	.101	-.0125	.2529
		PE	.02381	.05369	1.000	-.1181	.1657
		T-PE	-.09649	.05179	.376	-.2334	.0404
	T-PE	Non-PE	.21668*	.02140	.000	.1601	.2732
		PE	.12030*	.02859	.000	.0448	.1958
		P-PE	.09649	.05179	.376	-.0404	.2334
I do not have a theoretical orientation	Non-PE	PE	-.00524	.02772	1.000	-.0785	.0680
		P-PE	.03388	.05428	1.000	-.1096	.1773
		T-PE	.06624*	.02313	.026	.0051	.1274
	PE	Non-PE	.00524	.02772	1.000	-.0680	.0785
		P-PE	.03912	.05801	1.000	-.1142	.1924
		T-PE	.07148	.03089	.125	-.0102	.1531
	P-PE	Non-PE	-.03388	.05428	1.000	-.1773	.1096
		PE	-.03912	.05801	1.000	-.1924	.1142
		T-PE	.03236	.05596	1.000	-.1155	.1802
	T-PE	Non-PE	-.06624*	.02313	.026	-.1274	-.0051
		PE	-.07148	.03089	.125	-.1531	.0102
		P-PE	-.03236	.05596	1.000	-.1802	.1155

* The mean difference is significant at the 0.05 level.

Model Comparisons

In a later stage of the study, participants were asked to model their evaluation practices via innovative online modeling software using a set of variables expected to be important to T-PE evaluation (see Appendix A). The inclusion of a variable in a model indicated a participant's endorsement of its importance. Greater variable usage indicated greater importance, and more frequently endorsed relationships between two variables ("links") elevated the importance of that theoretical relationship and its constituent variables.

The T-PE and P-PE models produced in the modeling phase were statistically compared using differential item functioning (DIF) analysis. DIF was used in this context to test whether groups were more or less likely to include specific links (e.g., Stakeholder Engagement linked to Educate) in their models. If the members of one group tended to include a particular link in their practice models more than the members of another group, that link would then help discriminate how the two groups conceived of their practices.

The parameters for the DIF analysis were estimated using an expanded Rasch model. A Rasch model is the simplest item response theory (IRT) model, in the sense that it takes into account only two variables, fewer than other IRT models (Embretson & Reise, 2000; Gargani, 2003). The DIF analysis adds an interaction term to the two terms found in the traditional Rasch model. The statistical model is presented in Equation 1:

$$\eta DIF = \theta_j - \delta_i + \lambda_{gi} \qquad (1)$$

Here, η is the log of the odds (logit) that person *j* will include link *i* in his or her practice model; θ_j is the level of model complexity preferred by person *j*, where complexity is operationalized as the number of links in a model. Respondents with higher θ_j estimates tended to construct models of their practice that were more complex (i.e., logic models with more links). Further, δ_i is the relative difficulty of including link *i* in a model, or alternatively the relative likelihood that a link will be excluded from a model. Links with *lower* δ_i estimates tended to be included in the practice models of *more* respondents (i.e., they were less difficult and more common among modelers). And λ_{gi} is a group-by-item interaction term, where membership in any group (*g*) interacts with the difficulty of endorsing an item (*i*). Items with positive, 0, or negative λ_{gi} estimates were found more, equally, or less often, respectively, in T-PE practice models than in P-PE practice models.

To determine whether λ_{gi} was statistically different from 0 (in which case the two groups included the link in their models with the same frequency), Wald tests were performed for each estimated λ_{gi}. A Wald test is a Z test that is typically used in DIF analysis. The cutoff for statistical significance (alpha) was set to 0.05. No adjustment was made for multiple inferences.

Those links endorsed by 10% or more of the total sample (56 links) were used to create a Rasch model. The output of Equation 1 is the likelihood that a link will be included in any particular model. Tables 9 and 10 present the links that the T-PE modelers were more likely to endorse and the links that T-PE modelers were less likely to endorse, respectively, compared to P-PE modelers.

