

You Want Me to Do What? Evaluators and the Pressure to Misrepresent Findings

American Journal of Evaluation
34(1) 57-70
© The Author(s) 2012
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1098214012457237
aje.sagepub.com



Michael Morris¹ and Brittany Clark²

Abstract

Evaluators' experiences with being pressured by a key stakeholder to misrepresent findings were investigated in an anonymous survey of a random sample of over 2500 American Evaluation Association members; a response rate of 37% was obtained. Overall, 42% of the respondents had encountered misrepresentation pressure, with 70% of this subgroup having faced it on more than one occasion. Misrepresentation pressure usually focused on making evaluation results look more positive or less negative than the evaluator thought was warranted. In nearly 40% of all pressuring episodes, respondents believed that the stakeholder in question knew that he/she was requesting misrepresentation. Respondents' ethical satisfaction with how pressuring episodes were resolved varied widely, with satisfaction lowest when changes were made to a report without the respondent's consent. In the opinion of respondents, stakeholder pressure led to actual misrepresentation in 16% of all episodes, a finding that, in conjunction with the study's other results, suggests that most pressuring incidents are handled in a fashion that produces reasonably satisfactory, if not ideal, outcomes when viewed from the perspective of professional ethics. Just under one-third of all pressured respondents believed that something could have been done to prevent the influence attempt they encountered, and offered a variety of suggestions for preventive action. Viewed as a whole, these recommendations provide the foundation for a comprehensive approach to preventing and responding to misrepresentation pressure and related ethical challenges in evaluation.

Keywords

ethics, guiding principles, evaluation findings, research on evaluation, evaluation politics

Experienced evaluators know that ethical challenges can arise at virtually any point in a project, from the entry/contracting stage (e.g., exclusion of legitimate stakeholders from the planning process) to the "end game" of dissemination/utilization of results (e.g., disputes over the ownership/distribution of a final report). One of the most frequently reported conflicts in the literature is the pressure to misrepresent findings (e.g., Buchanan & MacDonald, 2011; Morris & Cohn,

¹ Department of Psychology, University of New Haven, West Haven, CT, USA

² The Brien Center for Mental Health and Substance Abuse Services, Pittsfield, MA, USA

Corresponding Author:

Michael Morris, Department of Psychology, University of New Haven, 300 Boston Post Road, West Haven, CT 06516, USA.
Email: mmorris@newhaven.edu

1993; Turner, 2003). Such pressure strikes at the core of the Guiding Principle of Integrity/Honesty, which asserts that “evaluators should not misrepresent their procedures, data, or findings” (American Evaluation Association [AEA], 2004). It also threatens the Program Evaluation Standard of Communication and Reporting (A8), which stipulates that “evaluation communications should have adequate scope and guard against misconceptions, biases, distortions, and errors” (Yarbrough, Shulha, Hopson, & Caruthers, 2011, p. 217), as well as Transparency and Disclosure (P5), which calls for evaluators to “provide complete descriptions of findings, limitations, and conclusions to all stakeholders unless doing so would violate legal or propriety obligations” (Yarbrough et al., 2011, p. 139).

Given the frequency of perceived misrepresentation pressure and its relevance to evaluation’s core principles and standards, it is important that researchers supplement anecdotal accounts with more comprehensive investigations of the nature of this phenomenon and the dimensions on which it can vary. Accordingly, this article reports the results of a large-scale survey focusing on evaluators’ experiences with misrepresentation pressure. Previous investigations have documented the *existence* of misrepresentation pressure but have not systematically explored evaluators’ views of the specific varieties of that pressure, where it comes from, how it is interpreted, the manner in which it is dealt with, and how it might have been avoided. Thus, the goal of the current study is not just to produce a more detailed picture of this ethical challenge and its contexts but to use this portrait to identify strategies that evaluators might employ to prevent and manage such pressure.

Method

In February 2009, a random sample of 2,523 nonstudent members of the AEA received an e-mail invitation to participate in an Institutional Review Board–approved study of their experiences with being pressured to misrepresent evaluation findings, where misrepresentation was defined as the “giving of a false or misleading account.” The e-mail contained a link to the survey site, which was configured in a fashion that made all responses anonymous. This fact was referenced in the invitation, and in both follow-up reminders that the entire sample received (at 1 week and 3 weeks following the original invitation). A total of 940 individuals responded to the survey, generating an overall response rate of 37%.

