

Articles should deal with topics applicable to the broad field of program evaluation. Articles may focus on evaluation methods, theory, practice, or findings. In all cases, implications for practicing evaluators should be clearly identified. Examples of contributions include, but are not limited to, reviews of new developments in evaluation, descriptions of a current evaluation study, critical reviews of some area of evaluation practice, and presentations of important new techniques. Manuscripts should follow APA format for references and style. Length per se is not a criterion in evaluating submissions.

Reported Influence of Evaluation Data on Decision Makers' Actions

An Empirical Examination

Christina A. Christie
Claremont Graduate University

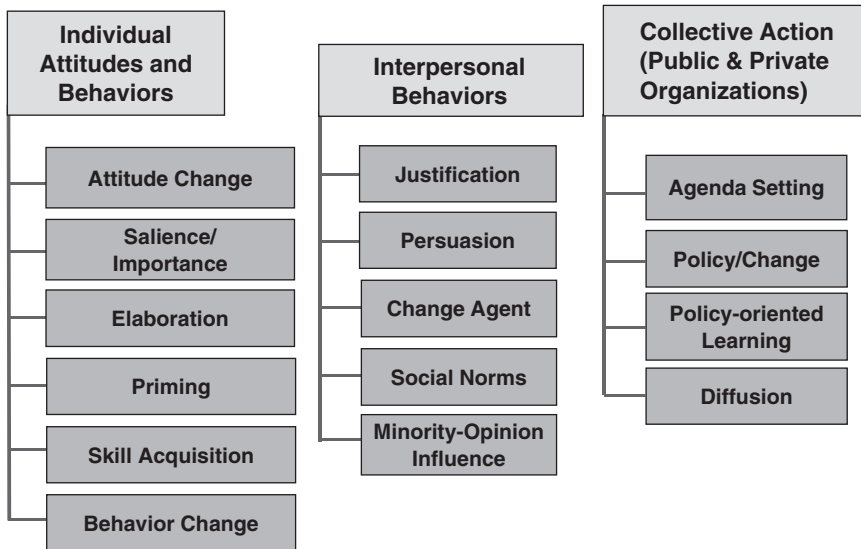
Abstract: Using a set of scenarios derived from actual evaluation studies, this simulation study examines the reported influence of evaluation information on decision makers' potential actions. Each scenario described a context where one of three types of evaluation information (large-scale study data, case study data, or anecdotal accounts) is presented and a specific decision needs to be made. Participants were asked to indicate which type of data presented would influence their decision making. Results from 131 participants indicate that participants were influenced by all types of information, yet large-scale and case study data are more influential relative to anecdotal accounts; certain types of evaluation data are more influential among certain groups of decision makers; and choosing to use one type of evaluation data over the other two depends on the independent influence of other types of evaluation data on the decision maker, as well as prior beliefs about program efficacy.

Keywords: *evaluation influence; decision making*

Introduction

Evaluation utilization is arguably the most researched area of evaluation and it also receives substantial attention in the theoretical literature. By this time, we have come to reasonable, but not universal, agreement on a definition of use, that is, the effect the evaluation has on the evaluand—the “thing” being evaluated—and those connected to the evaluand. Thus, use is a central outcome of any evaluation, because without it, evaluation cannot contribute to one of its primary objectives, social betterment.

Figure 1
Levels and Mechanisms Through Which Evaluation Produces Influences



Source: Henry and Mark (2003).

Terms such as *conceptual use*, *instrumental use*, *process use*, and *symbolic use* have seeped into our evaluation lexicon and help to describe this central dimension of evaluation. Kirkhart (2000), for example, and others (e.g., Henry, 2000; Henry & Mark, 2003; Weiss, Murphy-Graham, & Birkeland, 2005) have argued, however, that describing the changes that occur as a result of an evaluation as *evaluation use* is limiting and they are better understood if referred to as evaluation influence. The term *evaluation influence*, according to Kirkhart, moves past the term *use*, which was originally associated with data-based (or results-based) evaluation findings. She argues that *influence* is an umbrella term that addresses the conceptual limitations and simplifies the awkward language used to describe evaluation use beyond results-based use, such as process use and symbolic use.

Compared with evaluation use, the empirical literature on evaluation influence is sparse and relatively little is known about the specifics of how evaluation may influence decision makers' attitudes and actions. Henry and Gordon (2004) maintain that there is only limited research that "unpacks how evaluations wield their influence" (p. 4) and the influence of evaluation on program, policy, or participant improvement.

Mark (2004) argues that although the need for more empirically derived evaluation practice has been stressed for almost four decades, we are largely a field of "expert-based . . . evaluation practice, rather than an evidence-based evaluation practice" (p. 1). He poses several broad questions about evaluation practice and its effect, including, "How do we know which types of evaluation are more likely to make what kinds of contributions?" (p. 1). Mark also suggests that such

Christina A. Christie, Claremont Graduate University, 123 E. 8th Street, Claremont, CA 91711; phone: (909) 607-9020; e-mail: tina.christie@cgu.edu.

Author's Note: The author wishes to thank Katie Martin for her support with conducting this study and Sanjeev Sridharan for his helpful editorial feedback and comments.

ambitious research questions may be understood if we begin with more simple research questions that will build a body of evidence that could be considered collectively. This study is an attempt to understand a quite specific area of evaluation influence, that is, the reported influence of evaluation information on decision makers' potential actions, in hopes that we will develop a greater understanding of the relative benefits of different evaluation practices.

Henry and Mark (2003) and Mark and Henry (2004) offer a framework for developing a better understanding of evaluation influence, which serves as the conceptual foundation (and inspiration) for this study. Their framework, depicted in Figure 1, has three levels: individual, interpersonal, and collective. Each level is further explained by identifying specific mechanisms, measurable outcomes, and forms of influence.