Five of the links used by T-PE modelers were not used at all by P-PE modelers: Community Trust to Stakeholder Involvement, Stakeholder Involvement to Community Trust, Stakeholder Involvement to Increase Self-Critique, Increase Social Action to Increase Social Justice, and Build Capacity to Stakeholder Involvement. Six links were more likely to be endorsed by T-PE modelers than P-PE modelers (Table 9): Diverse Perspectives to Stakeholder Involvement; Educate to Stakeholder Involvement; Stakeholder Involvement to Increase Systematic Inquiry; Stakeholder Involvement to Build Capacity; Engage Marginalized Stakeholders to Stakeholder Involvement; and Develop Questions to Multiple Method Perspective.

Table 9
Variables More Likely to be Endorsed by T-PE Modelers than P-PE Modelers

From	To	T-PE Proportion	P-PE Proportion	Coef.	Standard Error	*Z*	*P*
Community Trust	Stakeholder Involvement	0.211	0	*			
Stakeholder Involvement	Community Trust	0.204	0	*			
Stakeholder Involvement	Increase Self-Critique	0.176	0	*			
Increase Social Action	Increase Social Justice	0.148	0	*			
Build Capacity	Stakeholder Involvement	0.148	0	*			
Diverse Perspectives	Stakeholder Involvement	0.183	0.063	0.624	0.220	2.836	0.0046
Educate	Stakeholder Involvement	0.183	0.063	0.624	0.220	2.836	0.0046
Stakeholder Involvement	Increase Systematic Inquiry	0.176	0.063	0.599	0.223	2.686	0.0072
Stakeholder Involvement	Build Capacity	0.232	0.125	0.392	0.200	1.960	0.0500
Engage Marginalized Stakeholders	Stakeholder Involvement	0.162	0.063	0.545	0.230	2.370	0.0178
Develop Questions	Multiple Method Perspective	0.155	0.063	0.517	0.234	2.209	0.0271

The T-PE group was less likely than the P-PE group to endorse nine links (Table 10): Report & Disseminate to Credible Findings; Develop Questions to Collect & Analyze Data; Develop Judgments & Recommendations to Report & Disseminate; Multiple Method Perspective to Develop Questions; Shared Understanding to Stakeholder Involvement; Collect & Analyze Data to Stakeholder Involvement; Develop Judgments & Recommendations to Credible Findings; Report

& Disseminate to Stakeholder Involvement; and Credible Findings to Outcomes Change.

Table 10
Variables Less Likely to be Endorsed by T-PE Modelers than P-PE Modelers

From	To	T-PE Proportion	P-PE Proportion	Coef.	Standard Error	*Z*	*P*
Report & Disseminate	Credible Findings	0.099	0.313	-0.772	0.257	-3.004	0.0027
Develop Questions	Collect & Analyze Data	0.162	0.375	-0.623	0.216	-2.88	0.0039
Develop Judgments & Recommendations	Report & Disseminate	0.232	0.438	-0.524	0.193	-2.715	0.0066
Multiple Method Perspective	Develop Questions	0.085	0.250	-0.692	0.274	-2.526	0.0116
Shared Understanding	Stakeholder Involvement	0.099	0.250	-0.603	0.260	-2.319	0.0204
Collect & Analyze Data	Stakeholder Involvement	0.120	0.250	-0.489	0.243	-2.012	0.0442
Develop Judgments & Recommendations	Credible Findings	0.120	0.250	-0.489	0.243	-2.012	0.0442
Report & Disseminate	Stakeholder Involvement	0.120	0.250	-0.489	0.243	-2.012	0.0442
Credible Findings	Outcomes Change	0.070	0.188	-0.597	0.298	-2.003	0.0451

An alternative way of comparing the groups created above is to examine any differences in variable endorsement by the groups. Figure 1 presents the distribution of the variables by how close each group's endorsement came to its *expected* value, represented as a percentage. If each variable was evenly endorsed (i.e., they were linked equally across groups) the points would all be at 100%. This radar chart shows each of the groups and how close to equal their endorsements fell on each variable. It is sorted lowest to highest by the T-PE group's values, beginning at 12 o'clock and moving clockwise.