The survey asked respondents if they had “ever felt pressured to misrepresent, in a written report or oral presentation, any of the findings from a study” they had conducted. Respondents were also asked to indicate the number of times in their evaluation career that they had faced such pressure. The remaining questions asked respondents to focus on various aspects of the *most recent instance* in which they felt pressured to engage in misrepresentation. The most recent instance was targeted in order to maximize the accuracy of the respondents’ recollections. These items addressed the following issues:

- Was the study an external evaluation or an internal evaluation?
- What type of evaluation, or combination of types, was being conducted (implementation, outcome, etc.)?
- Did the evaluation take place in the United States or elsewhere?
- Who applied the pressure to misrepresent?
- What was the nature of the misrepresentation being sought?
- What attributions did the respondent make concerning the *intentions* of the individual exerting pressure?
- Were any changes made in the report as a result of the pressure, and if so, who made them? Did the respondent view these changes as misrepresentation?

Table 1. Primary Discipline of Respondents ($N = 875$).

	Frequency	Percent
Education	198	22.6
Evaluation	149	17.0
Psychology	113	12.9
Public health/medicine/nursing	99	11.3
Public administration/political science/policy/international relations	77	8.8
Research/statistics	60	6.9
Sociology	50	5.7
Social work	44	5.0
Economics/business	14	1.6
Anthropology	13	1.5
Other	58	6.6

Table 2. Primary Employment Setting of Respondents ($N = 875$).

	Frequency	Percent
College/university	290	33.1
Private business/consulting	257	29.4
Nonprofit organization	190	21.7
Federal agency	48	5.5
State agency	35	4.0
Local agency	21	2.4
School system	14	1.6
Other	20	2.3

Table 3. Evaluation Experience of Respondents ($N = 838-847$).

Number of years	Frequency	Percent
1-5	187	22.3
6-10	218	26.0
11-15	148	17.7
16 or more	285	34.0

Number of evaluations	Frequency	Percent
1-5	100	11.8
6-10	160	18.9
11-19	186	22.0
20 or more	401	47.3

- Did the respondent seek advice during the period when he or she was being pressured, and if so, from whom?
- Ethically, how satisfied were respondents with how the situation was resolved?
- Did the respondent believe that anything could have been done to prevent the influence attempt, and, if so, what?

A final section of the survey collected respondent demographic data.

Table 4. Evaluation Experience by Pressure to Misrepresent.

	Pressured to misrepresent?			χ^2
	Yes	No	Unsure	
Years of experience				
1–5	32	61	7	13.233*
6–10	38	54	8	
11–15	44	49	7	
16 or more	47	48	5	
N = 838				
Number of evaluations conducted				
1–5	30	62	8	15.054*
6–10	34	57	9	
11–19	43	47	10	
20 or more	44	51	5	
N = 847				

Note. Table figures represent row percentages.

* $p \leq .05$.

Results

Of the study's 940 respondents, 93% answered all of the survey questions that pertained to them, while 7% failed to complete one or more relevant items. The percentages reported in the following sections are based on the number of valid responses to each question, excluding nonrespondents.

Respondent Characteristics

The highest degree obtained by the majority of respondents was a doctorate (58%), with holders of a master's degree accounting for 37%, and those at the bachelor's level comprising 5% ($N = 874$). Consistent with the interdisciplinary nature of evaluation, a wide variety of responses was generated by the question, "What is your primary discipline?" Education was the most frequent choice (23%), followed by evaluation (17%), psychology (13%), and health-related fields (11%). No other field accounted for as many as 10% of the respondents (Table 1).

In terms of primary employment (Table 2), the top three settings were colleges/universities (33%), private business/consulting (29%), and nonprofit organizations (22%). With respect to evaluation experience, 97% of the respondents ($n = 913$) had in fact conducted an evaluation. Of this subgroup, more than half (52%) had 11 or more years of experience, and nearly half (47%) had conducted 20 or more evaluations (Table 3).

Respondents reported much more experience with external evaluations than internal ones. Nearly half (45%) of all respondents had conducted over 80% of their evaluations in an external role, while only 18% had conducted a similar percentage in an internal role.

The only characteristic that allows a direct comparison between survey respondents and nonrespondents is gender. Among respondents, 68% were female and 32% male. The corresponding figures for nonrespondents are 62% and 38%. Thus, females are slightly, but significantly, overrepresented in the respondent group, $\chi^2(1, N = 2,523) = 8.51, p = .004$.

Experience With Pressure to Misrepresent

Overall, 42% of the respondents (377 of the 905) indicated that they had been pressured to misrepresent findings, 51% maintained that they had never experienced such pressure, and 7%

Table 5. Outcome of Pressuring Episode ($n = 341$).