This study examines the first level of evaluation influence, the individual. It is intended to provide a general understanding of the types of information decision makers (i.e., a specific actor in an evaluation) report as having influence on decisions related to programs. This study is intended to contribute to what I hope will one day become a larger body of literature on evaluation influence, which, as Mark suggests, can be examined collectively to develop an empirical literature on the evaluation process.

Presented in this article is a descriptive analog (Mark, 2004) or simulation study. As such, a set of scenarios derived from actual evaluation studies was constructed (as opposed to using real events experienced by study participants during an evaluation with which they had been involved). Each scenario describes a context where one of three types of evaluation information, large-scale study data, case study data (occasionally referred to as small-scale study data in this article), or anecdotal accounts, is presented and a specific decision needs to be made. Participants were then asked to indicate the extent to which the type of information presented would influence their decision. As such, this study offers a first step in understanding the complex web of factors that influence decision makers' choices.

The questions guiding this study are summarized:

- What is the likelihood that evaluation information will influence decision makers' actions?
- Are certain types of data reported to be more influential?
- When asked about relative influence, which data source do decision makers choose?

Brief Review of Relevant Literature

Evaluation influence is defined as "the capacity or power of persons or things to produce effects on others by intangible or direct means" (Kirkhart, 2000, p. 7). Influence has gained traction in recent years as a new term for describing the effect of an evaluation. Kirkhart (2000) authored a noteworthy article on the topic where she presents an integrated theory of evaluation influence and suggests the use of the word *influence* instead of *use* in an attempt to clarify the various definitions of evaluation use. Her integrated theory has three dimensions: source of influence, intention, and time. Source of influence is divided into two subgroups: results-based influence and process-based influence. Intention is also divided into two subgroups: intended influence and unintended influence. The final dimension, time, is divided into three subgroups: immediate, end-of-cycle, and long-term. Kirkhart presents nine potential applications for her integrated theory including strengthening the validity of studies of influence and to support theory building.

Mark and Henry (2003) have embraced the term *influence* and support Kirkhart's assertion that the field would be better served if the conversation about use was reframed and refocused on influence. They posit that Kirkhart's model offers a framework for empirically investigating and conceptualizing an evaluation's effect beyond use. To move this agenda forward, they

build from Kirkhart's model a theory of evaluation influence that identifies multiple levels, pathways, and mechanisms in an effort to explain influence. A version of this model is shown in Figure 1.

Henry and Mark (2003) and Mark and Henry (2004) identify three levels of influence: individual, interpersonal, and collective. At each level, they have identified between four and six mechanisms (change processes or outcomes), derived from social science literature, that may occur as a result of an evaluation. At the individual level, the authors refer to changes that occur in an individual's knowledge, attitudes, opinions, or actions as a result of the evaluation process or results. At this level, six mechanisms and measurable outcomes are identified: attitudinal change, salience, elaboration, priming, skill acquisition, and behavior change. The interpersonal level describes changes that occur as a result of interactions between individuals. Here, five mechanisms are identified: justification, persuasion, change agent, social norms, and minority-opinion influence. The third level, the collective, depicts the "direct or indirect influence of evaluation on the decisions and practices of organizations, whether public or private" (Henry & Mark, 2003, p. 298). Four mechanisms further define this level, which include agenda setting, policy-oriented learning, policy change, and diffusion.

This study focuses on decision making, a dimension of the behavioral change process mechanism at the individual level. Henry and Mark (2003) describe behavioral change as a mechanism that "connotes changes in what individuals do, in terms of the delivery of, or participation in, a program or process. Skills involve cognition and action" (p. 299). They suggest that this is an important mechanism when attempting to understand instrumental use, that is, the more direct or immediate use of evaluation information for decision making. To clarify how various kinds of information influence decision makers' actions, study participants were presented scenarios containing one of three kinds of information: large-scale study data, case study data, and anecdotal accounts. These three types of information do not exhaust the various categories or types of evaluation information that is generated by today's professional evaluators. However, they are qualitatively different from each other, thus offering distinct categories, and represent three general, important, and common categories of information used by decision makers (Tan & Halpern, 2006). Scenarios call for an immediate decision and, as such, link the influence of different types of information to instrument use. Thus, Henry and Mark's model provides a useful conceptual framework for this study, which investigates the influence of particular types of evaluation information on decision makers.

Method

Participant Recruitment and Study Procedure

The main study sample was recruited from two academic educational leadership advanced degree programs, administered at a large public research university located in California. These two highly regarded advanced degree programs have a long history of training educational leaders and, as such, offered a larger sample of individuals working in leadership roles and positions, with varying years of leadership experience, working in a variety of contexts. Education programs have been targets of evaluation inquiry for decades, and as a result, educational leaders are regularly asked to make decisions about programs presumably using some type of evaluation information. As such, they were a particularly appropriate sample for investigating the relationship between evaluation information and evaluation influence.

The program directors were contacted and asked for permission to survey current students as well as program alums. An e-mail soliciting voluntary participation was sent to a sample of 297 people from the program directors (but written by the researcher), which explained the purpose of the study and provided an html link to the online study instrument. The instrument was hosted

on a popular Web-based survey service. Participants did not receive compensation for participation. To ensure a reasonable sample size and to offer a comparative group of decision makers, a small sample ($n = 29$) of program directors was also recruited from a large California county health department. Directors of county health department programs were sent the identical e-mail from a senior program director. Data were collected over a 1-month period during fall 2004.