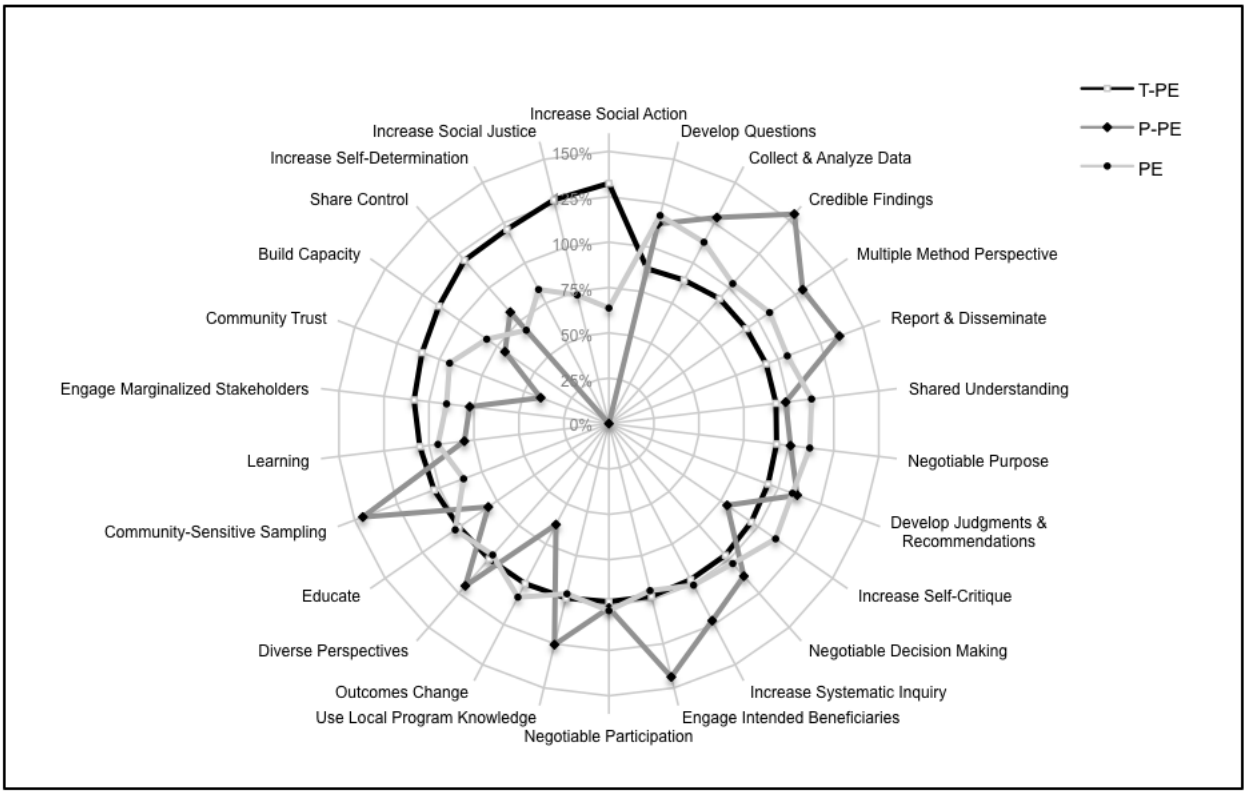

Figure 1. Principles, activities, and outcomes endorsement as a percentage of expected value

As would be expected from the philosophical importance of empowerment and social justice, Increase Social Action, Increase Social Justice, Increase Self-Determination, and Share Control, at about the 11 and 12 o'clock points of the radar chart, appear high above the expected value for the T-PE group. The first three of these do not show up in a P-PE model at all, and are therefore at zero in the radar chart. Instead, the P-PE group's endorsement of Credible Findings, Engage Intended Beneficiaries, and Community-Sensitive Sampling are higher than might be expected; Community Trust, Increase Self-Critique, and Outcomes Change are lower than might be expected.

