Journal of MultiDisciplinary Evaluation
Volume 15, Issue 33, 2019

JMDE
Journal of MultiDisciplinary Evaluation

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

What, How, and Why?
A Comparative Analysis of 12 Goal-Free Evaluations

Brandon W. Youker
Grand Valley State University

Background: Goal-free evaluation (GFE) is any evaluation in which the evaluator conducts the evaluation without reference to predetermined goals or objectives.

Purpose: The purpose of this study is to examine GFE in actual practice focusing on what GFE is, how it is conducted, and why the evaluators use it.

Setting: Not applicable.

Intervention: Not applicable.

Research Design: Document analysis.

Data Collection and Analysis: The researcher collected data from a non-random sample of 12 cases of GFE found in published and non-published papers, reports, and guidebooks. The researcher analyzed the documents using quantitative and qualitative content analysis.

Findings: The findings suggest that goal-free evaluators consider GFE an outcome evaluation that supplements GBE. These goal-free evaluators typically used an *ex post facto* evaluation design, non-random sampling of stakeholders, and semi-structured interviewing to collect data. The evaluators described using GFE to improve the evaluand, to find side-effects, and to evaluate highly complex evaluands.

Keywords: *goal-free evaluation; goal-based evaluation; comparison; outcomes; goals; side-effects.*

Background

Evaluation scholars have long recognized that defining the problem (Suchman, 1967) and specifying goals and objectives (Swisher & Abrams, 1973) are *sine qua non* for program administrators in establishing a sound program monitoring and evaluation strategy. Yet, whose goals and objectives matter? The program administrators? Staff? Funders? Designers? Consumers? What should the evaluator do if the stakeholders disagree on the goals? What if the stakeholders hastily devised or poorly stated the goals? Moreover, what do evaluators do when the exact problems or intended outcomes are not well defined or even known? Rather than becoming inhibited with these questions, goal-free evaluators conduct the evaluation without referring to the evaluand's intentions—its stated goals and objectives. Instead, goal-free evaluators investigate what the program did, or is doing—its actualized outcomes rather than its intended outcomes (Scriven, 1991).

Roughly, half a century ago, Michael Scriven (1972) proposed goal-free evaluation (GFE) yet the literature on GFE is sparse and sporadic. There are several influential evaluation books that introduce GFE (e.g., Alkin, 2004; Fitzpatrick, Sanders, & Worthen, 2004; Mertens & Wilson, 2012; Patton, 1997; Shadish, Cook, & Leviton, 1991; Worthen, 1990) and articles that present its theoretical pros and cons (e.g., Alkin, 1974; Irvine, 1979; Salasin, 1974; Scriven, 1973, 1974, 1991; Stufflebeam, 1974). Other scholars describe applying GFE to a particular field or practice area such as corporate job training (James & Roffe, 2000), curriculum development (Pipia, 2014), information systems (Cronholm & Goldkuhl, 2003), knowledge management (Hellström & Jacob, 2003), and agricultural extension (Manfredi, 2003), among others. Yet there is only a handful of scholars who offer general GFE principles or methodological guidance of substance (e.g., House, 1991; Scriven, 1974, 1991; Thiagarajan, 1975; Welch, 1976; Youker, 2013). Three decades later, Scriven himself considered GFE his most misunderstood contribution to evaluation (The Oral History Project Team, 2005). Consequently, throughout the years, scholars (e.g., Patton, 1997; Shadish et al., 1991; Smith, 2009; Welch, 1978) have criticized GFE for lacking methods by which to conduct it. In the 1970s, Welch (1978) called for "improved methodology" stating, "I believe additional applications of goal-free evaluation are needed to refine our methods and give the goal-free approach a fair and adequate test," (p. 12). Today, "additional applications" have come to fruition via published evaluation reports and scholarly papers. This study examines a non-random sample of applied GFE and asks the following three questions:

1. What is GFE?
2. How does one conduct a GFE?
3. Why use a GFE?

To answer these questions, this article consists of a comparative analysis examining a dozen cases of GFE implementation from the scholarly literature, evaluation guidebooks, and technical reports; this analysis focuses on GFE design, sampling methods, data collection methods, and purpose.