Simulation as a Method of Measuring Decision Making

The use of simulation studies in evaluation research is well documented in the literature. Some of the most notable studies were conducted by Brown, Braskamp, and Newman from 1978 to 1985 (Braskamp, 1980; Brown, Braskamp, & Newman, 1978, 1982; Brown & Newman, 1982; Brown, Newman, & Rivers, 1985; Ory & Braskamp, 1981). These studies are an important part of the evaluation empirical literature and have provided central insights into evaluation use. In particular, simulation studies provided a way to empirically study factors related to use being discussed in the literature (Alkin, Daillak, & White, 1979; Patton, 1997) in a systematic and controlled manner. For example, by randomly assigning stakeholders to different simulation contexts, Brown, Newman, and Rivers were able to examine the likelihood of evaluation utilization given varying evaluator characteristics (Braskamp, 1980; Brown et al., 1982), report styles (jargon free or jargon filled; Brown et al., 1978; Brown & Newman, 1982), and contentiousness of the evaluation context (Brown et al., 1985). As the use of simulations grew, Brown, Braskamp, and Newman (1982) conducted a meta-analysis of those simulation studies related to use.

The validation of simulation studies has been a topic of intense discussion in the literature of many disciplines using simulations (e.g., Dickson, Whitely, & Faria, 1990; O'Neil, Allred, & Dennis, 1997; Prohaska, Keller, Leventhal, & Leventhal, 1987; Rausser & Johnson, 1975; Ward, 1954). Simulations can be validated by examining their level of fidelity,¹ conceptual validity,² convergent validity,³ internal validity, and external validity (Feinstein & Cannon, 2002). Although the validation methods are meant to increase the realism and applicability of simulation studies, all researchers recognize that the main limitation of simulations is that participants are making "riskless choices" (Ward, 1954), meaning that when responding, participants do not have to deal with the real consequences of their decisions. Although this limitation reduces the realism of any simulation, studies that compared simulation results to other empirical methods found that simulation studies produced similar results to other methodological approaches (e.g., Cannon & Burns, 1999; O'Neil et al., 1997; Smith, Winer, & George, 1983). Thus, even though there are some limitations, simulations should still be considered a useful tool for the systematic examination of theoretical propositions in evaluation.

Study Instrument

The instrument was developed and pilot tested by the researcher. It was designed to capture the influence of three types, or sources, of evaluation information: large-scale evaluation study data, case study evaluation data, and anecdotes. The scenarios presented in the instrument were derived from real evaluation situations and address decisions related to the following kinds of programs: pedestrian safety, substance abuse prevention, literacy instruction, school-based consensus decision making, preschool programs, after-school programs, university-sponsored high school academic outreach programs, learning communities, civic participation, and mentoring programs. To develop the scenarios, the researcher conducted open-ended interviews with five practicing evaluators, asking each to describe at least three real situations that might be used for a simulation study examining the effect of different types of information on decision makers. Each interview lasted between 90 minutes and 2 hours. The situations described

by the evaluators were then used to develop the scenarios used in the instrument. Each evaluator was asked to review the scenarios derived from their own experiences for feedback and accuracy, and scenarios were then revised based on the evaluators' suggestions.

The instrument was divided into four sections. Section 1 contains 10 items intended to measure participants' beliefs about the efficacy of the programs that were presented in the scenarios. Items in section 1 were included to measure and control for biases about the programs presented in the scenarios that could have potentially influenced participants' responses.

Items in section 2 measured the effect of the three specific kinds of evaluation information on decisions. Three questions were included for each type of evaluation data, and the influence of information was measured using a 9-point scale that contained response options ranging from *not at all* to *somewhat* to *very much*.

Decision making is a complex process and is difficult to study in any context. Scenarios have a long history of use in the study of decision making (e.g., Einhorn & Hogarth, 1986; Ford & Richardson, 1994; Hodgkinson, Bown, Maule, Glaister, & Pearman, 1999; Pablo, 1994). Thus, a simulation study design was chosen for this study. However, it was important to communicate to participants that the scenarios presented were not intended to capture the complexity of decision making. To be sure that scenarios did not go beyond the scope of the purpose of the study, they were written to be focused on data type and, to avoid confusion, did not offer extensive detail about the situation beyond the most pertinent information. Rather, each scenario was designed to isolate a single factor, information type, on decision makers' potential actions. To ensure that participants were clear about the purpose and scope of the scenarios, the lead-in to section 2 stated:

The purpose of this study is to better understand the way people make decisions related to programs. To follow are 9 short scenarios. Following each scenario you are asked to imagine yourself in a decision making role, and then to indicate the extent to which the information presented in the scenario would influence your decision making. We understand that decision making is a complex process and often requires careful consideration of many factors and information from a variety of sources, including context, politics and personalities. The scenarios presented are simple by design. This is not meant to diminish the complexity of decision making, but to help us understand the impact of specific kinds of information on decisions.

Each of the nine items in section 2 presented a scenario describing a situation where evaluation information was to be used to make a decision. Scenarios had no specific mention of only quantitative or qualitative data. Instead, descriptions either implied the kind of data that were collected or, when there was specific reference to data, both quantitative and qualitative data were mentioned in the scenario. There were three scenarios for each of the three evaluation information types. An example of a scenario for each evaluation information type is presented in Table 1.

In section 3, participants were asked to imagine four different situations where they would need information for decision making. Following each of the four scenarios were three responses, one related to each of the three types of information, and participants were asked to choose the one that would most likely influence their decision making. These items were intended to measure relative evaluation information type preference. An example of an item from section 3 follows:

21. Imagine that last year your school implemented a substance-abuse prevention curriculum. This same curriculum was implemented across your state. Participating in a statewide evaluation of the curriculum are 1500 students from 50 schools. Your school, however, was NOT selected as one of the evaluation sites. You have to decide whether to revise the curriculum at your school for the following year.