	Frequency	Percent
No changes were made in the disputed section/sections	95	27.9
Some changes were made, but they did NOT constitute misrepresentations	143	41.9
Some changes were made and they DID constitute misrepresentations	53	15.5
Never saw the final data	8	2.3
Project was terminated	6	1.8
Resigned/let go	5	1.5
Evaluation still in process	4	1.2
Other	27	7.9

indicated that they were “not sure.” Within the pressured subgroup, 70% reported that they had been pressured to misrepresent findings in more than one evaluation. Indeed, 7% claimed that such pressure had occurred in *seven or more* of their evaluations. Not surprisingly, evaluation experience is positively correlated with being in the pressured subgroup, regardless of whether experience is measured in terms of years as a practicing evaluator or number of evaluations conducted (Table 4).

The remaining analyses in this and the following section concern the *most recent* instance in which respondents felt pressured to engage in misrepresentation. Answers to the open-ended item dealing with the nature of this episode reveal that a subgroup of respondents (35 of 377, or 9%) did not follow the instruction asking them to focus on their most recent experience. Rather, they mentioned that they had been pressured on multiple occasions. This calls into question their responses to survey items that are intended to address a single incident. To address this issue, analyses were conducted to determine whether results for this subgroup significantly differed from those for the other pressured respondents. No differences were found. Thus, the results presented here are for all members of the pressured subgroup.

Over two thirds (68%) of the pressuring episodes involved external evaluations rather than internal ones (32%; $n = 368$). In 71% of the episodes, an evaluation incorporating impact/outcome assessment was being conducted, while implementation/process analysis characterized 51% of the cases (recall that more than one type of evaluation could be relevant to a given incident). Far fewer pressuring episodes involved needs assessment (9%), cost–benefit/cost–effectiveness analysis (6%), or other types of evaluation (5%; $n = 366$).

In most cases (69%), the pressure to misrepresent came from the individual who hired the respondent or assigned him or her to conduct the evaluation. The next most frequent source of pressure (16%) was the director of the program being evaluated. The remaining 15% were distributed among program staff members, evaluation supervisors/leaders, evaluation coworkers, and other stakeholders ($n = 371$). In 39% of the episodes, respondents believed that the individual applying pressure *knew* that he or she was requesting misrepresentation, while in 35% of the cases respondents did not perceive deceptive intent. Respondents were unsure of intent in 25% of the episodes ($n = 343$).

Respondents indicated that no changes were made in the disputed sections of the report in 28% of the episodes. In 42%, changes were made, but in the opinion of respondents these changes did not constitute misrepresentation. However, perceived misrepresentation did occur in 16% of the incidents (Table 5). In cases where changes were made ($n = 255$), over half of the time (54%) the respondent made them. Another party made them, with the consent of the respondent, in 13% of the incidents. In 21% of the cases changes were made without the respondent’s consent. “Other” responses accounted for 12% of the total (e.g., *Report got buried*).

A slight majority (53%) of the pressured respondents sought advice from someone not connected to the evaluation while the situation was unfolding, while the remainder (47%) did not ($n = 344$). Advice seekers ($n = 161$) were, by far, most likely to contact a colleague (47%), while over a quarter

Table 6. Suggested Preventive Actions ($n = 101$).

	Frequency	Percent
Develop fuller understanding of goals/purposes/roles	20	19.8
Better educate stakeholders in methodology/statistics	13	12.9
Have independent third party intervene	12	11.9
Collaborate more with stakeholders	11	10.9
Establish formal evaluation protocol	10	9.9
Discuss the possibility of negative findings at beginning	8	7.9
Emphasize evaluator's responsibility to stand by the data	8	7.9
Improve data collection/documentation	6	5.9
Other	13	12.9

Table 7. Content of Misrepresentation Request ($n = 340$).

	Frequency	Percent
Present findings more positively	130	38.2
Omit or downplay negative findings	85	25.0
Change language—neither positive nor negative	43	12.6
Use inappropriate methodology or statistical procedures	21	6.2
Draw different conclusions	20	5.9
Show inappropriate concern for implications of results	12	3.5
Use invalid or old data	11	3.2
Other	18	5.3

(28%) contacted representatives of more than one group (e.g., colleagues, mentors, supervisors, other professionals).

When asked for their satisfaction on an ethical level with how the pressuring episode was resolved,¹ respondents displayed wide variation on the 5-point Likert-type item (1 = *not at all satisfied*, 5 = *very satisfied*), including 20% who were not at all satisfied, 24% moderately satisfied, and 22% very satisfied ($M = 3.08$, $n = 335$).