Methods

The researcher used qualitative and quantitative descriptive analyses to identify themes and perspectives present within GFE-related documents. The sample consists of 12 purposively sampled cases of GFE. The primary inclusion criterion was that there was enough available material on each case to answer all three research questions. Excluded from the analysis are papers where the evaluators conducted the GFE for non-evaluative purposes. Examples of excluded cases include those who use GFE as a criteria or values defining instrument (e.g., Moro, Cassibba, & Costantini, 2007), cases where the evaluators only passingly alluded to GFE (e.g., Bry & George, 1979), and those that provide too little information for adequate analysis (e.g., Whiting & Roberts, 2016). Thus, the GFEs included in this study fail to represent all published cases, let alone all GFEs ever conducted; nevertheless, these samples provide valuable insight regarding past GFE practice.

The researcher modeled the coding framework after the approach employed by Youker, Ingraham, and Bayer (2014) and

Youker, Ford, and Bayer (2017) to map GFE practice using evaluation documents as the source. The researcher reviewed the published and non-published materials on GFE identifying key components of GFE for content analysis. From here, the researcher developed content categories and then employed an inductive approach (i.e., open coding) to determine which categories were applicable to GFE methods and practice. This involved reading the materials and noting the environment and context surrounding the GFE, and the goal-free evaluators' reported methodology and their stated opinions in relation to the key research questions. To enhance the content validity of the coding categories, the researcher established coding rules for translating the segments of text into codes. For example, the researcher developed contextual translation rules for combining comments about identifying side-effects and unexpected, unknown, and unanticipated outcomes under one code, while considering statements related to identifying actual outcomes as distinct from the mentions of unanticipated outcomes. After identifying the documents for review, the researcher applied the coding framework across all GFE materials coding for the general evaluation context, what GFE is, how to conduct it, and why do it.

Table 1 presents each GFE case, the main textual sources and type of texts used in the analysis along with a brief description of the evaluand. The evaluands ranged from education, training, and human service programs to international development initiatives. Four cases, i.e., Most Significant Change (MSC), Outcome Harvesting (OH), Participatory Assessment of Development (PADev), and Qualitative Impact Protocol (QUIP), are fairly established in international development evaluation and possess their own evaluation guidebooks and publicly available evaluation reports; furthermore, they are the subject of several scholarly publications. Consequently, instead of referencing every evaluation publication or report as a case, this analysis treats each of these models as an individual GFE case.

Table 1
GFE in Practice

Case	Document Source	Document Type	Evaluand
History Test	Mueller & Colley (2015)	Journal article	City high school history department's new history testing system
Respite	Perry & Bontinen (2001)	Journal article	8-bed overnight respite program for person with Alzheimer's and dementia
OT Materials	Welch (1978)	Published paper presentation	Undergraduate occupational therapy program's materials and documents
Camp Training	Youker, Zielinki, Hunter, & Bayer (2016)	Journal article	Training of counselors at outdoor camp for individuals with disabilities
MIW	Berkshire, Kouame, & Richardson (2009)	Technical evaluation report	City's welfare and homelessness prevention program
MSSEP	Youker (2005)	Technical evaluation report	School district's summer school program for middle school students
Ke Aka Ho'ona	Matsunaga & Enos (1997) Stufflebeam, Gullickson, & Wingate (2002)	Technical evaluation report Technical evaluation report	Hawaiian community self-help housing program
Colleges	Evers (1980)	Doctoral dissertation	Colleges' efficiency/cost reduction initiative
MSC	Davies & Dart (2003, 2005) Lennie (2011)	Journal article; Guidebook Guidebook	Multiple in international development

Case	Document Source	Document Type	Evaluand
OH	Wilson-Grau & Britt (2012) World Bank (2014)	Guidebook World Bank publication	Multiple in international development
PADev	Dietz et al. (2013) Dietz, van der Geest, & Obeng (2013)	Guidebook Book chapter	Multiple in international development
QUIP	Copestake (2014) Copestake & Remnant (2015)	Guidebook Non-scholarly brief	Multiple in international development

Findings

Table 2 presents the evaluators' definitions for GFE. The first column is the GFE case, the second column is the author of the definition, the third column is the reference that inspired the author's definition, and the fourth column is the quoted definition. The researcher excluded four GFE cases as these evaluators did not recognize that their evaluation models were goal-free and therefore failed to provide a definition of GFE. Of the eight definitions, six included phrasing related to disregarding goals or intentions or serving as a counter to goal-based evaluation (GBE); and four definitions mentioned searching for actual effects and outcomes or searching for side-effects.