Table 1
Sample Survey Items: Scenarios

Item	Data Type
14. A study funded by the Department of Education collected data from over 24,000 students throughout the United States. The results indicated that after-school programs have no statistically significant impact on children's academic achievement. However, the study found after school programs to have a positive impact on students' social behavior. Imagine that you have been asked to increase funding for an after-school program at your school. Please indicate the extent to which this information would influence your decision to increase funding for after-school programs.	Large-scale evaluation study data
12. A case study of school governance conducted at a local district found that when a consensus decision-making model was adopted, schools were managed more effectively, as indicated by both qualitative and quantitative measures. Imagine that you are in charge of a large urban school and a state mandate has come down requiring you to specify and implement a new school governance model. Please indicate the extent to which you would consider adopting a consensus decision-making model based on these case study data.	Case study/small-scale evaluation study data
16. A local newspaper published an article highlighting a community college's experience with learning communities. It was reported that learning communities increase the retention and academic success of at-risk students. Imagine that you are an administrator at a community college where the majority of students are at risk. Please indicate the extent to which this information would influence your decision to make funding available for faculty who were interested in implementing learning communities in their classrooms.	Anecdotal accounts

Please check the information that would most likely influence your decision.

- Remarks about the curriculum provided to you from teachers
- A case study examining the curriculum, conducted at another school similar to yours
- Data from the statewide evaluation

As a measure of consistency, section 3 also included a repeat item, that is, an item already asked earlier in the instrument (thus, only four of the five items included in section 3 were unique items). The correlation between the two identical items was .9.

Section 4 of the instrument included six demographic items: gender, race/ethnicity, age category, highest degree obtained, subject area of highest degree, and type of institution at which employed.

The instrument was pilot tested for readability and understandability using a sample of 14 program directors and academic deans from a Southern California community college. Each pilot study participant was asked to complete the instrument and note any words or phrases that needed revision in any way. A focus group was held immediately after participants completed the instrument about the instrument and its administration. Two focus groups were conducted, one with eight and the other with six participants. Data from the focus groups were used to revise the instrument.

Sample

A total of 326 individuals were contacted to participate in this study, and 131 (40%) completed the questionnaire online. All graduates of the two educational leadership programs described previously and all of the program directors at a large county health department were asked to participate in the study. The majority of the study sample was female and either Caucasian or Latino/a. Most participants also had an advanced degree in education or school

Table 2
Characteristics of the Sample (n = 131)

	<i>n</i> ^a	%
Age		
Younger than 30	16	15.2
30-39	50	47.6
40-49	20	19.0
50+	19	18.1
Race/ethnicity		
Black	2	1.9
Latino/a	21	20.1
White	70	67.3
Asian/Pacific Islander	4	3.8
Mixed race/ethnicity	7	6.7
Sex		
Female	69	65.7
Male	36	34.2
Highest degree obtained		
Bachelors	12	11.4
Masters	69	65.7
Doctorate	24	22.8
Degree program		
Art/music	1	0.9
Business	2	1.9
Education	66	62.8
Evaluation/research methods	1	0.9
Psychology	7	6.6
Public health	3	2.8
School administration	10	9.5
None of the above	15	14.2
Institution employed		
College/university	20	19.0
For-profit	4	3.8
Government agency	2	1.9
Public K-12 district office	2	1.9
K-12 school	72	68.5
Other	5	4.7

a. Column totals may not add to the total sample size because of missing data.

administration and, at the time of survey administration, were employed in a related field (see Table 2).

Analysis

Using natural cutoff points internal to the scale and defined a priori, three categories were created for responses to each question in section 2 to summarize the influence of each type of data on decision making. The categories, *greatly influenced*, *somewhat influenced*, and *not at all influenced*, were then used to describe the extent to which decision makers were influenced by information from large-scale evaluation studies, case studies, and anecdotes.

Analysis proceeded in four stages to accomplish the goals and answer the questions laid out at the beginning of the study. In the first stage, the researcher looked at characteristics of the sample and the extent to which participants were influenced by types of evaluation information.

In the second stage, differences in the influence of evaluation data were examined across demographic groups. Third, relationships between types of data were examined. Finally, factors that were associated with choosing to use one type of information over the other two were examined. This article presents results from logistic regression models and uses odds ratios and 95% confidence intervals to describe and interpret associations. Readers should keep in mind that these analyses are descriptive in nature and thus are not intended to provide predictive conclusions about what types of information influence decision makers.

Results

What Is the Likelihood That Evaluation Data Will Influence Decision Makers' Actions?

The questions included in section 2 of the study instrument presented a scenario, and then participants were asked how this information would influence their decisions when implementing or planning for various health and educational programs. Descriptive analyses indicate that, overall, participants were likely to use all three types of evaluation data. The vast majority of participants reported that large-scale evaluation study data, case study data, and anecdotes would greatly or somewhat influence their actions. Furthermore, each type of information greatly influenced approximately 50% of participants. Very few participants reported that data would not influence their decisions at all, with the exception of anecdotal accounts (see Table 3).

Are Certain Types of Data Reported to Be More Influential?

Descriptive analyses indicate that, overall, a similar distribution of participants reported being greatly, somewhat, or not at all influenced by each type of evaluation information (see Table 3). However, type of evaluation information was found to be more influential among certain groups of decision makers. In particular, it was found that the extent of influence differed by educational background, sector of employment, and the degree to which decision makers were influenced by other types of evaluation data (presented in Tables 4 and 5).

As expected, differences in the influence of data from large-scale evaluation studies, case studies, and anecdotes were not observed across age, sex, and race (these results are not reported in tables). Although degree of influence did not vary markedly across these demographic groups, the influence of large-scale evaluation study data differed between participants who received degrees or were employed in the field of education and the rest of the sample (see Table 4). The odds of being greatly influenced by large-scale study data were much less (OR = .35) among participants who were employed at public K-12 schools or district offices. Similarly, it was observed that participants who received their degree in education or school administration were .57 times less likely to be influenced by large-scale study data.

Furthermore, strong positive associations were observed between the influences of large-scale and case study evaluation data. The influence of case study evaluation data was also strongly associated with that of anecdotes. However, an association between the influence of information from large-scale evaluation studies and anecdotes was not found. Thus, participants who were greatly influenced by large-scale evaluation study data were not more likely to use anecdotal accounts in the decision-making process and vice versa. However, decision makers who were greatly influenced by case study data were 5.77 times more likely to report that they would use large-scale evaluation study data and almost equally as likely (OR = 5.48) to report using anecdotes to inform their decisions.