An additional way of considering the validity of this grouping mechanism is to compare participants' responses to the practice-oriented survey questions. The P-PE evaluators were more active (43.8% worked on seven or more evaluations per year vs. 33.1% of T-PE and 31.7% of PE); reported being more experienced (43.8% had 16 or more years in evaluation vs. 21.8% of T-PE and 29.3% of PE); and considered their knowledge and experience "at an advanced level" (62.5% vs. 33.8% of T-PE and 40.2% of PE). Also, more were external evaluators (43.8% vs. 28.9% of T-PE and 40.2% of PE).

Limitations

The statements used to distinguish T-PE practitioners from other evaluators, while deductively developed in close consultation with three experienced participatory evaluation theorists, may not have been restrictive enough and, as a result, may have identified a sample that was not strictly T-PE-focused. The questions were piloted for clarity and understanding, but were not tested to determine how well they separated T-PE evaluators from other types of evaluators. The identification process was tested using other data gathered in this research, but the strength of these tests was limited by the fact that they used the data within the sample for testing. In fact, most of the participants indicated they agreed with all eight statements and the subgroups created were limited in their distinctiveness. Furthermore, the survey obtained only an 18.8% response rate, which limits the external validity of these findings.

Conclusions

Participatory evaluation has been classified as having at least two forms—one with an intention towards usefulness and practicality, and the other more inclined towards social justice and empowerment. The former benefits from the extensive use- and utilization-focused evaluation research; the latter may be actively researched under different monikers outside North America, but has decidedly less research focus in the North American evaluation literature. Individual observations (e.g., case examples of evaluations) provide some understanding of practice, and efforts are afoot to synthesize these (e.g., Cousins & Chouinard, 2012). Beyond this, however, empirical research is absent. Practitioners working with a transformative approach may look to Mertens (e.g., 2009) for philosophical guidance, but for the nuances of participatory practice with a transformative approach, there is little guidance.

The evaluation literature explains the transformative paradigm in a variety of ways. Cousins and Whitmore (1998) described something similar to Mertens' (2009) transformative evaluation (TE) but stopped short of discussing the critical engagement with power struggles that is central to the underlying theories of TE (e.g., critical theory, feminist theory, indigenous theory, critical race theory, etc.). Sabo Flores, in her dissertation (Sabo, 1999) and book (2008), discussed a participatory evaluation model focused less on the broader social justice issue of TE and more on the transformation of the individual. Likewise, other writers who have discussed youth participatory evaluation also focused on participation's value in affecting individual level change. This level seems neglected in the PE literature. In fact, this researcher struggled at the outset of the study with how "transformation" was defined in T-PE. It is the conclusion of this researcher that the transformation in T-PE first evolves from the philosophical perspective of the evaluator and that perspective directs the kinds of evaluations he or she undertakes. The nature of those evaluations is then oriented toward social justice and supports transformative axiology, epistemology, and ontology.

This research adds to the empirical knowledge of participatory evaluation by focusing on transformative participatory evaluation (Cousins & Whitmore, 1998). To that end, a panel of evaluation theorists collaborated with the researcher to develop principles central to the practice of T-PE. Eight core statements emerged, and these were used in a survey in the second phase to identify T-PE evaluators from other participatory evaluators. The statements had acceptable internal reliability but limited construct validity. Though the theorists who created them were very familiar with this form of evaluation, the discrimination strength the statements provided to distinguish T-PE evaluators and other sub-groups of PE evaluators was tenuous. More support was provided by the quantitative comparison of the resultant T-PE and P-PE models reported more extensively elsewhere (Harnar, 2012). Through Rasch modeling and differential item functioning analyses, endorsements were compared and 20 links showed significant endorsement differences between the groups. The T-PE group's links were more stakeholder- and community trust-based and the P-PE group's were more activity- and outcome-focused. This is congruent with the underlying philosophy of engagement of diverse perspectives of T-PE and therefore supports both the validity of the models and the statements' ability to discern T-PE evaluators from P-PE evaluators.

