

Article

Don't you forget about me: Independence of evaluations from the perspective of US evaluation clients—An exploratory study

Evaluation

2023, Vol. 29(1) 110–132

© The Author(s) 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/113563890221138876

journals.sagepub.com/home/evi**Fabienne Helen Schmidli** 

Zurich University of Applied Sciences (ZHAW), Switzerland

Lyn Ellen Pleger

Zurich University of Applied Sciences (ZHAW), Switzerland

Susanne Hadorn

University of Bern, Switzerland

Abstract

Policy evaluation literature has stressed the importance of *independence* of evaluations to guarantee objective evidence collection. The evaluator–client relationship is critical in this respect, since it contains inherent tensions due to the necessity for independent assessments alongside requirements for increased responsiveness to clients' interests. Despite this distinct relationship, the client perspective has only recently received attention in research. This article presents findings from a survey among US evaluation clients and compares these to existing evidence from Switzerland. Unlike previous studies, we distinguish between constructive and destructive client influences. We show that professional experience and client familiarity with evaluation standards increase the likelihood of constructive influences aimed at improving evaluation results. Nevertheless, the findings indicate that dissatisfaction with an evaluation increases client's attempts at influence that may be destructive. By discussing both motives behind influence and potential preventive measures, this article seeks to contribute to the increased social impact of policy evaluations.

Keywords

evaluation clients, evaluation ethics, evidence-based policy making, independence, US evaluation practice

Corresponding author:

Fabienne Helen Schmidli, Center for Arts Management, School of Management and Law, Zurich University of Applied Sciences (ZHAW), Bahnhofplatz 12, CH-8401 Winterthur, Switzerland.

Email: fabienne.schmidli@zhaw.ch

Introduction

Scientific independence is regarded as an essential prerequisite for evaluations (Mayne, 2017; Perrin, 2018; Pleger et al., 2016: 1; Pleger and Hadorn, 2018). At the same time, evaluations take place within a political context (Leeuw, 2012) in which clients of evaluations often pursue a specific interest concerning the evaluation results (Barnett and Camfield, 2016: 528). Evaluators are thus often confronted with the conflicting requirements of ensuring scientific independence while meeting the expectations of the client (Pleger et al., 2016: 12). To guarantee independent and objective evaluation processes, it is essential that the client does not exert any influence on the evaluation process that violates scientific standards. However, empirical studies have repeatedly shown that evaluators often feel under pressure and that the most influential stakeholder is the client (Morris and Clark, 2012; Pleger et al., 2016; Stockmann et al., 2011; Turner, 2003). Pressure on evaluators can take many forms and might jeopardize scientific independence (Pleger et al., 2016: 11) and “ethical challenges can arise at virtually any point in a project, from the entry/contracting stage [. . .] to the [. . .] utilization of results” (Morris and Clark, 2012: 57).

However, influence is not necessarily destructive, but can also be constructive (Perrin, 2018; Pleger and Hadorn, 2018). In this respect, motives, intentions, and consequences of the influencing clients are decisive. Hence, in the understanding of this article, independence can still be high in evaluation contexts where all relevant stakeholders are included in the evaluation process at various stages. For instance, it is usual for evaluators to request data from clients and evaluated persons and provide an opportunity to give feedback on results from independent evaluations. Importantly, evaluators then interpret the data without clients (or other stakeholders) pushing for a specific interpretation or outcome. Independence is only jeopardized when such inputs not only lead to a better understanding of the subject under evaluation or a correction of factual errors in evaluation reports but also distort interpretations.

To our knowledge, the view of clients on the influence on evaluators has so far only been examined empirically in one study, whereas the others investigated evaluator perspectives (e.g. Barnett and Camfield, 2016; Berriet-Sollicet et al., 2014; Desautels and Jacob, 2012; Morris and Clark, 2012; Newman et al., 1995). Pleger and Hadorn (2018) analyzed the perceptions of Swiss evaluation clients on evaluation independence by means of a survey. The results show that the clients surveyed were rarely addressed by evaluators about the pressure they exert, even though the relationship between evaluators and clients is often characterized by conflict (Pleger and Hadorn, 2018). To broaden the understanding of the perception of evaluation clients and gain new insights into the important area of independence from evaluations, this study presents findings from an online survey among US evaluation clients. By applying the BUSD model by Pleger and Sager (2018), this study nuances our understanding of client influence on evaluators by distinguishing between destructive and constructive forms of influence. The aim of the present study is threefold: First, it aims to provide first insights on the evaluation client perspective in the evaluation environment in the United States to broaden our understanding of the causes and forms of pressure on evaluators. Second, we examine the validity and transferability of the findings from the Swiss study to the US context, to identify similarities and differences between the two countries' evaluation environments. Third, the present study adopts an innovative approach by discussing different forms of influence.

The article consists of several sections and is structured as follows. After the introduction, we summarize the current body of literature on the independence of evaluations and introduce

the theoretical framework, including the BUSD model (section “Pressure on evaluators and independence of evaluation”), followed by the study design in section “Survey design.” The results are discussed in section “Results” before some concluding remarks in the final section.

Pressure on evaluators and independence of evaluation

As societal problems become more complex and governments are increasingly challenged to find effective policy solutions, the notion of evidence-based policymaking (EBP) has come to the forefront of public policy research and practice. Continuously collecting facts and learning from what has and has not worked previously is crucial for designing measures that, when implemented, actually lead to the desired outcome (Head, 2008). Policy evaluations play a crucial role in this development toward evidence-based policies since they are one of the main instruments through which evidence is collected, and thus provide the factual basis for planning public action (Berriet-Sollicet et al., 2014). Discussions about the *independence* of evaluations are rooted in the fact that, despite the use of scientific methods, evaluations take place for the most part in a political, and thus not neutral, context (Desautels and Jacob, 2012), in which evaluation results might benefit one party and harm the other. Unsurprisingly, scholars and practitioners are concerned with the issue that evaluations are subject to potential influence (as a consequence of the potential “losers” and “winners” in the process) (see, for example, Morris, 2007; Morris and Cohn, 1993; Turner, 2003).

Evaluation independence encompasses a variety of aspects, a key feature of which is evaluation governance, for example, in the form of evaluation protocols or codes of conduct. Evaluation governance sets the norms and standards by which evaluations are conducted and is thus a key aspect of independence. In practice, these codes of conduct are usually formalized in the form of evaluation standards. The normative principle of integrity of the American Evaluation Association (AEA) describes the preservation of independence within the framework of professional, ethical conduct by emphasizing the importance of evaluators behaving transparently and honestly to guarantee the integrity of the evaluation (AEA, 2021b: 3). Independence is also set down as a cornerstone in the Organisation for Economic Cooperation and Development’s (OECD, 2010) “Quality Standards for Development Evaluation” with the goal of ensuring quality for the evaluation process and the evaluation product and is an integral part of various other national evaluation standards such as the Swiss (Swiss Evaluation Society, 2022), German (DeGEval, 2022), Australian (Australian Evaluation Society (AES), 2022), Canadian (Canadian Evaluation Society (CES), 2022), or UK Evaluation Society (2022).