Table 3
Percentage of Participants Endorsing Each Data Type (*n* = 131)

	<i>n</i> ^a	%
Section 2		
Influence of large-scale study data		
Greatly	63	60.0
Somewhat	42	40.0
Not at all	0	0.0
Influence of case study evaluation data		
Greatly	52	48.6
Somewhat	55	51.4
Not at all	0	0.0
Influence of anecdotes		
Greatly	52	50.0
Somewhat	50	48.0
Not at all	2	1.92
Section 3		
Relative influence		
Data chosen to implement a pedestrian safety program		
Large scale	46	43.8
Case study	47	44.7
Anecdotes	12	11.4
Data chosen to implement a substance abuse prevention program		
Large scale	42	40.3
Case study	27	25.9
Anecdotes	35	33.6
Data chosen to implement a program within one's organization ^b		
Large scale	43	40.9
Case study	43	40.9
Anecdotes	19	18.1
Data chosen to implement a program within one's organization ^b		
Large scale	76	72.3
Case study	24	22.8
Anecdotes	5	4.7
Data chosen to implement a program within one's organization ^c		
Large scale	69	65.7
Case study	31	29.5
Anecdotes	5	4.7

a. Column totals may not add to the total sample size because of missing data.

b. When suspected to receive resistance from staff.

c. When in favor of program that has not been proven successful.

When Asked About Relative Influence, Which Data Source Do Decision Makers Choose?

When examining the frequency distributions of responses to the five items asking participants to select one type of information over the other two, no clear trends were observed. It should be noted, however, that for four of the five items, a smaller number of participants chose to use anecdotal information over large-scale evaluation and case study

Table 4
The Influence of Information From Large-Scale Evaluations:
Differences by Degree Program and Employment Sector^a

	Greatly Influenced (<i>n</i> = 63)		Not Greatly Influenced (<i>n</i> = 42)		OR ^b	95% CI	<i>p</i> Value
	<i>N</i>	%	<i>N</i>	%			
Degree program							
Education or school administration	41	65.0	34	80.95	0.59	0.22, 1.61	0.02
Other degree	22	34.9	8	19.05	1		
Institution employed							
Public K-12 school or district office	38	60.3	35	83.33	0.37	0.14, 1.00	0.05
Other institution	25	39.6	7	16.67	1		

Note: 1 indicates the referent group.

a. Analysis adjusted for the participants' beliefs about the efficacy of programs in the questions.

b. Odds ratios computed from one logistic regression model.

Table 5
Associations Between Types of Evaluation Data: Odds of Being
Greatly Influenced by Large-Scale Study, Case Study,
and Anecdotal Information^a

	Large Scale			Case Study			Anecdotes		
	OR ^b	95% CI	<i>p</i> Value	OR	95% CI	<i>p</i> Value	OR	95% CI	<i>p</i> Value
Large scale									
Greatly influenced				5.77	2.1, 15.2	0.0004	0.74	0.2, 1.9	0.54
Not greatly influenced				1			1		
Case study									
Greatly influenced	7.12	2.5, 19.8	0.0002				6.10	2.3, 15.7	0.0002
Not greatly influenced	1						1		
Anecdotes									
Greatly influenced	0.74	0.2, 1.9	0.54	5.48	2.1, 13.8	0.0003			
Not greatly influenced	1			1					

Note: 1 indicates the referent group.

a. Analysis adjusted for the participants' beliefs about the efficacy of programs in the questions.

b. Odds ratios model the odds of being greatly influenced.

data (see Table 3). Also of interest, relative influence of large-scale evaluation study data was highest when participants were asked about implementing a program in their own organization (see Table 3).

Using logistic regression, no differences were found in the relative influence of evaluation data across demographic groups, which, again, was expected (not reported in table).

As seen in Table 6, it was found that relative influence was associated with the independent influence of certain types of data, specifically when examining the decision to implement a pedestrian safety program. Being greatly influenced by large study and anecdotal information corresponded with choosing to use that particular type of information in decision making. It was

Table 6
The Relative Influence of Three Data Sources on Implementing a Hypothetical Pedestrian Safety Program: Odds of Choosing to Use One Type of Information Over the Other Two^a

Level of Influence	Relative Choice			Large Scale			Case Study			Anecdotes		
	Large (n = 46)	Case Study (n = 47)	Anecdotes (n = 12)	OR ^a	95% CI	p Value	OR	95% CI	p Value	OR	95% CI	p Value
Large scale												
Greatly influenced	31	24	6	2.44	0.9, 6.2	0.06	0.64	0.2, 1.5	0.31	0.29	0.0, 1.3	0.11
Not greatly influenced ^b	14	22	6	1			1			1		
Case study												
Greatly influenced	20	22	8	0.71	0.2, 1.9	0.49	0.95	0.3, 2.4	0.92	2.72	0.5, 13.1	0.21
Not greatly influenced ^b	26	25	4	1			1			1		
Anecdotes												
Greatly influenced	18	25	8	0.42	0.1, 1.0	0.72	1.61	0.6, 3.9	0.30	3.18	0.6, 15.8	0.16
Not greatly influenced ^b	27	20	4	1			1			1		
Program beliefs												
Agreed ^c	41	34	7	3.69	1.1, 11.8	0.03	0.65	0.2, 1.8	0.41	0.19	0.0, 0.8	0.03
Didn't agree	5	13	5	1			1			1		

Note: 1 indicates the referent group.

a. Odds ratios model the odds of using one type of information over the other two.

b. Category includes *somewhat influenced* and *not at all influenced*.

c. *Agreed* means that on a 4-point scale with two additional responses for *don't know* and *no opinion*, the participant agreed or somewhat agreed that the type of program is effective (versus *disagreed* or *somewhat disagreed*).

found that the odds of using large study results compared with case study results or anecdotal information was 2.44 times greater for participants who were greatly influenced by large study information. Similarly, the odds of using anecdotal information compared to information from large or case studies were 3.18 times greater for participants who were greatly influenced by anecdotal information.