Table 2
GFE Definitions

Case	Author	Reference	GFE Definition
History Test	Mueller & Colley (2015)	Scriven (1991)	A goal free evaluation does not focus on the degree to which the intended outcome of a program was achieved but rather looks to determine the unintended side effects (both positive and negative) of a program. (p. 10698)
Respite	Perry & Bontinen (2001)	Patton (1990)	Goal free evaluation, in its search for actual effects, is an inductive and holistic strategy aimed at countering the logical deductive limitations inherent in the usual quantitative goal based approach to evaluation. (Patton, 1990, p. 116) (Perry & Bontinen, pp. 86-86)
OT Materials	Welch (1978)	Scriven (1972)	The goal-free approach is an experimental venture which capitalizes upon objectivity, reduces personal bias, and increases potential for detecting side-effects. (p.4)
Camp Training	Youker, Zielinski, Hunter, & Bayer (2016)	Scriven (1991)	Goal-free evaluation is any evaluation in which the evaluator lacks the knowledge of or simply disregards the evaluand's stated goals and objectives. Rather the goal-free evaluator investigates the evaluand's actual outcomes—past and present—not its stated intentions. (p.28)
MIW	Berkshire, Kouame, & Richardson (2009)	Scriven (1991, 2007)	In conducting a goal-free evaluation, rather than measuring the success of the program in meeting a pre-specified set of target goals, evaluators attempt to learn about the program and its results without being aware of specific objectives of the program. The evaluation focused on how successful the program was in achieving positive client outcomes that could be directly attributable to client participation in the program. (p. 5)

Case	Author	Reference	GFE Definition
MSSEP	Youker (2005)	Scriven (1991)	Goal-free evaluation: An evaluation in which the evaluation team deliberately avoids learning what the goals are or were so as to avoid being overly focused on intended outcomes. The rationale behind this approach is that both intended and unintended effects are important to include in an evaluation. (p. 17)
Ke Aka Ho'ona	Matsunaga & Enos (1997)	None provided	This is an evaluation process by which the observer identifies and assesses the full range of program impacts without being constrained by knowledge of program goals and design. (p. 1)
Colleges	Evers (1980)	Scriven (1972)	To do goal-free evaluation, one is not being asked to do fact-free or information-free evaluation. Information, both descriptive and judgmental, is as necessary to operationalize summative, goal-free evaluation as any other approach one might take. What one is specifically to guard against is information that each specific project poses as intended goals. (p. 86)

Table 3 displays the stated or implied GFE design elements. All of the evaluations were outcome evaluations as the primary evaluation task was to identify changes or impacts caused by the intervention. Eleven out of 12 GFEs were formative evaluations conducted for program improvement purposes. Three evaluations were both formative and summative while only the Colleges evaluators used GFE solely for summative reasons (i.e., accountability and decision-making). All but one was *ex post facto* meaning the evaluators conducted their GFEs during or after the intervention and therefore did not employ any form of experimental design. Half of the cases explicitly referred to instituting goal screening or blinding processes to shield the evaluators from program goals and objectives. Five of the evaluations stated or implied using a goal-dismissive form of GFE where the evaluators did not take particular precautions to avoid the goals; instead, they simply ignored the goals by not asking about or referring to them. All but one of the GFEs supplemented GBE as they were a component, instrument, or phase of a much larger goal-based monitoring and/or evaluation strategy.

Table 3
GFE Designs

	Outcome	Formative	Summative	Ex post facto	Goal blinding/ screener	Goal-dismissive	Supplement GBE
History Test	X	X		X		X	X
Respite	X	X		X			
OT Materials	X	X		X	X		X
Camp Training	X	X			X		X
MIW	X	X		X	X		X
MSSEP	X	X		X	X		X
Ke Aka Ho'ona	X	X		X	X		X
Colleges	X		X	X	X		X
MSC	X	X		X		X	X
OH	X	X	X	X		X	X
PADev	X	X	X	X		X	X
QUIP	X	X	X	X		X	X
Totals	12	11	4	11	6	5	11

The evaluators provided their purposes for selecting GFE. Table 4 shows that five of the 12 cases reported employing GFE to identify unanticipated outcomes or for situations when the outcomes were poorly defined or unknown, and five reported using their GFEs for examining complex evaluands. Three cases mentioned broadening perspectives or potential outcomes and three identified reducing potential for bias as reasons for GFE. Other GFE purposes included identifying or assessing consumer or beneficiary needs, identifying or determining evaluation criteria or stakeholder values, identifying the intervention's actual outcomes, providing context for understanding the consumer, and understanding the specific mechanisms and processes underlying change.