Until recently, research on the independence of evaluations has focused on the role of evaluators and their reactions to pressure (Morris and Clark, 2012; Stockmann et al., 2011; LSE GV314 Group and Page, 2014). As research in this area progressed, it became increasingly evident that the main source of influence is the evaluation client (Morris and Clark, 2012; Pleger et al., 2016; Stockmann et al., 2011; Turner, 2003). This is rooted in the power and position that enables the client to influence or even determine individual aspects of the evaluation process, such as the scope, methodology, and resources, but also individual results (Barnett and Camfield, 2016: 532).

Evaluations have inherent dependencies that are different to those in “regular” commissioned research. In a nutshell, it is the core mission of evaluators to generate objective evidence and make assessments of policies, which often also include criticisms that the clients may disagree

with. It can even be argued that governments (often acting as clients) want to avoid bad (or even seek good) publicity when commissioning evaluations (LSE GV314 Group and Page, 2014). At the same time, it is in the interest of the evaluators that the clients are satisfied with the evaluation, in particular, because the former are usually dependent on future contracts, offering “career benefits, access to research material, and the prestige that comes of having one’s research credentials endorsed by government” (LSE GV314 Group and Page, 2014: 224). Overall, Smith (1998) has long ago argued that “increasing client demands, shifting evaluator roles and methodological innovations have all influenced evaluators to become increasingly responsive to client interests” (p. 117). This may lead to situations where contracting principals might appear as clients (according to the maxim “the customer is always right”) as opposed to neutral commissioners. Such dynamics are clearly a danger to evaluation independence and can be detrimental to the quality of the evidence generated. At the same time, however, addressing the needs of clients (and other stakeholders) can make evaluation results more user-centered and thus carry more weight in subsequent decision-making (see, for example, Patton, 1997). Evaluators must therefore find ways to deal with (inappropriate) influence from their clients in order to be able to produce high-quality evaluations, without producing evidence that will not be utilized. Given these specific challenges, there is a clear terminological distinction between *commissioner* and *client*.¹ In this article, however, the two terms have deliberately been used as synonyms for the sake of simplicity (such as, for example, Azzam et al., 2021; Gullickson, 2020), whereby the aforementioned problematic dependencies are always considered in the analysis.

Evaluators are thus faced with the, sometimes conflicting, priorities of meeting the expectations of the clients and at the same time maintaining the independence of evaluations. The independence of evaluations is at risk as soon as clients exert pressure on the evaluators, such as to change or misinterpret certain evaluation results (Pleger et al., 2016: 12). To a certain extent, this area of tension lies in the nature of evaluations. Evaluations habitually take place in an organizational and political context where stakeholders in evaluations have a specific interest in the results of the evaluation (Barnett and Camfield, 2016: 528). More precisely, evaluations are usually commissioned directly by third parties such as civil servants or policy-makers with a specific interest within the political context (Pleger and Hadorn, 2018: 2). As a result, the evaluation process, despite its scientific basis, is still prone to distortion and can never be completely neutral (Desautels and Jacob, 2012: 437; Eliadis et al., 2011). Besides knowing stakeholder potential to influence evaluations (Eckhard and Jankauskas, 2019), it is furthermore desirable and legitimate to identify the valence of influence to optimize evaluation quality and, as a result, to preserve the independence of evaluations.

To better understand the drivers behind the exertion of influence, Pleger and Hadorn (2018) surveyed Swiss evaluation clients. The study thus examines the general client perspective for the first time and, due to the similarity of the study design and the fact that several US studies have examined the evaluators’ perspective (see, for example, Morris and Clark, 2012; Morris and Cohn, 1993; Perrin, 2018), serves as a suitable basis for a first-time country comparison between the United States and Switzerland focusing on the client side.

Their findings show that clients are often unaware that evaluators feel pressured to change evaluation results. Communication problems as well as tension between evaluator responsibility to be neutral and the clients’ desire to provide input all seem to be part of standard evaluation practice. Much of the literature on the independence of evaluations is based on the presumption that clients intervening to influence the evaluation and evaluators responding to

		Explicitness of influence	
		<i>Explicit, direct</i>	<i>Implicit, indirect</i>
Direction of influence	<i>Destructive</i>	<p><i>Distortion</i> Negative, destructive pressure</p>	<p><i>Undermining</i> Negative persuasion/counsel</p>
	<i>Constructive</i>	<p><i>Betterment</i> Positive, constructive pressure</p>	<p><i>Support</i> Positive persuasion/counsel</p>

Figure 1. Reprinted from Lyn Pleger and Fritz Sager (2018), Copyright (2021), with permission from Elsevier.

these demands are harmful to the evaluation outcome since they distort the independent analytical process of the evaluators. This article applies the BUSD model (Figure 1) of Pleger and Sager (2018) to the collected data on US client perspectives to comply with the assumption that influence can indeed be destructive, but that in other cases, it can also be constructive. Specifically, the motives behind client influence can be manifold and may have either a positive or negative impact on the quality of the evaluation results. The BUSD model specifies this argument and proposes two dimensions along which the influencing actions of clients can be categorized. The model distinguishes between destructive and constructive influences (i.e. the *direction of influence*), which can be exerted explicitly (directly) or implicitly (indirectly) (i.e. the *explicitness of influence*).

We argue that categorizing client influences on evaluation processes helps to understand better the realities observed in evaluation practice. Rather than looking at independence in a unidirectional way, differentiating between different forms of influence helps to nuance our understanding of the potential consequences of influence and clarifies the need for action to improve the quality of evidence provided by evaluations. Overall, according to the BUSD model, a total of four forms of influence are distinguished: *Distortion* and *Undermining* are destructive influence types, in contrast to *Betterment* and *Support*, which represent constructive influence types (Pleger and Sager, 2018: 169–170). The model takes these ambivalent characteristics into account and allows a clear distinction between positive and negative influences, while the terms “influence” and “pressure” are treated as functional equivalents, both of which can be understood as an ethical challenge during the evaluation process (Pleger and Sager, 2018: 167).