Conversely, participants who were greatly influenced by anecdotal information had less than half the odds (OR = 0.42) of using large study information over case study and anecdotal information. And, participants who were greatly influenced by large-scale study information had 0.29 times the odds of using anecdotal information over large and case study information. No correspondence was found between the degree of influence of case study information and choice to use case study information over other types of information.

Participants who agreed that pedestrian safety programs were important had 3.69 times the odds of using large study information rather than case study or anecdotal information, and had 0.19 times the odds of using anecdotal information over the other two. This was consistent with results of other conditions.

Discussion

This article contributes to a better understanding of a specific but relevant question about evaluation influence: Which types of information reportedly influence decision makers' potential actions? The results presented herein offer an empirical measure of the reported effect of evaluation data on decision making and aid in answering the three questions presented in the introduction of this article.

Participants were influenced by all three types of information, yet, in general large-scale and case study data, they tend to be more influential relative to anecdotal accounts.

Descriptive analyses indicate that the vast majority of participants were somewhat or greatly influenced by all types of evaluation data and very few participants reported not being influenced at all. These results suggest that people in decision-making roles rely on different types of information when making decisions. Thus, it can be presumed that when faced with real, live decisions, the complexities presented by the various social and political factors (which were not seriously considered in the scenarios presented in this study) may well be of greater influence than type of evaluation information (Granville, 1977). This suggests that although an evaluator may be aware of some of the particular factors influencing decision makers' decisions, when developing studies, it may be most advantageous for evaluators to ask decision makers specifically to identify the particular types of evaluation information that they believe would be most influential given the social and political context of the evaluation.

Results indicate that when examining relative influence of evaluation information type, large-scale study data were always chosen over anecdotal accounts. In one of the five scenarios, case study data were chosen most frequently, relative to the others, and in another scenario, large-scale and case study were evenly selected by participants. This points to the decision maker's desire for both broad and more in-depth information when making decisions about programs, thus reflecting the need for evaluation studies that are designed to yield both types of information.

In one scenario, an anecdotal account was chosen more frequently relative to case study information (however, in this scenario, large study data were chosen most frequently relative to the other two types of information). Although this occurred in only one scenario, this suggests

that, although not very frequently, anecdotal evidence may be seriously considered by decision makers in particular circumstances. This finding substantiates what has been described as the power of a good anecdote. For example, when presenting a paper on applied psychology, Diane Halpern, then president of the American Psychological Association, told a story of going before a U.S. Congressional committee to present research study findings, and immediately prior to her appearance before the group, a congressional aid advised her “to tell a good story” when describing the implications of the research (Tan & Halpern, 2006). Thus, evaluators should not necessarily dismiss the influence of anecdotal evidence but rather acknowledge that it does, in some cases, have influence. Therefore, distinguishing between case studies (which involve rigorous systematic data collection) and anecdotal accounts (which are better described as case examples) and the limitations related to each is advantageous when decision makers are considering such information for program decisions.

Choosing to use one type of evaluation data over the other two may depend on the independent influence of other types of evaluation data on the decision maker, as well as his or her prior beliefs about program efficacy.

The influence of each type of information was associated with the influence of one or both of the other types of evaluation data. This study found that participants who were greatly influenced by large-scale evaluation study data were not more likely to use anecdotal accounts in the decision-making process and the reverse was also true. This suggests that those who prefer large-scale data were more likely to discern the methodological differences between case studies and anecdotes (or case examples). It may be that those who are influenced by large-scale data better understand the strength and rigor of a case study compared to that of an anecdotal account. However, decision makers who were greatly influenced by case study data were likely to report that they would use both large-scale evaluation study data and anecdotes to inform their decisions. It may be that those who are influenced by case study data are more likely to use a “story” or account in some decision-making situations but, in other situations, see the value of using large-scale study information to inform decision making. This points to the need for a better understanding of how case study information and anecdotal accounts are used, and conditions under which those who tend to be more influenced by case studies use information yielded from large-scale studies for decision making. In sum, the results from the analysis of relative influence highlight a need for a better understanding of how types of evaluation data are related, are used alone, or are used in combination with each other.

This study also found that prior beliefs about program efficacy were associated with relative influence of evaluation information. Compared with those who did not agree, participants who agreed that pedestrian safety programs are effective and their implementation beneficial had increased odds of using data from large-scale evaluation studies and decreased odds of using anecdotes to implement such a program. This suggests that when decision makers believe in a program, they are less likely to be influenced by a story, be it an in-depth case study or an anecdotal account. Once convinced of a program’s efficacy, decision makers were more likely to use large-scale study data to influence their decisions, thus suggesting that large-scale study data are more influential when one is looking to confirm prior beliefs. This finding offers a potentially interesting avenue for further empirical work.

Certain types of evaluation data may be more influential among certain groups of decision makers.

For the most part, the influence of evaluation data did not vary across subgroups of the sample. However, participants with degrees in education and working in fields related to

education were less likely to use large-scale study evaluation data. This finding is worthy of discussion and, I believe, may be best understood in light of today's politics.

As most evaluators are aware, the No Child Left Behind Act (NCLB) offers preference to funding educational programs that have been, are currently, or will be tested using randomized control trials (RCT). The academic debate on this topic has been lively and emotionally charged at times (for examples, see Donaldson & Christie, 2005; "Scientific Culture," 2002). Many school administrators have opposed the legislation for a variety of reasons, including the RCT stipulation. When examining the results of this study in this political context, it may be that those opposed to the NCLB RCT stipulation have chosen to reject evidence from studies resembling experiments (or quasi-experiments) and what is being observed is a backlash on the part of decision makers incensed by NCLB. This, of course, is just speculation. Alternatively, because many educational evaluation studies have not employed RCT designs, it may be that educational decision makers are less familiar with the strengths offered by these study designs or are less accustomed to or comfortable with information generated from these types of studies and thus are less likely to be influenced by it. Whether or not these explanations are even in part legitimate, this finding has important (negative) implications for evaluators designing large-scale quasi- or experimental designs for answering questions about educational program effectiveness.