Table 4
GFE Purposes

	Ke Aka Ho'ona	MSSEP	MIW	Camp Training	OT Materials	Respite	History Test	Totals
					X	X	X	5
			X				X	5
	X			X				3
					X			3
		X	X					2
				X				2
		X					X	2
								1
								1

	QUIP	PADev	OH	MSC	Colleges
Identifying unexpected/ unknown outcomes			X	X	
Examining complex evaluands	X		X	X	
Broadening perspectives/outcomes				X	
Preventing bias	X				X
Identifying/ assessing needs					
Determining criteria/ values				X	
Identifying actual outcomes					
Providing context		X			
Understanding change			X		

The goal-free evaluators described their sampling methods as displayed in Table 5. In four of the cases, the evaluators chose not to sample but reported using a census; in each of these cases, there were relatively discrete and identifiable impactee groups and therefore the evaluators attempted to collect data from all impactees. Nine of the GFEs used non-random sampling methods to select participants or data sources; PADev was the only to employ stratified non-random sampling to identify diverse samples of community members from which to collect data. The most frequently reported sampling method was purposive sampling, which targeted key stakeholders, and impactees particularly intended beneficiaries. Of the eight GFEs that use purposive sampling, three specified the type of purposive sampling. Perry and Bontinen (2001) used diversity sampling for History Test; Berkshire, Kouame, and Richardson (2009) report using modal instance sampling with Making It Work (MIW); and Lennie (2011) used outlier sampling with MSC. Two GFEs used random sampling; PADev reported utilizing simple random sampling while QUIP utilized stratified random sampling.

Table 5
GFE Sampling Methods

	Non-random sampling			Random sampling	
	Census	Purposive	Stratified non-random	Simple random	Stratified random
History Test	X				
Respite		X			
OT Materials		X			
Camp Training	X	X			
MIW		X			
MSSEP	X				
Ke Aka Ho'ona		X			
Colleges		X			
MSC		X			
OH		X			
PADev	X		X	X	
QUIP					X
Totals	4	8	1	1	1

The GFEs employed nine distinct methods of data collection (see Table 6). The mean number of data collection methods used per GFE case was 3.33 (*SD* = 1.07). Camp Training reported incorporating the greatest number of data collection methods with six, while Colleges and PADev employed the fewest with two. Ten out of 12 evaluations incorporated semi-structured interviews typically with the program consumers but on occasion with program staff or the broader community as well. Two-thirds used direct observation or participant observation as a data collection method while over half reported reviewing program or consumer documents or archives. Half of the GFEs employed focus groups usually with program consumers or other key stakeholders. Three used surveys or questionnaires and three used a checklist approach where evaluators marked off benchmarks as being complete or incomplete. Unique to their GFEs, a single GFE case reported using a case study approach (i.e., History Test), another was the only to apply a pretest-posttest (i.e., Camp Training), and PADev was the sole GFE to incorporate collaborative ranking exercises which the evaluators used as a way to determine the significance of the outcomes.

Table 6
GFE Data Collection Methods

	Semi-structured interviews	Direct or participant observation	Document/ archival analysis	Focus groups	Surveys	Checklists	Case studies	Pretest-posttests	Collaborative ranking tasks	Totals
History Test	X	X					X			3
Respite	X	X		X						3
OT Materials			X	X		X				3
Camp Training	X	X		X	X	X		X		6
MIW	X	X	X							3
MSSEP	X	X	X							3

Ke Aka Ho'ona	X		X			X				3
Colleges	X		X							2
MSC	X	X	X	X						4
OH	X	X	X		X					4
PADev				X					X	2
QUIP	X	X		X	X					4
Totals	10	8	7	6	3	3	1	1	1	

Discussion

The GFE cases presented represent what goal-free evaluators report about their GFEs focusing on the following questions: (1) What is GFE? (2) How does one conduct a GFE? (3) Why use a GFE? The findings from this analysis lead to the following conclusions summarized in Table 7.

Table 7
Study Conclusions

What is GFE?	How does one conduct a GFE?	Why use a GFE?
GFE is an outcome evaluation	Goal-free evaluators typically use an ex post facto design	GFE is used for formative evaluation
GFE is a supplement to GBE	Goal-free evaluators typically use non-random sampling	GFE is used to uncover side-effects
	Goal-free evaluators typically use semi-structured interviewing	GFE is used for evaluating complex evaluands

What is GFE?