To help evaluators distinguish between constructive and destructive influences, three differentiators, *Awareness*, *Intention*, and *Accordance*, were developed (Pleger and Sager, 2018: 170). The differentiators—hierarchically ordered and built on each other—serve to determine the valence of an influence through a specific question. *Awareness* asks whether the influencing party is aware of his or her actions; *Intention* asks whether the consequences decrease the quality of the evaluation; and *Accordance* asks whether the consequences violate scientific

standards. The more questions answered in the affirmative, the more destructive the influence. Conversely, if all three questions can be answered in the negative, the influence is constructive. *Accordance* has the strongest effect on the intention to influence (Pleger and Sager, 2018: 171). To capture the theoretical premises outlined above, we included the BUSD model in the survey as explained below. In the following section, we present the survey design.

Survey design

Data were collected in April and May 2019 and consisted of a six-part email distribution via Qualtrics² survey software. For the recruitment of respondents, a total of 55,683 email addresses of potential evaluation clients were collected (54% public sector, 35% universities and colleges, 8% non-governmental organization (NGO), 4% private sector). In addition, the online survey was sent to 1000 AEA members by email. Despite the high number of emails sent and considerable efforts to increase the response rate, the number of respondents remained very low, especially in relation to the email addresses targeted. Nevertheless, due to the exploratory nature of this study, even the small sample size, which is not representative, can contribute to a gain in knowledge as it gives first-time insights on how the landscape of evaluation commissioners is composed in the United States. Based on this explorative approach, future research can build on our findings by capturing and describing the client side more systematically and in greater detail, and by including other influencing factors (e.g. governance) in further analyses. Two hundred fifty-three respondents were recruited, of which 189 did not meet the criteria³ required for the survey. The final data set consisted of 62 individual cases. The online survey consisted of 55 questions largely corresponding with those of the Swiss survey in Pleger and Hadorn (2018) to ensure comparability of the Swiss and US data. Additional questions were included in the US survey to capture the distinction between constructive and destructive influences.

To operationalize influence, questions were derived from the theoretical BUSD model along the “direction of influence” and “explicitness of influence” dimensions. Here, the operationalized forms of influence refer to the whole evaluation process instead of focusing on individual stages, such as data collection or communication of the results (Pleger and Sager, 2018). This understanding is based on the finding that influence occurs at every stage of the evaluation phase, that is, from the contracting phase of the evaluation to the use of the utilization of results (Morris, 2007, 2015). For the operationalization of the constructs “destructive influence,” the four-level scale of the Swiss study was adapted (Pleger and Hadorn, 2018), and the construct was operationalized based on the BUSD model, measured with eight items. For the operationalization of the “constructive influence” construct, a new four-level scale was formed theoretically based on the BUSD model. The answers for both scales could be classified according to four frequency categories.⁴ Based on the scale, an additive index with a value range from 0 to 24 was formed, with these in turn being assigned to the four frequency categories. The degree of homogeneity of the “destructive influence” scale is in a low range and could be increased to the alpha value of .635 by excluding one item. Given the broad construct, this value is accepted in the present case. The degree of homogeneity of the scale of the “constructive influence” type lies in a very good range, with a Cronbach’s alpha of .842. Derived theoretically from the BUSD model, an additional item was created for each of the four influence forms “Betterment,” “Support,” “Undermining,” and “Distortion”: The item of the direct, positive influence form “Betterment” measures the proactive illustration of

improvement potential without distorting the results, and the indirect, implicit influence form “Support” the general commitment on the part of the client to optimize the evaluation quality. The item of the direct, negative influence form “Distortion” measures the possible request that individual evaluation results be shortened (i.e. a reduced relevance of those results), and the indirect, implicit influence form “Undermining” that there may be a delay in the evaluation process due to the proposed changes. Answers could be given through the four frequency categories mentioned.

For the operationalization of the construct “dissatisfaction,” not only the reasons were measured but also the extent of dissatisfaction with regard to seven different aspects of an evaluation (items). An additive index was formed, and the degree of homogeneity of the scale is very high, with a Cronbach’s alpha value of .853, which could only be improved by excluding the item “Keeping the schedule” (Cronbach’s alpha: .916). The “conflict relationship” construct was operationalized as a variable with different values, based on an 11-step scale capturing how conflict-prone clients generally perceive the relationship to evaluators in their work. The “familiarity” construct with evaluation standards was operationalized via the four-level scale adopted from the Swiss study (Pleger and Hadorn, 2018). Client “work experience” is measured by the variable “professional experience,” which is an additive index composed of the variables “evaluation years” and “number of evaluations.” The degree of homogeneity of the scale is very low, with a Cronbach’s alpha of .550. Due to the short scale, no items can be excluded, and the value is accepted accordingly.

Results

This section first briefly presents the exploratory findings of the individual characteristics of US clients in Table 1 and their use of different forms of constructive and destructive influence (self-reported data), followed by a presentation of the findings concerning the perceived relationship between US clients and evaluators, and finally compares the exploratory findings with the Swiss data. In order to allow comparability between different findings, the results were deliberately expressed using percentages and numbers.

Individual characteristics of US clients

Table 1 compares the findings concerning characteristics of US and Swiss clients relating to gender, years of commissioning evaluations, number of evaluations commissioned in the past, and sector affiliation.

In the US sample, there is a slight overrepresentation of female survey participants, whereas in the Swiss sample, men are more strongly represented. The two samples are very similar in terms of the number of years of commissioning evaluations. While Swiss clients come almost exclusively from the public sector, the public sector also dominates for US clients, followed by NGOs and the higher education sector. It should be noted that the sample is not representative and may be biased due to the relatively higher representation of clients that have commissioned only five or fewer evaluations in the past. Therefore, the results are not representative, as the answers are not representative of the general client landscape but refer to the clients who participated in the survey. In the following, individual characteristics of the clients are considered regarding their perceived constructive and destructive influence in the evaluation process and their dissatisfaction with the evaluation process and results.

Table 1. Client characteristics in the United States and Switzerland.

Characteristic	US clients	Swiss clients
Gender	39% male, 61% female (N=31)	76% male, 24% female (N=38)
Education ^a	3% college, 6.5% bachelor's degree, 52% master's degree, 6.5% professional degree, 32% doctorate (N=31)	3% no education, 16% technical college degree, 47% university degree, 32% doctorate, 3% habilitation (N=38)
Years of commissioning evaluations ^a	11% 0 or less than 1 year, 46% 1–5 years, 27% 6–10 years, 5% 11–15 years, 11% more than 15 years (N=56)	M=9 years (N=53)
Number of evaluations commissioned in the past	39% 1–5 evaluations, 37% 6–20 evaluations, 24% more than 20 evaluations (N=54)	39% 1–5 evaluations, 44% 6–20 evaluations, 17% more than 20 evaluations (N=52)
Sector affiliation	40% public sector; 27% NGOs; 13.5% university or college; 8% private sector; 11.5% other sector (N=52)	92% public sector; 8% other sectors (N=52)

NGOs: non-governmental organizations.