Overall, less than a third (30%) of the pressured respondents believed that something could have been done to prevent the attempt to pressure them. Half (50%) felt that nothing could have been done, while 20% were unsure ($n = 342$). Those who thought something could have been done suggested a variety of possible preventive actions (Table 6). These include establishing a formal protocol at the beginning of the evaluation for how findings (and disputes over findings) would be dealt with (e.g., *A clearly defined, written and signed contract agreeing on the exact details of the evaluation including its focus and report format—there should be no midstream changes!*); informal discussion of such matters during the entry/contracting stage (e.g., *Discuss ahead of time what would happen if the results didn't turn out the way the organization was expecting*); better education of stakeholders in methodology and statistical procedures (e.g., *Educating our client as to how analyses are derived*); developing a fuller description and understanding of the evaluation's purpose, goals, and stakeholder roles (e.g., *Obtaining a clearer sense from the beginning of the evaluation who the intended audiences were for the final report*); and collaborating more thoroughly with stakeholders throughout the evaluation (e.g., *Better communication along the way. . . could have helped them to understand the value of reporting even negative findings to stakeholders and. . . that some of what they were upset about was common for their type of organization*). Occasionally, respondents presented strategies that seemed more reactive than preventive in nature; for example, having an independent third party intervene in some fashion (e.g., *There should have been a sort of ombudsman role somewhere in the process*).

A major goal of this study was to explore the *nature* of the misrepresentation that respondents felt pressured to engage in. The open-ended responses to this question were initially read by the second author who developed the content analysis categories presented in Table 7.

Both authors then independently reviewed every answer and assigned each one to a category. Disagreement occurred in less than 10% of the cases and was resolved through discussion. Results of the content analysis indicate that portraying findings as more *positive* than they actually were was the most frequent request (38%). Another 25% of the respondents stated that the pressure focused on inappropriately downplaying the evaluation's negative results. Thus, pressure that directly addressed the positive–negative evaluative dimension is cited in over 60% of the episodes.

Pressure involving the positive–negative dimension can take a variety of forms. In some cases it focuses on the *manner* in which findings are presented. For example, “The program director. . . encouraged me to put a more positive ‘spin’ on the information so the program would look better in the eyes of its funding agency.” Another respondent reported that he or she “was asked to present the findings in such a way that only those who read the report page-by-page would notice the negative findings.” And a third noted being “asked to highlight those findings which made the program look best (even though they were only mediocre in scale). I was asked to footnote those findings that did not show the program to be effective.”

In other cases, the positive/negative pressure involved the inclusion or deletion of findings, as in “I was directed by a superior to cut out parts of the report”; “Pressure was put on me to only put in positive findings and minimize the recommendations for changes”; and “Executive Director wanted to delete all negative information from report and not disclose it to anyone.” Finally, there are occasionally instances where the respondent perceives pressure to engage in outright falsification, such as “The director wanted us to report on more participants than the program had official records of,” “Try to ‘make up’ the results in order to look more positive,” and “Falsify attendance because of attendance requirements in the grant.”

Another type of pressure involved the language used in presenting findings, where the respondent did not explicitly indicate a focus on the positive–negative dimension (e.g., *Reword-ing of the outcome and presentation of findings; The director of my division. . . consistently sent the evaluation back with “adjustments” to the content, e.g., changes in wording and structure; Disagreements over the phrasing of results*).² There were also cases where respondents believed that they were encouraged to use methodological or statistical procedures that would generate misleading results (e.g., *Kept suggesting that I change my statistical test to find “other” outcomes; I was asked to re-run analyses and run different types of analyses that were not necessarily appropriate for the study/data set; Run different statistical tests beyond the contract’s pre-specified research questions in order to find positive outcomes for future grant fund-ing*). Closely related to this category were episodes where evaluators felt pressured to employ invalid or old data (e.g., *I was asked to report older findings that were more positive because the new ones indicated a bad trend; Asked me to use old data and present it as current/present findings*).

In some instances, the perceived influence attempt focused more on the conclusions presented by the evaluator than on the findings themselves (e.g., *Reverse the overall conclusions of the evaluation; Generalize the evaluation findings to programs which were not within the scope of the program; I was encouraged to draw a certain set of conclusions*).

Finally, stakeholders sometimes draw attention to the potential consequences of an evaluation in a way that makes the evaluator feel pressured to do less than the “right thing.” For example, “Indicated that if the results were what the client was looking for we would be offered a \$500,00 contract”; “We were told that ‘this isn’t what the evaluation should say because the evaluation report needs to show we are having the desired impact to support us getting another grant.’”

Table 8. Misrepresentation Perceptions by Type of Evaluation.