A firm conclusion drawn from comparing Tables 2, 3, and 4 is that GFE is inextricably an outcome evaluation. All of the evaluators designed their GFEs to identify and gather data on the outcomes or the effects of the intervention. This was unsurprising as Scriven originally conceptualized GFE to be a way of examining program outcomes without the cueing as to which outcomes to search for. The goal-free evaluator looks at "what the program is actually doing" (Scriven, 1991, p. 180) rather than what it says it should be doing.

GFE is a supplement to GBE. The overwhelming majority of these evaluators agreed that it is best to view GFE as a complementary approach applied in conjunction with GBE. None advocate for GFE as a standalone evaluation; rather they solidly support GFE as part of a larger evaluation strategy. For example, the Occupational Therapy (OT) Materials, Camp Training, MIW, Middle School Summer Enrichment Program (MSSEP), and Colleges evaluators reported simultaneous GBE occurring independently of their GFE, while MSC and QUIP specified that there were previous GBEs that their GFE supplemented.

How does one conduct a GFE?

All but one of the GFEs was *ex post facto* in that the evaluator conducted the GFE during or after the independent variable (i.e., the evaluand, program, or intervention) and hence, the evaluator did not control or manipulate the independent variable. There is an inherent selection threat to internal validity with *ex post facto* designs, because—without random assignment to treatment or control groups—it is impossible to make definitive claims as to whether there is a causal relationship between the independent variable

and the outcome. Nevertheless, given contextual conditions, often evaluators have little choice in adopting *ex post facto* designs for demonstrating correlation or the possibility of a cause and effect relationship (Chatterji, 2007).

Table 5 shows that two-thirds of the goal-free evaluators non-randomly sampled individuals from whom to seek evaluation data while one-third used a census to collect data from all participants or intended beneficiaries. It is probable that the choice of non-probability sampling methods for several of the evaluators was a function of the size and scope of the evaluand and evaluation or of environmental conditions rendering random sampling infeasible or unnecessary. Regardless of sampling procedure, evaluators typically collected data from key stakeholders namely program participants, their family members, and program or project staff.

All of the goal-free evaluators used multiple methods to collect data yet 10 of 12 specifically used semi-structured interviewing as their primary data collection method. The goal-free evaluators interviewed program participants and/or key stakeholders asking about outcomes and changes seemingly attributable to the intervention. Evaluators used other common qualitative data collection methods such as observation, document analysis, and focus groups. Despite the fact that the goal-free evaluators in these cases relied heavily on qualitative data, Scriven (1991) maintains that GFE is combinative, as evaluators can combine it in full or in part with other evaluation methods such as "qualitative versus quantitative, survey versus experiment, multiple perspectives versus one right answer, etc." (p. 182). As an example, the Camp Training goal-free evaluators used a quantitative pretest-posttest to assess the participants' changes in knowledge. Scriven's point is that there is no reason to pigeonhole GFE by associating it exclusively with qualitative data.

Why use a GFE?

Table 3 shows that majority of evaluators in the sample use GFE for formative purposes with the occasional evaluator employing GFE for both formative and summative reasons. According to Scriven (1974, 1991), executives and administrators perceive GFE as threatening because the GFE approach forces them to relinquish control over the direction of the outcomes under examination and consequently much of the direction of the evaluation inquiry as a whole. If so, it is plausible that the evaluation client who agrees to use GFE does so with the genuine intention of critically observing and improving the evaluand instead of judging it.

Evaluators and evaluation clients use GFE for its ability to uncover side-effects which are any outcomes that occur unrelated to the goals, according to Scriven (2007). For example, a potential side-effect of involvement in the peer substance abuse support group, Alcoholic Anonymous (AA), is connecting the group participant with another participant who helps him land a job; although, another potential side-effect of AA is connecting the group participant with a different participant who becomes his new best drinking buddy. Both outcomes are unintentional, possible, and likely attributable to participation in AA, therefore side-effects. In fact, decades after introducing GFE, Scriven (1991) wrote that he "knows of no pure GFE that has failed to uncover new and substantial side-effects after a program has already been evaluated in a goal-based mode" (p.180). He asserts that for program people, GFE is entirely in their interest because they may find side-effects that are spoiling the program and are rectifiable, or discover side-effects that are positive and worthy of examination, support, and applause. Whether future research will conclude that GFE always provides value in terms of revealing side-effects as Scriven suggests, GFE's perceived ability to identify these unanticipated outcomes is what draws many to the goal-free approach (see Table 4). Several evaluators reported that because their evaluands were subject to previous traditional GBEs (see Table 3), they therefore had more opportunity and freedom to explore broadly for outcomes, beyond those previously stated, intended, expected, or known.