Note: Data for Switzerland were taken from Pleger and Hadorn (2018: 6). Education was measured with different levels in both studies. In the Swiss study, years of commissioning evaluations were measured differently with an open question directly asking for the specific number of years. Hence, data for these variables are not directly comparable across samples. N=total number of all valid cases. N varies in the following tables due to the questionnaire design, which did not force responses, and thus respondents' answers were not mandatory.

Constructive and destructive influence

Clients were asked about the relevance of different forms of influence in their evaluation practice. The different perceived influencing actions presented in Tables 2 and 3 have been categorized as either constructive or destructive influences. This allocation is based on the principles of the BUSD model by Pleger and Sager (2018). However, we are aware that certain actions categorized as destructive (e.g. requesting a more positive presentation of the evaluation results; see Table 3) can, in some cases, also be considered constructive if they help to eliminate factual errors in the evaluation reports.

As presented in Table 2, all clients have exercised *constructive influence* on the evaluation process at least once, with as many as 61 percent of respondents stating that they have done so many times before (N=23). Sample sizes vary in the following tables due to the questionnaire design, which did not force responses, and thus respondent answers were not mandatory.

All clients indicated that they have supported the work of the evaluators by providing all information relevant to the evaluation at an early stage more than once and that they had already held a constructive dialogue with the evaluators at least once on how to improve the quality of the evaluation. The statements that conclusions were discussed neutrally with evaluators, that ideas for improving the quality of the evaluation were exchanged, that evaluation methods were discussed to improve the quality of the evaluation, and that the results could be presented to improve the understanding of the target group were answered positively by all respondents except one. Interestingly, four clients (15%) indicated that they have never proactively illustrated improvement potential without distorting the evaluation results. The construct *Betterment* (i.e. the direct, constructive influence type), which measured the proactive illustration of improvement potential without distorting the results, indicates that 93 percent

Table 2. Overview of the results on constructive influence type.

	0	1	2	3	n
Discussion of evaluation methods to improve evaluation quality	4 (1)	15 (4)	52 (14)	29 (8)	27
Illustration of potential for improvement without distorting the evaluation results	15 (4)	7 (2)	37 (10)	41 (11)	27
Discussion of the results presentation to improve the understanding of the target group	4 (1)	12 (3)	42 (11)	42 (11)	26
Exchange of ideas for quality improvement	4 (1)	7 (2)	41 (11)	48 (13)	27
Neutral discussion of conclusions	4 (1)	–	38 (10)	58 (15)	26
Constructive dialogue on improving evaluation quality	–	4 (1)	32 (8)	64 (16)	257
Support through early information delivery	–	–	19 (5)	81 (21)	26
<i>Constructive influence type</i>	–	–	39 (9)	61 (14)	23
<i>Betterment^a</i>	7 (2)	17 (5)	31 (9)	45 (13)	29
<i>Support^a</i>	–	10 (3)	26 (8)	64 (20)	31

Note: Scale of 0 (=No, I have never done that before), 1 (=Yes, I did that once), 2 (=Yes, I did that more than once), 3 (=Yes, I did that many times); n=total number of all valid cases. The numbers represent percentages; the respective frequencies are given in parentheses.

^a*Betterment* was measured with the item "I generally show the evaluators which aspects can be improved to achieve a better quality in the evaluation without distorting the results." *Support* was measured with the item "I am generally committed to ensuring an optimization of the evaluation's quality."

Table 3. Overview of the results on destructive influence type.

	0	1	2	3	n
More positive presentation of the evaluation results	74 (20)	7 (2)	19 (5)	–	27
Rephrasing of evaluative sentences	37 (10)	4 (1)	29 (8)	30(8)	27
Communication of expected results	74 (20)	11 (3)	11 (3)	4 (1)	27
Omission of certain parts	70 (19)	19 (5)	7 (2)	4 (1)	27
Withholding important information	96 (26)	–	–	4 (1)	27
Renegotiation of fees	74 (20)	4 (1)	11 (3)	11 (3)	27
Reference to the approach of others	62 (16)	16 (4)	19 (5)	4 (1)	26
<i>Destructive influence type</i>	16 (4)	65 (17)	19 (5)	-	26
<i>Undermining^a</i>	63 (19)	7 (2)	23 (7)	7 (2)	30
<i>Distortion^a</i>	66 (19)	21 (6)	10 (3)	3 (1)	29

Note: Scale of 0 (=No, I have never done that before), 1 (=Yes, I did that once), 2 (=Yes, I did that more than once), 3 (=Yes, I did that many times); n=Total number of all valid cases. The numbers represent percentages; the respective frequencies are given in parentheses.

^a*Undermining* was measured with the item "It may happen that the evaluation process is somewhat delayed due to my proposed amendments." *Distortion* was measured with the item "I may demand that evaluators shorten certain evaluation results."

of the clients have already applied this form of constructive influence in their evaluation practice, whereas only two clients (7%) have never done so ($N=29$). Similarly, all clients ($N=31$) stated that they have applied the indirect, constructive influence type of *Support* during their evaluation practice and thus are generally committed to optimizing the quality of the evaluation.

As regards the use of destructive influence, the results show that 84 percent of the respondents stated that they exercised a destructive influence at least once (*Undermining* and/or *Distortion*) (see Table 3) on evaluators ($N=26$).

Requested particularly frequently was the reformulation of individual evaluative sentences in the evaluation report. Also, in many cases, clients referred how other evaluators would have proceeded ($N=27$). While 37 percent of the clients reported that the evaluation process was delayed at least once by proposing adaptations of evaluation results (indirect form of destructive influence, *Undermining* ($N=11$)), 34 percent of the evaluators reported that certain evaluation results (direct form of destructive influence, *Distortion*) were shortened at least once ($N=10$). Thirty percent of the respondents said they had asked to leave out certain parts at least once. Likewise, more than a quarter of the respondents suggested at least once that the results be presented in a more positive light, and one person reported that they told the evaluators in advance what result they expected ($N=27$).

The findings show that the range of perceived, possible types of influence in the evaluation process—whether constructive or destructive—is extensive. Thus, the results indicate that the influence of clients can be not only negative and destructive but also positive and constructive.

Results concerning the relationship between US clients and evaluators

This section describes the results concerning the perceived relationship between US clients and evaluators concerning dissatisfaction and potential conflicts.