By engaging evaluators with a T-PE perspective, this research further develops our understanding of this evaluation theory by identifying those who actually prefer to practice this form of evaluation. No longer are T-PE evaluators only a theoretical type of evaluator. They were identified and are represented by this research, and the importance of their philosophical perspectives cannot be undervalued. These perspectives inform the choices they make and the evaluations they undertake. The T-PE identifying statements presented in this research should help the field better explain practice by facilitating the identification of practitioners to observe. Observations and other data collection methods can then build a stronger case for the bifurcation of PE into transformative and practical forms.

References

Alkin, M. C. (2004). Comparing evaluation points of view. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 3–11). Thousand Oaks, CA: Sage.

Brisolara, S. (1998). The history of participatory evaluation and current debates in the field. *New Directions for Evaluation*, *80*, 25–41. doi:10.1002/ev.1115

Burke, B. (1998). Evaluating for a change: Reflections on participatory methodology. *New Directions for Evaluation*, *80*, 43–56. doi:10.1002/ev.1116

Checkoway, B. N., & Gutierrez, L. M. (2006). Youth participation and community change: An introduction. In B. N. Checkoway & L. M. Gutierrez (Eds.), *Youth participation and community change* (pp. 1–9). New York, NY: The Howarth Press.

Checkoway, B., & Richards-Schuster, K. (2003). Youth participation in community evaluation research. *American Journal of Evaluation*, *24*(1), 21–33. doi:10.1177/109821400302400103

Cousins, J. B., Whitmore, E., & Shulha, L. (2013). Arguments for a common set of principles for collaborative inquiry in evaluation. *American Journal of Evaluation*, *34*(1), 7–22. doi:10.1177/1098214012464037

Cousins, J. B., & Chouinard, J. A. (2012). *Participatory evaluation up close: A review and integration of the research base*. Charlotte, NC: Information Age Press.

Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis*, *14*(4), 397. doi:10.2307/1164283

Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, *80*, 5–23. doi:10.1002/ev.1114

Crano, W. D., & Brewer, M. B. (2002). *Principles and methods of social research* (2nd ed.). Mahwah, NJ: Lawrence Earlbaum Associates.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Earlbaum Associates.

Estrella, M., & Gaventa, J. (1997). Who counts reality? Participatory monitoring and evaluation: A literature review. *International Institute for Rural Construction*, *5*, 1–70.

Fetterman, D. M. (2005). A window into the heart and soul of empowerment evaluation: Looking through the lens of empowerment evaluation principles. In D. M. Fetterman & A. Wandersman (Eds.), *Empowerment evaluation principles in practice* (pp. 1–26). New York, NY: The Guilford Press.

Fleischer, D. N., & Christie, C. A. (2009). Evaluation use: Results from a survey of U.S. American Evaluation Association members. *American Journal of Evaluation*, *30*(2), 158–175. doi:10.1177/1098214008331009

Gargani, J. (2003). *A novel quantitative method for developing and analyzing program theory*. Unpublished manuscript.

Gaventa, J. (1993). The powerful, the powerless, and the experts: Knowledge struggles in an information age. In P. Park, M. Brydon-Miller, B. Hall, & T. Jackson (Eds.), *Voices of change: Participatory research in the United States and Canada* (pp. 2–46). Toronto, Canada: OISE Press.

Greene, J. C. (1988a). Stakeholder participation and utilization in program evaluation. *Evaluation Review*, *12*(2), 91–116. doi:10.1177/0193841X8801200201

Greene, J. C. (1988b). Communication of results and utilization in participatory program evaluation. *Evaluation and Program Planning*, *11*(4), 341–351. doi:10.1016/0149-7189(88)90047-X

Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco, CA: Jossey-Bass.

Greenwood, D. J., Whyte, W. F., & Harkavy, I. (1993). Participatory action research as a process and as a goal. *Human Relations*, *46*(2), 175–192. doi:10.1177/001872679304600203

Harnar, M. A. (2012). *Theory Building Through Praxis Discourse: A Theory- And Practice-Informed Model Of Transformative Participatory Evaluation*. (Unpublished doctoral dissertation). Claremont Graduate University, Claremont, CA.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury.

King, J. A. (1998). Making sense of participatory evaluation practice. *New Directions for Evaluation*, *80*, 57–67. doi:10.1002/ev.1117

Mertens, D. M. (2009). *Transformative research and evaluation*. New York, NY: The Guilford Press.