Evaluators and evaluation clients utilize GFE for evaluating complex evaluands (see Table 4) because of its adaptability to unspecified or changing outcomes. For instance, The United States Agency for International Development (2013) endorses

GFE as a method of "complexity-aware monitoring" (p. 14) for humanitarian aid projects. There are other evaluators who champion GFE for its ability to deal with complexity in international development projects (Davies & Dart, 2005; Rogers, 2011; Wilson-Grau & Britt, 2012), natural disaster relief efforts (Belanger, 2006), and international peacebuilding projects (Brusset, de Coning, & Hughes, 2016). Belanger (2006) summarizes the rationale as follows. Traditionally the goals and objectives of the program drive everything and because they have to be measurable, the program administrators must articulate them well ahead of time. This is problematic for complex and multifaceted interventions especially when there are numerous interaction effects, the outcomes are slow to manifest, and the sources for the effects can be difficult to discern even after several years. Brusset, de Coning, and Hughes (2016) espouse GFE for evaluating peacebuilding initiatives basing their justification "on the assumption that there are many entry points in seeking to affect a conflict, and that they cannot be defined in advance," (p. 13).

Conclusion

Guidance for conducting a GFE is scant; therefore, this investigation examined a dozen completed GFEs to identify how the evaluators conceptualized GFE, to describe how they conducted their GFE, and to infer why they chose GFE. The findings from this inquiry are descriptive rather than prescriptive, continuing to shift the dialog from what goal-free evaluators proclaim one ought to do to what the evaluators really did. Historically, in summarizing GFE, one might describe gathering data on the evaluand's outcomes while avoiding or blinded from the evaluand's stated goals. With the inclusion of these dozen cases, GFE expands to include the following. GFE is an outcome evaluation that supplements GBE. GFEs often use an *ex post facto* evaluation design, non-random sampling of stakeholders, and semi-structured interviewing for collecting data. Lastly, evaluators report that they have used GFE for its ability to improve the evaluand, to find side-effects, and to evaluate highly complex evaluands.

After 50 years, GFE maintains its relevance and it warrants further study; yet in the immediate future, widespread adoption of and empirical studies on GFE seem unlikely. Until that day, examining past GFE practice and learning from the experiences of those who used GFE will continue to be essential for studying the outcomes and consequences of the GFEs themselves. Through reviewing GFE-related documents, reports, and cases, evaluation scholars might more accurately describe how evaluators conduct GFE, further the operationalization of GFE, learn more about its strengths and limitations, and extrapolate regarding its suitability in diverse contexts and under varying conditions. Therefore, this is a call for evaluators to conduct GFEs and to publish their methods, instruments, findings, context and considerations, etc. and share their technical reports. Ernest House was possibly the very first to try a GFE after Scriven proposed it; and decades later, he still recommends that all evaluators give GFE a go:

> Here's my main suggestion for you [evaluators]. Go out and try to do one of these on a complex program. You should go out—go out and try to do it. You will then have insights you cannot possibly have without attempting to do one (E. House, personal communication, November 6, 2014).

References

Alkin, M. C. (1974). Comments by Marvin C. Alkin: Wider context goals and goal-based evaluators. In W. J. Popham (Ed.), *Evaluation in education: Current applications* (pp. 52-55). Berkeley, CA: McCutchan.

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

Belanger, K. H. (2006, November). *Goal-based vs. goal-free evaluation in crises: Lessons from 2 hurricanes.* Paper presented at the meeting of the American Evaluation Association, Portland, OR.

Berkshire, S., Kouame, J., & Richardson, K. K. (2009). *Making It Work: Evaluation report.*

Kalamazoo, MI: Western Michigan University, The Evaluation Center.

Babbie, E. (1999). *The basics of social research.* Belmont, CA: Wadsworth Publishers.

Brusset, E., de Coning, C., & Hughes, B. (Eds.). (2016). *Complexity thinking for peacebuilding practice and evaluation*. London, UK: Palgrave Macmillan.

Bry, B. H. & George, F. E. (1979). Evaluating and improving prevention programs: A strategy from drug abuse. *Evaluation and Program Planning, 2*, 127-136.

Chatterji, M. (2007). Grades of evidence: Variability in quality of findings in effectiveness studies of complex field interventions. *American Journal of Evaluation, 28*(3), 239-255.