Dissatisfaction and destructive types of influence. To investigate possible causes for the destructive nature of the influence, clients were asked about their experiences with dissatisfaction with the evaluation process and results. Overall, 57 percent of the clients ($n=20$; $N=35$) had at least once been dissatisfied with an evaluation. Eighty-one percent of the clients cited the unsatisfactory quality of the evaluation and 57 percent the non-fulfillment of expectations as reasons for dissatisfaction. For 29 percent, the reasons for dissatisfaction were that the time schedule was not adhered to and the evaluators did not understand the evaluation context ($N=21$). The additional open question revealed further reasons for dissatisfaction. One respondent stated that the “complexity of the program structure and funder interests complicated the evaluation,” and it was further noted that “the evaluation report did not emphasize the strengths and successes of the project as much as [the client] would have liked” and that “the evaluator muted the report findings that [were] heard during the verbal briefing.” Other reasons for dissatisfaction were the lack of quality of the analysis and synthesis, “the evaluator [lacking] experience on the evaluation field,” such as, specifically, “the inability to help link [the] evaluation to [the] strategy.” Statistical analysis (two-sided Fisher’s exact test) confirmed a positive correlation between *dissatisfaction* with the *destructive influence type*, or in other words, that clients who experienced dissatisfaction with an evaluation were more frequently found to use destructive influence (Fisher’s exact test, $p=.021$). In contrast, no empirical evidence was found to connect dissatisfaction and a constructive influence type.

Conflict relationship and incentives. Furthermore, clients were asked whether they have ever had a conflict with an evaluator and how they reacted (through what we call here positive or negative “incentives”) when they were dissatisfied with the evaluation they had commissioned in

the past. About half the clients reported that they had conflicted with an evaluator due to their work or working methods (49%, $N=37$). Reasons given included that the “report was not delivered on [the] promised date,” “the evaluator had no experience in the field of evaluation! [sic],” and a “misuse of findings or misrepresentation of findings out of context—broad generalization with small/bias[ed] cohort.”

At the same time, findings show that 56 percent of the clients said they never reacted when dissatisfied with an evaluation by setting (mostly negative) incentives for evaluators. Rather, dissatisfaction in some cases had primarily internal consequences for the clients’ organizations, with one respondent stating that they “decided in the future we will do our own evaluation to maximize learning and relationship building inside our organization and with community partners.” Hence, dissatisfaction did not always affect the existing client–evaluator relationship. Nevertheless, 39 percent of respondents indicated they had created incentives once, and only one person (6%) had done so more than once. According to clients, the negative incentive that the evaluators will not be considered for future assignments was used most frequently. Twenty-two percent of the clients had done so at least once. Seventeen percent told the evaluators that the contract could be withdrawn, and another 17 percent proposed clear incentives to trigger changes in results. This took the form of a threat that the report would not be published in one case. Eleven percent asked clients at least once to change evaluation results while simultaneously implying negative consequences for evaluators actions ($N=18$). Findings revealed a significantly positive correlation between the conflict relationship and the incentive system (Fisher’s exact test, $p=.030$). That means that an evaluation relationship perceived as conflicted by the clients is positively related to negative incentives when clients are dissatisfied with an evaluation.

Evaluation clients’ perspectives and on the relationship between clients and evaluators in a country comparison

The following country comparison includes the US and Swiss clients’ expectations of independence of evaluations, the clients’ knowledge and familiarity of evaluation standards, as well as the relationship between clients and evaluators regarding the reaction to requests for amendments, the imputations of pressure/influence and preventive measures.

The selection of the two countries, United States and Switzerland, as comparison groups is based on data availability. So far, a comparable study among evaluation commissioners has only been conducted in Switzerland, which is why the results from Switzerland are compared with those from the United States. Nevertheless, this comparison is still compelling for two substantive reasons: First, a study already exists that compares the views of evaluators on independence from different countries, including the United States and Switzerland (Pleger et al., 2016). It is therefore possible to compare both the results on the evaluators’ perspectives and the clients’ perspectives from these two countries. Second, in terms of comparability between the two countries’ evaluation cultures, both countries were found to have a high degree of evaluation culture maturity (Jacob et al., 2015).

Client expectations of independence, knowledge, and familiarity of evaluation standards. Table 4 illustrates the expectations that clients from the United States and Switzerland have concerning the independence of evaluations.

Table 4. Overview of the results on the expectation of independence in country comparison.

A good evaluator . . .	Clients US			Clients CH		
	M	SD	N	M	SD	N
. . . is primarily obligated to follow defined evaluation standards	5.2	1.2	38	5.0	1.1	42
. . . only presents the actual findings	4.7	1.4	37	5.3	0.9	42
. . . must not be influenced regarding his or her methodical approach	4.6	1.4	37	4.6	1.1	41
. . . relentlessly reveals the weaknesses identified during the evaluation	4.5	1.5	39	5.0	1.2	44
. . . has first and foremost a moral responsibility toward his or her clients	3.9	1.8	36	2.9	1.5	39
. . . adjusts his or her results according to the expectations and needs of the clients	2.9	2.1	39	3.6	1.7	43
. . . weakens negative evaluation results toward clients to make them more acceptable*	1.7	1.1	39	–	–	–

SD: standard deviation.

Note: Scale from 0 (= do not agree at all) to 5 (=fully agree). For comparability, the values from 0 to 5 were adjusted to the values 1 to 6 (see Pleger and Hadorn, 2018). Data for Switzerland were taken from Pleger and Hadorn (2018: 8).

N=total number of all valid cases.

*Since this item was added as part of the present study, no data exist for Switzerland.

The average expectation of evaluation independence for the clients in the United States is 4.0 (out of 5) and in Switzerland 4.4 (out of 5)—both values indicating a high expectation. According to clients of both countries, they expect evaluators to be committed to defined evaluation standards. Both the US clients and the Swiss clients stated they expected evaluators to present only the results actually obtained. The clients of both countries also have relatively high expectations that evaluators will not be influenced in their methodological approach and will relentlessly disclose the evaluated policy's weaknesses. The expectation that evaluators have a moral responsibility toward stakeholders is stronger among US clients than Swiss clients. Nevertheless, the latter strongly expect that evaluators align the results with the expectations and needs of the clients. The expectation that evaluators will mitigate negative results so that they are more likely to be accepted is not widespread among US clients.

While 67 percent of Swiss respondents ($N=39$) said they knew the national evaluation standards, only 30 percent of the US clients ($N=30$) did. Of those clients who knew the evaluation standards, 78 percent in the US ($N=9$) and 63 percent in Switzerland ($N=26$) claimed to be familiar or very familiar with them.