Meyer, L. H., Park, H., Grenot-Scheyer, M., Schwartz, I., & Harry, B. (1998). Participatory research: New approaches to the research to practice dilemma. *The Journal of the Association for Persons with Severe Handicaps*, *23*(3), 165–177.

Miller, R. L. (2010). Developing standards for empirical examinations of evaluation theory. *American Journal of Evaluation*, *31*(3), 390–399. doi:10.1177/1098214010371819

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

Preskill, H., & Caracelli, V. (1997). Current and developing conceptions of use: Evaluation Use TIG Survey Results. *American Journal of Evaluation*, *18*(1), 209–225. doi:10.1177/109821409701800122

Sabo Flores, K. (2008). *Youth participatory evaluation: Strategies for engaging young people*. San Francisco, CA: Jossey-Bass.

Sabo, K. (1999). *Young people's involvement in evaluating the programs that serve them*. (Unpublished doctoral dissertation). City University of New York.

Shadish, W. R. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, *19*(1), 1–19. doi:10.1177/109821409801900102

Streiner, D. L., & Norman, G. R. (2008). *Health measurement scales: A practical guide to their development and use* (4th ed.). New York, NY: Oxford University Press.

Suarez-Balcazar, Y., & Harper, G. W. (2003). Community-based approaches to empowerment and participatory evaluation. *Journal of Prevention & Intervention in the Community*, *25*(3), 1–4.

Whitmore, E. (1988). *Participatory evaluation approaches: Side effects and empowerment*. (Unpublished doctoral dissertation). Cornell University, Ithaca, NY.

Appendix A

Principle	**Definition**
Community Trust	Evaluator works to build lasting trust by developing working relationships with a broad range of stakeholders.
Negotiable Purpose	The purpose of the evaluation is negotiated with stakeholders.
Multiple Method Perspective	Evaluator applies multiple methods as appropriate to the evaluation context.
Diverse Perspectives	Evaluator ensures representation of diverse perspectives by including concerns, values, and interests of stakeholders.
Negotiable Decision Making	Technical decision-making (e.g., survey instrument selection, statistical analyses, data presentation) is negotiated with stakeholders.
Negotiable Participation	Scope of stakeholder participation in evaluation is not decided ahead of time. Barriers to and supports necessary for participation are identified and negotiated.
Community Sensitive Sampling	Sampling procedures account for community diversity.
Engage Marginalized Stakeholders	Evaluator engages marginalized program stakeholders (e.g., those who might otherwise lack representation) in meaningful participation.
Engage Intended Beneficiaries	Evaluator engages intended program beneficiaries in meaningful participation.
Activity	**Definition**
Build Capacity	Evaluator trains stakeholders in the necessary skills to participate in the evaluation.
Share Control	Evaluator negotiates the giving of control of the evaluation to program stakeholders.
Educate	Evaluator educates stakeholders on the value of evaluation.
Use Local Program Knowledge	Evaluation decisions are made using local program knowledge.
Develop Questions	Evaluator collaborates with stakeholders in defining evaluation purpose and evaluation questions.
Collect & Analyze Data	Evaluator collaborates with stakeholders in data collection and analysis.
Develop Judgments & Recommendations	Evaluator collaborates with stakeholders in interpreting findings, and formulating judgments and recommendations from the data.
Report & Disseminate	Evaluator collaborates with stakeholders in reporting and disseminating the findings.
Outcome	**Definition**
Shared Understanding	All participants develop shared understanding of program functions and processes.
Learning	All participants learn new skills.
Credible Findings	Participants see evaluation findings as credible.
Increase Systematic Inquiry	Increase capacity for participants to engage in and use systematic inquiry.
Increase Self-Critique	Increase participants' capacity for self-critique.
Increase Self-Determination	Increase individual self-determination, emancipation and empowerment.
Increase Social Justice	Enhance social justice.
Increase Social Action	Increase social action.
Outcomes Change	Program outcome expectations change as a result of the process.