Henry and Mark's Framework: Some Empirical Insights

When considering the results of this study within the context of Henry and Mark's framework, it should be recognized that many factors, which have been identified by others (e.g., Alkin et al., 1979; Patton, 1997; Patton et al., 1977), influence the individual level of their model. This study addresses, within the context of this framework, the reported influence of different types of information on decision makers' (i.e., individuals') decisions. Findings offer empirical insight into the kinds of information reported to be influential, as well as the relative influence of such information. Of interest was the extent to which specific types of evaluation information are influential given a decision maker's prior beliefs about program efficacy. Related to this is the relationship between types of evaluation and this framework.

The Henry and Mark framework focuses on an individual's attitudes and behavior (see Figure 1). This study examined decision makers' actions, which are behaviors, and the extent to which prior beliefs about a program's effectiveness may affect an action or behavior, but did not examine attitudes. Study findings indicate that all three types of information influence decision makers' decisions. Decisions are actions (i.e., a behavior), thus suggesting that each of the three information types tested in this study may influence an individual's behavior. Prior beliefs about program efficacy were also found to influence decision makers. Therefore, when examining the various pathways to individual influence within the context of the framework, it may be valuable to also consider the influence that evaluation information may have on individuals' beliefs, in addition to what the current framework identifies: attitudes and behaviors.

In light of the study findings, there are some additional points to consider within the context of the Henry and Mark framework. Henry and Mark (2003) suggest that behavior change at the individual level "appears to be an important outcome in some evaluation theories, including some participatory and empowerment approaches" (p. 301). Although this study focused specifically on decision makers, findings may offer insight into the types of information that influence individuals involved in evaluations more generally. Thus, we might imagine that the findings of this study would also apply to other important program stakeholders, for example, program staff. Thus, when individual behavior change is a prescriptive theoretical outcome of an evaluation, we might consider the types of information that are more or less likely to influence an individual's behavior change.

Henry and Mark (2003) also suggest that in practice, the outcomes identified in their model are connected “like the links in a causal chain” (p. 305). For example, an individual-level attitudinal change might then cause changes in persuasion at the interpersonal level. Borrowing from the health behavior literature (e.g., Rosenstock, Strecher, & Becker, 1988), we know that changes in individuals’ attitudes and behaviors are linked. As a consequence, the findings of this study related to the influence of different types of information at the individual behavioral level are also relevant when considering changes in attitude at the individual level. It would be likely, then, that the same type of information that decision makers report influencing behavior would also influence attitude change, and perhaps even salience, and elaboration.

Results from this study can inform further research on evaluation influence. For example, future studies might examine the extent to which information generated from an evaluation influences individuals’ attitudes, behavior, and beliefs and identify which of the three is most influenced by evaluation information, and under what conditions. Extending the study sample to include program stakeholders other than decision makers would help identify whether certain types of evaluation information are more likely to influence individuals with varying roles and decision-making authority within program and organizational contexts. Finally, further research should look to examine how changes at the individual level of the Henry and Mark framework might be linked to changes in other levels of the framework.

Study Limitations

The “unreal” nature of a simulation study raises concerns about the generalizability of findings from these studies to real-world situations. A shortcoming of using this method for this study is that participants did not have to make an actual decision but, rather, were asked to speculate about the extent to which a particular type of information would be influential. Admittedly, there are many factors not included in the scenarios that influence decision makers, as is often the case with any simulation study design. For example, all of the cases presented in the scenarios assumed that rigorous, quality studies had been conducted (when examining larger scale and case study data). The scenarios in this study were not designed to consider the trustworthiness of the studies, which would be a consideration in a real-world context, thus posing a limitation on the extent to which the scenarios may in fact reflect a more real context for some participants. Efforts were taken to minimize such effects by using real evaluation studies to derive the scenarios; however, there is no substitute for using situations in which participants were actual actors and decisions were consequential.

Still, an argument can be made that there are benefits to developing an understanding of the reported effect of different types of evaluation information via a simulation study. For example, an advantage of a simulation study is that participants are asked to respond to the scenarios focusing solely on the evaluation information type, which were free of the social and political confounds that a real-life situation might bring. For the purpose of this investigation, simulation helped to isolate the effects of evaluation information. Nevertheless, the results from this study should be considered in the context of its limitations and should be tested further to see whether other methodological approaches would produce similar results.

In addition, the study sample was limited primarily to educational leaders. Findings suggest that the political context in which educational decision makers are currently working may have influenced participants’ responses. Although there is no specific reason to assume that educational decision makers are qualitatively different from decision makers in other contexts, because the sample was indeed limited to this particular population, I would again limit generalizing the results of this study beyond educational decision makers. I would also

suggest that further research be conducted with a broader sample of decision makers to validate the findings of this study across populations and contexts.

Notes

1. Simulations can take many forms, ranging from “high fidelity” simulations, which are defined by their high level of realism (such as flight simulators or the presence of actors to interact with), to “low fidelity” simulations (Feinstein & Cannon, 2002), which include presenting participants with a description of a hypothetical situation and asking them to respond to the description. Evidence suggests that there are no real differences between high fidelity and low fidelity simulations in terms of the reliability of the results (Alessi, 1988; Feinstein & Cannon, 2002).