	External/internal		χ^2
	External	Internal	
Were changes made?			
No changes made	34	35	5.473*
Changes with no misrepresentation	51	39	
Misrepresentation	15	26	
<i>n</i> = 275			
Who made changes?			
I (we) made them	65	56	8.084**
Someone made them WITH my consent	19	12	
Someone made them WITHOUT my consent	16	32	
<i>n</i> = 197			
Incident preventable?			
Yes	27	34	6.599**
No	55	40	
Not sure	18	26	
<i>n</i> = 342			

Note. Table figures represent column percentages.

* $p \leq .065$. ** $p \leq .05$.

Table 9. Changes Made by Misrepresentation Perceptions.

	Request content			χ^2
	More positive/less negative	Other		
Changes made				
No changes made	30	40	6.48*	
Changes made with no misrepresentation	47	48		
Misrepresentation	22	11		
<i>n</i> = 271				
	Source of change			
	I (we) made them	Someone else made them WITH my consent	Someone else made them WITHOUT my consent	χ^2
Changes made with no misrepresentation	85	89	18	56.57**
Misrepresentation	15	11	82	
<i>n</i> = 159				

Note. Table figures represent column percentages.

* $p \leq .05$. ** $p \leq .001$.

Relationships Between Factors

The status of the evaluation as internal or external was related to several features of the pressuring episode. In external evaluations, pressure was more likely to come from individuals who hired or assigned the evaluator (74%) than was the case in internal evaluations (58%), $\chi^2(1, n = 368) = 9.49, p < .002$. In a finding that approached significance ($p < .07$), internal evaluators were more likely than external ones to believe that the pressuring episode resulted in changes in the evaluation report that constituted misrepresentation. Consistent with this result, internal evaluators were more likely than external ones to state that changes were made in the

Table 10. Misrepresentation Perceptions by Advice Seeking ($n = 340\text{--}342$).

	Sought advice?		χ^2
	Yes	No	
Misrepresentation request			
More positive/less negative	41	59	7.529*
Other	57	43	
Incident preventable?			
Yes	61	39	6.988*
No	38	62	
Not sure	48	52	

Note. Table figures represent row percentages.

* $p \leq .01$.

Table 11. Misrepresentation Perceptions by Ethical Satisfaction With Episode Outcome.

	Mean ethical satisfaction ^a	F
Intent		
Individual KNEW he or she was asking for misrepresentation	2.88	4.496*
Individual did NOT believe he or she was asking for misrepresentation	3.39	
Not sure	2.96	
$n = 334$		
Source of change		
I (we) made them	3.28	38.54**
Someone else made them WITH my consent	2.82	
Someone else made them WITHOUT my consent	1.63	
$n = 200$		
Incident preventable?		
Yes	2.77	7.12**
No	3.37	
Not sure	2.84	
$n = 335$		

^a From an ethical perspective, how satisfied were you with how this incident was resolved? (1 = not at all satisfied, 5 = very satisfied).

* $p \leq .05$. ** $p \leq .001$.

evaluation report without their consent. Not surprisingly, given these findings, internal evaluators were less ethically satisfied than external evaluators with how the pressuring episode was resolved. ($M_s = 2.67$ vs. 3.26 , respectively, $p < .001$, $n = 335$). Internal evaluators, however, were more likely than external ones to believe that something could have been done to prevent the pressuring episode (Table 8).

Pressuring incidents resulting in changes that constituted misrepresentation in the eyes of respondents were related to several variables in the study. These episodes were more likely to involve pressure that focused on the positive/negative dimension of the findings. Misrepresentation was also more likely when respondents indicated that changes were made without their consent (Table 9). Finally, those who perceived misrepresentation were less ethically satisfied with how the incident was resolved ($M = 1.59$) than those who reported no changes in the report ($M = 3.82$) or changes that did not involve misrepresentation ($M = 3.40$, $F = 62.55$, $p \leq .01$, $n = 269$).

Seeking advice during the incident was also linked to several variables. Females were twice as likely to solicit advice than males (56% vs. 28%), $\chi^2(1, n = 337) = 22.12$, $p = .000$. When the episode involved

the positive/negative dimension of the findings, respondents were less likely to seek advice than when the pressure focused on other dimensions. Finally, those who sought advice were more likely than other respondents to believe that the pressuring incident could have been prevented (Table 10).

Ethical satisfaction with the episode's outcome was related in predictable ways to two variables and in a less predictable fashion to a third. When respondents believed that the pressuring individual was *knowingly* asking for deception, the level of satisfaction was lower. Moreover, respondents were most satisfied when they were the ones who made changes in the report as a result of the episode, and least satisfied when someone else made changes without their consent. A finding that may appear counterintuitive at first glance is that respondents who believed the pressuring incident could *not* have been prevented report greater ethical satisfaction with the outcome than those who saw the incident as avoidable (Table 11).