Relationship between clients and evaluators in country comparison. To observe differences and similarities concerning the perceived relationship between clients and evaluators across countries, the findings of Swiss clients by Pleger and Hadorn (2018)—regarding reactions to requests for amendments, imputation of pressure and influence, difficulties in cooperation, and preventive measures—were integrated into the following country comparison.

According to the information provided by the clients in the United States and Switzerland, a compromise between the evaluators and clients was found in the majority of cases after the latter had proposed a change (see Table 5). This applies to 70 percent of US clients ($N=30$) and 53 percent of Swiss clients ($N=32$). One client in each country indicated that the proposed changes had been made. Twenty-seven percent ($n=8$) of US clients and 44 percent ($n=14$) of Swiss clients reported that the evaluators had not made the requested amendments. The clients

Table 5. Overview of the results on reaction and imputation in country comparison.

	Clients US		Clients CH	
	Percent	N	Percent	N
Reaction^a to requests for amendments				
The amendments were made	3	1	3	1
The amendments were not made	27	8	44	14
A compromise was found	70	21	53	17
		30		32
Imputation^b pressure/influence				
No, I have never been accused of this before	94	30	92	36
Yes, I have been accused of this before	0	0	3	1
Don't know	6	2	5	2
		32		39

Note: Data for Switzerland are taken from Pleger and Hadorn (2018: 10–12).

^aThe variable *Reaction* was measured with the following question: How did the evaluators you commissioned generally react when you asked them for amendments?

^bThe variable *imputation* was measured with the following question: Have you ever been accused by an evaluator of pressuring or influencing him or her to present results incorrectly or inaccurately?

were also asked whether the evaluators had ever confronted them with the accusation that they, as clients, had put pressure on the evaluators. The results show that in almost no cases did evaluators confront their clients with the imputation of influence or pressure. While no US client confirmed that they had already been confronted with such an accusation, one client in Switzerland did ($N=39$).

Although the comparison of the two countries is not based on the same measurement, the percentages indicate that the same main causes are responsible for the difficulties in the cooperation between clients and evaluators (see Table 6).

About half of the Swiss clients and half of the US clients indicated that the evaluator's lack of understanding of the organization being evaluated (or a lack of mutual understanding between the evaluators and their clients) was a reason for the problematic cooperation. About one-fifth of the Swiss clients agreed with the statement that independence of evaluation is difficult to guarantee without having too much influence on the evaluator ($N=36$). For the US clients, the evaluators' lack of critical professional competencies and resources (44%, $N=34$), and personal factors were named as causes of difficulties in cooperation (37%, $N=35$). In the open-ended question to the US clients, the point was raised that "it can be challenging to find evaluators that have both contextual expertise/insights into the perspective of the program beneficiaries as well as the methodological expertise." In addition to difficulties on the part of the clients, such as language barriers, lack of time to manage the evaluators well, or "to integrate the evaluators without burden on the organization or grantees," difficulties on the part of the evaluators were also mentioned. One client stated that "some evaluators are only theoretical model users, but do not see actual transactions" or do "not want to get involved in the data collection process."

Preferred preventive measures in country comparison. The survey provides important insights on the reasons for conflicts between clients and evaluators. When asked for the causes of

Table 6. Overview of the results on the difficulties in cooperation in country comparison.

Cooperation with evaluators is generally difficult because . . .	Clients US		Clients CH	
	Yes	N	Yes	N
. . . evaluators often do not understand the organization being evaluated	51 (18)	35	42 (15) ^a	36
. . . of a lack of mutual understanding between the evaluator and the client	46 (16)	35		36
. . . the independence of evaluations is difficult to guarantee without me having too much influence on the evaluator	27 (9)	34	19 (7)	36
. . . evaluators lack important expertise	44 (15)	34	14 (5)	36
. . . evaluators' personal views might impact the evaluation (e.g. political attitudes, personal preferences)	37 (13)	35	14 (5)	36
. . . evaluators' resources for their work (e.g. time) are too limited	41 (13)	32	11 (4)	36
. . . evaluators are not motivated*	9 (3)	33	–	–

Note: For the United States, the variable *Difficulties* was measured with the following question: "Where do you generally see the biggest difficulties in working with evaluators?" Respondents were asked to indicate with No (=0) or Yes (=1) whether the statements (in this table) apply by completing the related sentence "Cooperation with evaluators is generally difficult because . . ." *N*=total number of all valid cases. The figures represent percentages; the respective frequencies are given in parentheses; data for Switzerland came from Pleger and Hadorn (2018: 9). There, the variable was surveyed using an open-ended question, with the answers being assigned to individual categories.

^aThe present study measured the first two items unidimensionally (previously one category).

*Since this item was added as part of the present study, no data exist for Switzerland.

conflicts, lack of essential competencies on the part of the evaluators (61%), the poor quality of the evaluation results (67%), and the lack of understanding of the requirements (72%) were the most frequently given reasons for conflicts by the clients (*N*=18). Clients were then asked about appropriate measures to prevent disputes between themselves and evaluators; the corresponding findings on the perception of US clients concerning selected preventive measures, as well as a comparison with Swiss clients' responses, are displayed in Figure 2.

The clients of both countries most frequently selected the creation of greater mutual understanding (US: 82%, CH: 59%) and closer cooperation between clients and evaluators (US: 55%, CH: 33%) as appropriate preventive measures. The introduction of a formal evaluation protocol (46%) was ranked third among US clients, while only six percent of Swiss clients were convinced of the usefulness of this measure. The discussion of possible negative results (US: 39%, CH: 18%), the improvement of the data documentation (US: 33%, CH: 24%), the emphasis on the responsibility of the evaluators (US: 30%, CH: 18%), as well as the option to allow evaluation clients to make a formal statement on the evaluation results, which will then be published together with the evaluation report (US: 18%, CH: 12%), were considered to be further potentially useful prevention measures. For Swiss clients, the two measures of neutral intervention by third parties (3%) and meta-evaluation by independent third parties (6%) are relatively less important, whereas 18 percent of US clients cited these measures. An internal or external reporting unit was selected by 20 percent of Swiss clients and 12 percent of US clients (US, *N*=33; CH, *N*=34).

The results show that clients who had not yet experienced a conflict tend to select the preventive measures more frequently than the clients who have already had confrontations with evaluators. In contrast, neutral interventions by third parties and the creation of greater mutual understanding were selected more frequently by clients experiencing conflicts.