2. Conceptual validity is concerned with ensuring that the simulation closely represents a real-world system.

3. Convergent validity looks at how the simulation results compare with other measures of comparable competencies.

References

- Alessi, S. M. (1988). Fidelity in the design of instructional simulations. *Journal of Computer-Based Instruction, 15*(2), 40-47.
- Alkin, M. C., Daillak, R., & White, B. (1979). *Using evaluations*. Beverly Hills, CA: Sage.
- Braskamp, L. A. (1980). Credibility of a local education program evaluation report: Author source and client characteristics. *American Educational Research Journal, 15*, 441-450.
- Brown, R. D., Braskamp, L. A., & Newman, D. L. (1978). Evaluator credibility and acceptance as a function of report styles: Do jargon and data make a difference? *Evaluation Quarterly, 2*, 331-341.
- Brown, R. D., Braskamp, L. A., & Newman, D. L. (1982). Studying evaluation utilization through simulations. *Evaluation Review, 6*, 114-126.
- Brown, R. D., & Newman, D. L. (1982). An investigation of the effects of different data presentation formats and order of arguments in a simulated adversary evaluation. *Education Evaluation and Policy Analysis, 4*, 197-204.
- Brown, R. D., Newman, D. L., & Rivers, L. S. (1985). Does the superintendent's opinion affect school boards' evaluation information needs? An empirical investigation. *Urban Education, 20*(2), 204-221.
- Cannon, H. M., & Burns, A. C. (1999). A framework for assessing the competencies reflected in simulation performance. *Developments in Business Simulation & Experiential Exercises, 26*, 40-44.
- Dickson, J. R., Whitely, R. R., & Faria, A. J. (1990). An empirical investigation of the internal validity of a marketing simulation game. *Developments in Business Simulation & Experiential Exercises, 17*, 47-52.
- Donaldson, S. I., & Christie, C. A. (2005). The 2004 Claremont debate: Lipsey vs. Scriven. Determining causality in program evaluation and applied research: Should experimental evidence be the gold standard? *Journal of Multidisciplinary Evaluation, 3*, 60-77.
- Einhorn, H. J., & Hogarth, R. M. (1986). Decision making under ambiguity. *Journal of Business, 59*(4), 225-250.
- Feinstein, A. H., & Cannon, H. M. (2002). Constructs of simulation evaluation. *Simulation & Gaming, 33*(4), 425-440.
- Ford, R. C., & Richardson, W. D. (1994). Ethical decision making: A review of the empirical literature. *Journal of Business Ethics, 13*(3), 205-221.
- Granville, A. (1977). *Where do school decisions really come from? Sources of influence and the influence of evaluation*. Unpublished doctoral dissertation, University of California, Los Angeles.
- Henry, G. T. (2000). Why not use? *New Directions for Evaluation, 88*, 85-98.
- Henry, G. T., & Gordon, C. S. (2004, November). *Coming full circle: Assessing evaluation's influence on program performance*. Paper presented at the American Evaluation Association annual conference, Atlanta, GA.
- Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation, 24*, 293-314.
- Hodgkinson, G. P., Bown, N. J., Maule, A. J., Glaister, K. W., & Pearman, A. D. (1999). Breaking the frame: An analysis of strategic cognition and decision making under uncertainty. *Strategic Management Journal, 20*(10), 977-985.
- Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation, 88*, 5-23.
- Mark, M. M. (2004, November). *Building a better evidence-base for evaluation theory: Examining questions of feasibility and utility and specifying "how to."* Paper presented at the American Evaluation Association annual conference, Atlanta, GA.

- Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation, 10*, 35-57.
- O'Neil, H. F., Allred, K., & Dennis, R. A. (1997). Validation of a computer simulation for assessment of interpersonal skills. In H. F. O'Neil, Jr. (Ed.), *Workforce readiness: Competencies and assessment* (pp. 229-254). Mahwah, NJ: Lawrence Erlbaum.
- Ory, J. C., & Braskamp, L. A. (1981). Faculty perceptions of the quality and usefulness of three types of evaluation information. *Research in Higher Education, 15*, 271-282.
- Pablo, A. L. (1994). Determinants of acquisition integration level: A decision making perspective. *Academy of Management Journal, 37*(4), 803-836.
- Patton, M. Q. (1997). *Utilization focused evaluation*. Thousand Oaks, CA: Sage.
- Patton, M. Q., Grimes, P. S., Guthrie, K. M., Brennan, N. J., Franch, B. D., & Blyth, D. A. (1977). In search on impact: An analysis of the utilization of federal health evaluation research. In C. H. Weiss (Ed.), *Using social research in public policy making* (pp. 141-163). Lexington, MA: D. C. Heath.
- Prohaska, T. R., Keller, M. L., Leventhal, E. A., & Leventhal, H. (1987). Impact of symptoms and aging attribution on emotions and coping. *Health Psychology, 6*(6), 495-514.
- Rausser, G. C., & Johnson, S. R. (1975). On the limitation of simulation in model evaluation and decision analysis. *Simulation & Games, 6*(2), 115-150.
- Rosenstock, I. M., Strecher, V. J., & Becker, M. H. (1988). Social learning theory and the health belief model. *Health Education Quarterly, 15*(2), 175-183.
- Scientific culture and educational research. (2002). *Educational Researcher, 31*(8).
- Smith, H. W., Winer, J. L., & George, C. E. (1983). The relative efficacy of simulation experiments. *Journal of Vocational Behavior, 22*(1), 96-104.
- Tan, S. J., & Halpern, D. F. (2006). Applying the science of psychology to a public that distrusts science. In S. I. Donaldson, D. E. Berger, & K. Pezdek (Eds.), *Applied psychology: New frontiers and rewarding careers* (pp. 153-170). Mahwah, NJ: Lawrence Erlbaum.
- Ward, E. (1954). The theory of decision making. *Psychological Bulletin, 51*, 380-417.
- Weiss, C. H., Murphy-Graham, E., & Birkeland, S. (2005). An alternative route to policy influence: How evaluations affect D.A.R.E. *American Journal of Evaluation, 26*, 12-30.