Discussion

This study builds on previous investigations (e.g., Buchanan & MacDonald, 2011; Morris & Cohn, 1993; Turner, 2003) in demonstrating that being pressured to misrepresent findings is a common occurrence in evaluation. Over 40% of the respondents reported experiencing such pressure, and a majority of that subgroup indicated that they had encountered it on more than one occasion. Most often, and especially in the case of external evaluations, the pressure came from the individual responsible for hiring the evaluator or assigning him or her to the project. Given the political power inherent in the hiring/assigning role, these influence attempts can represent significant forces in an evaluation. Making the situation even more challenging is the frequent perception by evaluators that a deliberate intent to deceive underlies stakeholder efforts to influence the presentation of findings.

Although it is true that misrepresentation pressure most often focuses on how positive or negative the findings are, in many cases (nearly 40% in this study) the issue takes a different form, at least nominally. The dispute may involve language/wording, statistics/methodology, or the nature of the evaluator's conclusions, for example. Interestingly, perceptions of the stakeholder's *intent* is not related to the content of the request; that is, stakeholders who press for changes on the positive/negative dimension are no more likely to be viewed as "acting in bad faith" (i.e., believing they are asking for misrepresentation) than those who lobby in other domains. This finding suggests a glass half full, or half empty, depending on one's perspective. Benign motivation can be seen as underlying just about any type of pressuring request, offering hope for a positive resolution of the conflict, but so can malignant intent, which is likely to generate more pessimistic expectations.

It is certainly the case that some pressuring incidents end badly. In 15% of the episodes, respondents report that misrepresentation did in fact occur, and in another 3% respondents indicated that they resigned from or were fired from the evaluation, or the project was terminated. Not surprisingly, misrepresentation was more likely to occur when changes in the report were made without the respondent's consent, which took place in 21% of all the episodes in which changes were made. Given these results, it is understandable that 20% of the respondents gave the lowest possible rating (*not at all satisfied*) when registering their satisfaction with how the episode was resolved in ethical terms.

On the other hand, in nearly 70% of the episodes either no changes were made in the disputed section of the report, or the changes that were made were not viewed by the respondent as constituting misrepresentation. In these circumstances, respondents were more likely to be ethically satisfied with the episode's outcome. Indeed, nearly two thirds of the respondents (65%) were at least moderately satisfied, on an ethical level, with how the pressuring incident turned out. Thus, as distressing as pressuring episodes may be to evaluators, the evidence indicates that negative outcomes are by no means inevitable. In fact, the data suggest that, more often than not, they are handled in a fashion that produces a reasonably satisfactory, if not ideal, result when viewed from the vantage point of professional ethics.

Of course, to an undetermined extent rationalization processes might have contributed to this optimistic finding. Some respondents may have succumbed to a temptation to interpret the outcome of their pressuring episode in relatively positive terms in order to protect their self-image as ethical practitioners (i.e., *I would never have been a party to a violation of the standards of my profession*). The variation in reported ethical satisfaction found in the study, however, suggests that a motive to engage in this “self-serving bias” did not uniformly trump other considerations when respondents reflected on their experiences.

The results also indicate that misrepresentation pressure applied to internal evaluators generates more perceived ethical damage than cases involving external evaluators. In the former, changes made to reports without the consent of the evaluator are more likely to occur, and the changes made are more likely to constitute misrepresentation in the evaluator’s eyes. Overall, these episodes lead to less ethical satisfaction with how the incident was resolved. Given these findings, more attention to the ethical challenges faced by internal evaluators, and how to prevent and/or handle such challenges, would appear to be sorely needed.

It is noteworthy that respondents who thought the pressuring episode was not preventable were more ethically satisfied with the episode’s outcome than those who saw the incident as avoidable. Attribution processes may be playing a role in these cases. To the extent that respondents believed that they failed to engage in actions that could have prevented the pressuring episode, they may have been more likely to hold themselves at least partially responsible for any negative ethical consequences of the incident. Insofar as ethical satisfaction is a reflection of, among other things, one’s own perceived role in events, these dynamics would lead to lower satisfaction than would be found in cases where respondents felt there was nothing they could have done to prevent misrepresentation pressure.