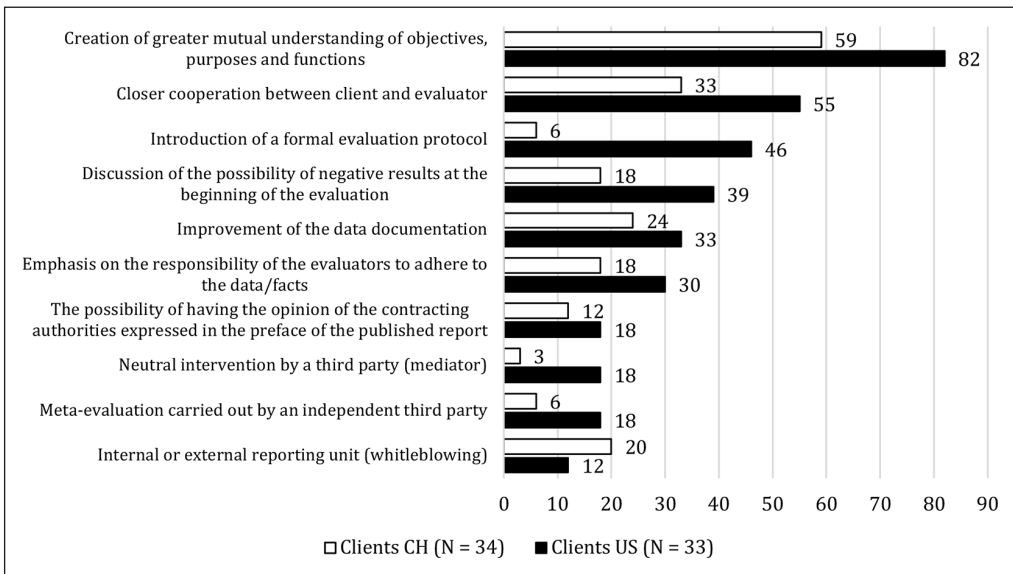


Figure 2. Values represent percentages; multiple answers were possible for the clients as respondents, but not for the evaluators. Data for Switzerland are from Pleger and Hadorn (2018: 13). Data for the US clients were collected in April and May 2019 via the online survey (see section “Survey design”).

Discussion

The results confirm the findings of previous studies, which indicate that there is a broad spectrum of possible ways in which clients influence the evaluation process. Here it is essential to distinguish between both constructive and destructive forms as well as between indirect and direct influences to understand whether influence leads to an improvement or distortion of evaluation results. Only this way can a distortion of evaluation results be prevented by eliminating destructive influences, but at the same time evaluations improved by allowing constructive influences. On the one hand, the BUSD model provides a theoretical basis to operationalize the different types of influence on the evaluation process and allows for the development of future research. On the other hand, the BUSD can be applied in practice both by evaluators and, as extended in this study, by clients to better assess the implications of their relationship in practice.

In contrast to Pleger and Sager (2018), who had formed the types of influence from the BUSD model ex post, in this article, the various forms of influence were operationalized ex ante and measured empirically for the first time. Among US clients, the constructive forms of influence *Betterment* and *Support* are comparatively more strongly represented than the destructive influence types *Distortion* and *Undermining*. In this article, the categorization of the different influencing actions by clients to either constructive or destructive influences is based on the BUSD model. However, certain influencing actions categorized as destructive (e.g. asking repeatedly for modifications) can in some cases also be considered constructive if they help to eliminate factual flaws in the reports. This limitation, which might result in a slight overestimation of single destructive influencing actions, must be considered when

interpreting the results. Moreover, the findings are based on self-reported data from the client perspective, which should also be considered.

Destructive influence and how to prevent it

About half of both US clients have, similar to findings from Swiss evaluation clients (Pleger and Hadorn, 2018), been dissatisfied with evaluations in the past. Accordingly, the relationship between clients and evaluators in the United States and Switzerland is not free of conflicts, with about half of clients indicating that they have at some point had a conflict with evaluators. The most frequently cited causes of conflict among US clients are the perceived lack of essential competencies on the part of the evaluators and the poor quality of the evaluation reports. That (perceived) insufficient competencies on the part of evaluators and poor-quality reports lead to dissatisfaction on the part of the clients does not seem surprising. Clients hire evaluators because they seek the expertise not available within their own organizations. If the evaluators do not bring this expertise as expected, this inevitably leads to conflict.

Importantly, the findings reveal a positive, statistically significant relationship between dissatisfaction and the destructive type of influence. This indicates that dissatisfaction with various aspects of an evaluation increases clients' urge to take control over the evaluation outcome, for instance, because they think that evaluators cannot adequately evaluate without client support. This then opens the door for destructive attempts to influence the evaluators. In turn, this implies that a reduction of dissatisfied clients would also help to solve the problem of destructive influence on evaluation results. Therefore, preventive measures must take effect early in the evaluation process so that potential dissatisfaction on the client's part can be prevented. This is the only way to ultimately reduce the perceived need for influence on the part of the client.

Consistently, when asked about their preferred preventive measures, the clients named the following: (1) creating a greater mutual understanding of objectives, purpose, and functions of evaluations and (2) fostering closer cooperation between both parties. The preferred preventive measures thus clearly aim to decrease the most frequently experienced reasons for conflict: a lack of competence by the evaluators and the poor quality of the evaluation reports (both as perceived by clients). Preventive measures could consist of a more substantial dialogue between both parties to reduce the client perceptions that evaluators do not have the necessary skills and that the quality of the reports is poor. An essential part of the dialogue is listening, including characteristics such as making an effort to understand the other party's views and requests or appropriate responses to others after careful consideration. Listening is relevant not only for effective communication because it fosters mutual understanding (of expectations) but also for relationship building and rebuilding of trust. Trust can be fostered through transparency as an element of communication when clients and evaluators have a conflict or distrust each other (Hung-Baesecke and Chen, 2020: 6–7). However, even if an improved dialogue between clients and evaluators helps decrease a mismatch of expectations, genuine cases of poor evaluator skills remain a potential source of conflict. Accordingly, Picciotto (2019) argued that there is also a need for evaluators to develop their competencies (p. 95). To fill the growing gap of specialized evaluation in times of rapid change, evaluators need to expand their instruments and improve their competence. To meet the expectations of the clients, evaluators would have to pursue innovative approaches and at the same time be better embedded in management systems and social processes.

Interestingly, clients who had not yet experienced conflicts with evaluators tended to be more open toward the use of preventive measures than those who had experienced conflict. Nevertheless, the intervention of neutral third parties to solve disputes and the creation of greater mutual understanding were frequently perceived as appropriate preventive measures by persons with conflict. These findings suggest that clients without conflict may have indicated the very measures they already use in their evaluation practice, since the frequency of past behavior has been found to predict people's future activities (i.e. people tend to perceive actions they have already experienced as appropriate) (Marien et al., 2018: 54). Those who have experienced conflict might already show some resignation and see the problems in others rather than in their own actions. They are likely to consider third-party intervention as appropriate or emphasize the commitment of both players to reaching a mutual understanding. This mechanism can be attributed to so-called attribution processes suggesting that individuals tend not to see their own behavior as the cause of problems but to blame external sources for what they experience (Kelley, 1973).