Copestake, J. G. (2014). *Full guidelines for the Qualitative Impact Protocol (QUIP)* (Working Paper). Retrieved from Centre for Development Studies, University of Bath: http://opus.bath.ac.uk/36678/5/QUIP_Guidelines_Oct_14.pdf

Copestake, J. & Remnant, F. (2015). *The qualitative impact assessment protocol (QUIP)* (Briefing Paper). Retrieved from University of Bath Centre for Development Studies website: http://www.bath.ac.uk/cds/projects-activities/qualitative-impact-assessment-protocol-QUIPS11/index.html

Cronholm, S. & Goldkuhl, G. (2003, September). *Six generic types of information systems evaluation.* Paper accepted to the 10th European Conference on Information Technology Evaluation, Madrid, Spain.

Davies, R. & Dart, J. (2003). A dialogical, story-based evaluation tool: The Most Significant Change technique. *American Journal of Evaluation, 24*(2), 127-155.

Davies, R. & Dart, J. (2005). *The 'Most Significant Change' (MSC) technique: A guide to its use*. Retrieved from http://www.mande.co.uk/docs/MSCGuide.pdf

Dietz, T. et al. (2013). *PADev guidebook: Participatory assessment of development*. Leiden: KIT Publishers.

Dietz, T., van der Geest, K. & Obeng, F. (2013). Local perceptions of development and change in Northern Ghana. In J. Yaro (Ed.), *Rural development in Northern Ghana* (pp. 17-36). New York: Nova Science Publishers.

Evers, J. W. (1980). *A field study of goal-based and goal-free evaluation techniques*. Unpublished doctoral dissertation, Western Michigan University, Kalamazoo.

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

Hellström, T. & Jacob, M. (2003). Knowledge without goals? Evaluation of knowledge management programs. *Evaluation, 9*(1), 55-72.

Henry, G. T. & Mark, M. M. (2003). Toward an agenda for research on evaluation. *New Directions for Evaluation*, *97*, 69-80.

House, E. (1991). Confessions of a responsive, goal-free evaluator. *Evaluation Practice, 12*(1), 109-114.

Irving, J. F. (1979). Goal-free evaluation: Philosophical and ethical aspects of Michael Scriven's model. *CEDR quarterly, 12*(3), 11-14.

James, C. & Roffe, I. (2000). The evaluation of goal and goal-free training innovation. *Journal of European Industrial Training, 24*(1), 12-20.

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd ed.). Thousand Oaks, CA: Sage.

Lennie, J. (2011). The Most Significant Change technique: A manual of M&E staff and others at Equal Access. Washington, D.C.: USAID.

Manfredi, T. C. (2003). Goal based or goal free evaluation? Growing New Farmers website. http://www.smallfarm.org/uploads/uploads/Files/Goal_Based_or_Goal_Free_Evaluation.pdf

Matsunaga, D. S. & Enos, R. (1997). *Goal-free evaluation of the Alger Foundation's Wai'anae Self-help Housing Project: Evaluation report.* Kalamazoo, MI: Western Michigan University, The Evaluation Center.

Mertens, D. M. & Wilson, A. T. (2012). *Program evaluation theory and practice.* New York, NY: The Guilford Press.

Moro, G., Cassibba, R., & Costantini, A. (2007). Focus groups as an instrument to define evaluation criteria. *Evaluation, 13*(3), 340-357.

Mueller, G. W. & Colley, L. M. (2015). An evaluation of the impact of end-of-course exams and ACT-QualityCore on U.S. history instruction in a Kentucky high school. *Journal of Social Studies Research, 39*, 95-106.

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

Perry, J. & Bontinen, K. (2001). Evaluation of a weekend respite program of persons with Alzheimer disease. *Canadian Journal of Nursing Research, 33*(1), 81-85.

Pipia, E. (2014). The effectiveness of goal-free evaluation in curriculum development and in quality assurance process. *Journal of Education, 3*(2), 29-31.

Rogers, P. (2011). Implications of complicated and complex characteristics for key tasks in evaluation. In K. Forss, M. Marra, and R. Schwartz (Eds.), *Evaluating the complex: Attribution, contribution, and beyond.* New Brunswick, New Jersey: Transaction Publishers.

Salasin, S. (1974). Exploring goal-free evaluation: An interview with Michael Scriven. *Evaluation, 2*(1), 9-16.

Scriven, M. (1972). Prose and cons about goal-free evaluation. *Evaluation Comment, 3*(4), 1-4.