The proportion of pressuring incidents that end positively, or not occur at all, might well be increased if more evaluators followed the advice offered by respondents who claimed that steps could have been taken to preempt the influence attempt. These recommendations were heavily weighted toward the entry/contracting stage of the evaluation and are consistent with the general emphasis in the organizational consulting field on the crucial importance of this phase (e.g., Block, 2011). When viewed as a whole, respondents’ suggestions indicate that a comprehensive approach to preventing and responding to misrepresentation pressure would involve the following components:

- During the entry/contracting stage, engage key stakeholders in a meaningful and frank discussion of misrepresentation pressure. Describe the various forms that this pressure can take, and the role that professional principles, standards, and guidelines should play in framing an evaluator’s response to this issue. Invite stakeholders to share their perspectives on how to structure the evaluation process so that threats to the integrity of the findings are minimized, while allowing for differences of opinion to be aired in a collaborative environment. As one respondent puts it, this would entail “a discussion about the real possibility of negative findings, my responsibility to report them, how the client can use them to improve the program, and how the funder and other stakeholders might react and how their reaction can be managed.” Although this conversation might be a difficult one, most evaluators would probably agree that it is better to have a moderately challenging interaction early in the relationship than a much more awkward and painful one later on.
- In some cases, it may be appropriate to formalize the understanding that has been reached in the previous discussion by means of a written protocol. Such a step would guard against the disputes that conflicting memories of a shared conversation can produce, especially concerning such a sensitive issue as the integrity of evaluation findings. Indeed, when contemplating “what might have been” in his or her misrepresentation episode, one respondent asserted that “something in the contract should have stipulated that as a contractor we have the right to document the results based on our standards of professional integrity and our expertise and technical skill.”

- Discussions of misrepresentation are likely to be most productive when they take place within the context of a broader exploration by evaluators and stakeholders of the goals, roles, and boundaries that should guide the evaluation. Although one might assume that such an exploration would occur as a matter of course in any evaluation, it appears that this is not necessarily the case, especially in evaluations involving misrepresentation pressure. As one respondent vividly puts it:

Clear boundaries needed to be set from the very beginning of the funding period such that the client understood that while the evaluators were being paid out of the grant, it did not mean that the client “owned” the evaluators. The evaluators and client needed to have an understanding from the start that the client and evaluator need to work together respectfully, which means that each must do their part and the client should not expect the evaluator to be at their beck and call.

- Once the evaluation is under way, keep the stakeholders informed of emerging findings and educated, to the greatest extent possible, about the methods and statistical procedures underlying those findings. In the straightforward words of one respondent, “As an evaluator I should have actively engaged the client through providing report-backs on activities.” Once again, this represents standard advice for evaluators, but it can be easy to forget—or not implemented due to avoidance—when findings are unlikely to be welcomed by stakeholders.
- Taking the preceding steps promises to reduce the incidence of misrepresentation pressure but will not eliminate it entirely. When such pressure does occur, revisiting with stakeholders the formal and/or informal understandings that had been negotiated during the entry/contracting phase probably represents the most appropriate initial response. Seeking advice from colleagues can also be useful, and many evaluators in the study did exactly that. But what if the conflict remains unresolved? Although respondents were not asked to address this question, some offered comments that were relevant to it. One recommended approach was simply not to give in to the influence attempt, even if the consequences would have been negative for the evaluator. For example, a respondent noted that “I could have just resisted the pressure. Maybe that would have cost me my contract with them and the possibility of future contracts.” Another maintained that “I should have resisted more and as an organization we should have bitten the bullet.” In other cases respondents recommended interventions by a third party (e.g., *Intervention from a corporate officer above me who could explain the structure and rationale behind objective third-party evaluations without jeopardizing my ongoing working relationship with the client*).

The “simply resist” strategy highlights an issue that often goes unaddressed in discussions of ethical challenges in evaluation: the occasional need for moral courage. Acts of moral courage encompass three components: a commitment to ethical principles, an awareness of the personal risk associated with displaying that commitment, and a willingness to take that risk (Kidder, 2005). Sometimes, when even one’s most diligent efforts to prevent and/or respond to misrepresentation pressure have failed, and common ground between evaluator and stakeholder/stakeholders cannot be found, the only reason for “doing the right thing” is that, indeed, it *is* the right thing to do.