Scriven, M. (1973). Goal-free evaluation. In E. R. House (Ed.), *School evaluation: The politics and process* (pp. 319-328). Berkeley, CA: McCutchan.

Scriven, M. (1974). Prose and cons about goal-free evaluation. In W. J. Popham (Ed.), *Evaluation in education: Current applications* (pp. 34-67). Berkeley, CA: McCutchan.

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

Scriven, M. (2007, February). Key evaluation checklist. Retrieved from https://wmich.edu/sites/default/files/attachments/u350/2014/key%20evaluation%20checklist.pdf

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

Smith, C. (2009). The Scottish enlightenment, unintended consequences and the science of man. *The Journal of Scottish Philosophy, 7*(1), 9–28.

Stufflebeam, D. L. (1974). Comment by Daniel L. Stufflebeam: Should or can evaluation be goal-free? In W. J. Popham (Ed.), *Evaluation in education: Current applications* (pp. 43-46). Berkeley, CA: McCutchan.

Stufflebeam, D. L., Gullickson, A., & Wingate, L. (2002). *The spirit of Consuelo: An evaluation of Ke Aka Ho'ona.* Retrieved from Western Michigan University Evaluation Center website: www.wmiche.du/evalctr

Suchman, E. A. (1967). *Evaluative research: Principles and practice in public service and social action programs.* New York: Russell Sage Foundation.

Swisher, J. D. & Abrams, A. (1973). Specifying objectives. In L. A. Abrams, E. F. Garfield, & J. D. Swisher (Eds.), *Accountability in drug education.* Washington, D. C.: Drug Abuse Council.

The Oral History Project Team, O. H. P. T., & Miller, R. L. (2005). The oral history of evaluation, part 3: The professional evolution of Michael Scriven. *American Journal of Evaluation*, *26*(3), 378-388.

Thiagarajan, S. (1975). Goal-free evaluation of media. *Educational Technology, 15*(5), 38-40.

United States Agency for International Development. (2013). *Discussion note: Complexity-aware monitoring*. Retrieved from: http://usaidprojectstarter.org/sites/default/files/resources/pdfs/Complexity-Aware-Monitoring-2013-12-11-FINAL.pdf

Welch, W. W. "Goal Free Evaluation Report for St. Mary's Junior College" (Unpublished Report) Minneapolis: 1976.

Welch, W. W. (1978, March). *Goal-free formative evaluation-an example*. Paper presented at the meeting of the American Education Research Association, Toronto, Canada.

Whiting, L. S. & Roberts, S. A. (2016). "We want to influence": An evaluation of the nation health service England youth forum [Abstract]. *Archives of Disease in Childhood, 101*, A365-A366.

Wilson-Grau, R. & Britt, H. (2012). *Outcome harvesting*. Retrievedfrom http://www.managingforimpact.org/sites/default/files/resource/outome_harvesting_brief_final_2012-05-2-1.pdf

World Bank. (2014). *Cases in outcome harvesting: Ten pilot experiences identify new learning from multi-stakeholder project to improve results*. Washington, D.C.: The World Bank. Retrieved from: http://documents.worldbank.org/curated/en/419021468330946583/pdf/901720WP0Box380n0Outcome0Harvesting.pdf

Worthen, B. (1990). Program evaluation. In H. Walberg & G. Haertel (Eds.), *The international encyclopedia of educational evaluation* (pp. 42-47). Toronto, ON: Pergammon Press.

Youker, B. W. (2005). Goal-free evaluation of the 2005 Kalamazoo Public School Middle School Summer Enrichment Program: Final report. Kalamazoo, MI: Western Michigan University, The Evaluation Center.

Youker, B. W. (2013). Goal-free evaluation: A potential model for the evaluation of social work programs. *Social Work Research, 37*(4), 432-438.

Youker, B. W., Ford, K., & Bayer, N. (2017). Dismissing the goals: A comparison of four goal-dismissive goal-free evaluations. *Evaluation and Program Planning*, *46*, 1-6

Youker, B. W., Ingraham, A., & Bayer, N. (2014). An assessment of goal-free evaluation: Case studies of four goal-free evaluations. *Evaluation and Program Planning*, *46*, 10-16.

Youker, B. W., Zielinski, A., Hunter, O. C. & Bayer, N. (2016). Who needs goals? A case study of goal-free evaluation. *Journal of Multidisciplinary Evaluation, 12*(27), 27-43.