Limitations

Although the study’s 37% response rate raises concerns about the generalizability of its findings, the respondents’ demographics are similar on a number of key dimensions to those obtained in an AEA-sponsored survey of the Association’s membership in 2007–2008 (response rate of 49%; American Evaluation Association, 2008). For example, 68% of the current study’s respondents were female, while the percentage of female respondents in the AEA membership scan was 67%. The percentages of the current sample holding a doctoral, master’s, or bachelor’s degree as their highest educational

credential were 58%, 37%, and 5%, respectively, while the corresponding figures for AEA's survey were 52%, 42%, and 7%. In terms of employment, 33% of the current sample worked in a college or university (vs. 29% in the AEA sample), 29% in private firms or consulting positions (vs. 35%), 12% in a federal, state, or local agency (vs. 12%), and 22% in a nonprofit organization (vs. 7%). Finally, with respect to evaluation experience, 34% of the current sample had over 15 years of experience, while only 27% did in the AEA membership scan. In contrast, only 22% of the current respondents had less than 6 years of experience, while 33% of the AEA sample did.

Overall, these comparisons suggest that senior evaluators, and those who work in nonprofit settings, may be overrepresented among the current study's respondents. Given that evaluation experience is associated with encountering misrepresentation pressure, this study probably uncovered more reports of such pressure than would have been the case with a less experienced, but more representative, sample.

It is also possible that evaluators who have encountered misrepresentation pressure are more motivated than the nonpressured to respond to a survey such as this one, resulting in an inflated estimate of the incidence of misrepresentation pressure among evaluators in general. Unfortunately, there are no published studies available with a comparable methodology and a higher response rate that can shed direct light on this possibility. What the research in this general domain does clearly indicate, however, is that misrepresentation pressure is one of the most frequent ethical challenges reported by evaluators, compared with other types of ethical problems they experience (e.g., Buchanan & MacDonald, 2011; Morris & Cohn, 1993; Turner, 2003). Thus, even if 42% exaggerates, to an unknown degree, the absolute percentage of all evaluators who have faced misrepresentation pressure, there is little reason to believe it exaggerates the relative importance of this challenge in evaluators' professional lives.

A third limitation is that it is only *evaluators'* perceptions of misrepresentation that are reported in this study. In particular, we do not know how other stakeholders viewed the pressuring episodes described by respondents. Understandably, their accounts could differ significantly from those offered by our sample, a possibility that is consistent with the results of research on the evaluation experiences of program officers at philanthropic foundations (e.g., Morris, 2007). More research is clearly needed here in order to generate a comprehensive portrait of misrepresentation in evaluation.

Conclusion

At this point in the history of the field, it is clear that being pressured by powerful stakeholders to misrepresent findings is part of the landscape of ethical challenges in evaluation. It is an occupational hazard that many, if not most, evaluators will encounter. These troubling episodes need not routinely result in dismal outcomes, however. This study suggests that evaluators do a reasonable job of upholding the AEA Guiding Principle of Integrity/Honesty in responding to misrepresentation challenges. If more attention were paid to the preventive strategies recommended by respondents, both new and experienced evaluators might succeed in reducing the incidence of such attempts.

Authors' Note

This is a revised version of an article originally presented at the annual meeting of the American Evaluation Association, November 12, 2009, in Orlando, FL.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Specifically, the item asked, “From an ethical perspective, how satisfied were you with how this incident was resolved? In this context, ‘ethical’ refers to ‘conforming to accepted professional standards of conduct.’”
2. It is likely that some answers coded as a “language/wording conflict” referred to a disagreement that involved a positive/negative dimension, but in these instances the respondent’s open-ended description was not detailed enough to warrant such an interpretation for coding purposes.

References

- American Evaluation Association. (2004). *Guiding principles for evaluators* (rev.). Retrieved August 31, 2011, from <http://www.eval.org/Publications/GuidingPrinciples.asp>
- American Evaluation Association. (2008, April). *AEA internal scan report to the membership*, by Goodman Research Group. Retrieved September 19, 2011, from www.eval.org
- Block, P. (2011). *Flawless consulting: A guide to getting your expertise used* (3rd ed.). San Francisco, CA: Jossey-Bass/Pfeiffer.
- Buchanan, H., & MacDonald, W. (2012, November). *Anytime, anywhere, evaluation ethics matter!* Presentation at the meeting of the American Evaluation Association, Anaheim, CA.
- Kidder, R. M. (2005). *Moral courage*. New York, NY: William Morrow.
- Morris, M. (2007). Foundation officers, evaluation, and ethical problems: A pilot investigation. *Evaluation and Program Planning*, 30, 410–415.
- Morris, M., & Cohn, R. (1993). Program evaluators and ethical challenges: A national survey. *Evaluation Review*, 17, 621–642.
- Turner, D. (2003). *Evaluation ethics and quality: Results of a survey of an Australasian Evaluation Society members*. Retrieved August 11, 2011, from http://www.aes.asn.au/about/Documents%20ongoing/ethics_survey_summary.pdf
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users*. Los Angeles, CA: Sage.