Constructive influence type

Nevertheless, results indicate that evaluations benefit even more frequently from constructive influence. Clients stated that they support evaluators, for instance, by providing relevant information early and conducting a constructive dialogue concerning the improvement of the quality of the evaluation. According to the clients, this was done not to distort but to improve the quality of the evaluation results. For instance, one respondent summarized their motives behind influence in the evaluation process as follows: "my proposed amendments are to improve the scientific standards to make the reporting more objective, factual, or evidence-based." Almost all the US clients stated they had a constructive dialogue and exchanged ideas (e.g. neutral discussion of conclusions, the presentation of results, and the evaluation methods) more than once with evaluators to improve the evaluation quality. One client explicitly stressed that they called for changes to the "recommendations purely to make it more understandable to the wider audience—NOT [*sic*] to change results or consequences of the findings."

Importantly, the results show that the clients' work experience (in terms of the number of evaluations carried out and years of commissioning evaluations) has a positive, statistically significant correlation with the constructive type of influence. The longer and more intensively a client has been active in their work, the more pronounced their constructive influence. In addition, the more familiar clients are with the evaluation standards, the higher their expectations of the independence of evaluations. This underscores how important it can be that clients know evaluation standards so that, conversely, the independence of evaluations can be guaranteed or improved. Against the background of the low level of awareness of these standards, however, there is a plea for the necessary reduction of this alarming information deficit on the part of the clients. However, because it has proven to be extremely difficult to approach the US clients within the scope of this study, this low level of knowledge of the standards among US clients does not seem to be a surprise. While evaluators are institutionally well-organized in the AEA, such an organization lacks on the evaluation client side. This greatly hinders the professionalization of this group, for example, because information on evaluation standards cannot be distributed systematically among evaluation clients. This exposes a fundamental problem relating to evaluation ethics since improving the understanding and acceptance of evaluation standards on the part of commissioning parties only works if they can be

reached. Studies have shown how important it is for both evaluators (Pleger et al., 2016) and clients (Pleger and Hadorn, 2018) to create a greater mutual understanding of objectives, purposes, and functions of evaluations. However, it will be difficult to achieve this goal without an institutionalized channel through which a systematic exchange to increase mutual understanding can be facilitated.

Both professional experience and familiarity with evaluation standards thus benefit evaluation processes and outcomes by increasing the odds of constructive influences or decreasing destructive influences because of their higher expectations of evaluation independence. Moreover, the finding that almost no amendments were made can be considered positive from an ethical point of view. Finally, the findings also indicate that evaluation standards contribute to quality assurance and that evaluators follow them, which is consistent with the findings of Pleger et al. (2016).

Attitude and future contracts as an explanation for differences in dealing with conflict relationship

In general, the question arises whether the interests of the clients and evaluators in the evaluation relationship are actually as different as assumed. From an evaluation-ethical perspective, both actors should pursue the same, higher interests of scientific independence and integrity, as well as compliance with evaluation standards and high evaluation quality (AEA, 2021b). In practice, however, diverging interests may be expressed in terms of difficulties in cooperation or manifest conflicts. One of the main difficulties during evaluations, according to US and Swiss clients, is the evaluators' lack of understanding for the organization being evaluated and the lack of mutual understanding between the two parties. This can potentially lead to different expectations, resulting in client dissatisfaction with evaluation processes and outcomes. In turn, as the survey results have shown, dissatisfaction fosters clients' destructive influence, potentially leading to a decrease in the quality of the evaluation. However—and this seems to be one of the key problems in evaluation practice both in the United States and in Switzerland—hardly any client has ever been accused by evaluators of exerting pressure or influence. These findings are surprising in light of the findings of Morris and Clark (2012), who found that 42 percent of evaluators have been pressured to misrepresent evaluation results (p. 61). Consequently, while tensions due to attempted destructive influence are not uncommon in evaluation practice (see, for example, Morris, 2007; Morris and Clark, 2012; Morris and Cohn, 1993; Pleger and Hadorn, 2018; Stockmann et al., 2011; LSE GV314 Group and Page, 2014; Turner, 2003), they are not explicitly addressed and therefore remain an undiscussed and often invisible issue.

The overly optimistic perception of US clients in terms of lacking imputation of exerting pressure might result from distorted rationalization processes and the effects of social desirability. Consequently, clients present the actual situation more positively to protect their image of ethical correctness. Similar behaviors were identified by Morris and Clark (2012) for the evaluators (p. 67). The fact that evaluators hardly ever explicitly address the problem of destructive influence by clients might be explained for the US evaluators by their attitude, as they appear to be less optimistic than Swiss evaluators about the chances of effectively preventing pressure (Pleger et al., 2016: 12). US evaluators might therefore be less inclined to address pressure and risk a conflict with the client because they have little hope of success. Instead, they fear the adverse effects of possible follow-up orders if they address the issue.

As also mentioned in Morris and Clark (2012) and discussed in Pleger and Hadorn (2018), evaluators are prone to yield to pressure or seek compromise to minimize the risk of losing future contracts. Nevertheless, it is clear that—as in Switzerland—the US evaluation landscape also lacks effective conflict communication (Pleger and Hadorn, 2018: 14).

Conclusion

The study aimed to provide an explorative overview of evaluation client perspectives on various forms of influence on the evaluation process in the United States. Thereby, in the specific context of evaluations, clients are not neutral actors, but rather have the character of clients who often have a clear idea of (and may apply different strategies to achieve) the desired evaluation outcome (Schoenefeld and Jordan, 2017). Employing an online survey of 62 evaluation clients (non-representative sample), we strive to gain a better understanding of client behavior by examining their viewpoints on the independence of evaluation in the context of evaluation processes, such as client dissatisfaction with evaluations and their relationships with evaluators. The study operationalized the various forms of influence systematically, differentiating between constructive and destructive forms of influence for the first time. This nuanced understanding of influence based on the BUSD model allows improved exploratory, non-representative insights into the evaluator/client relationship. Furthermore, by portraying their relationship, existing lines of conflict and their reasons can be mapped. The present investigation of the type of constructive influence, therefore, provides initial insights into the type of constructive influence and illustrates the importance of integrating this dimension into future research. Consequently, the study expands our knowledge about the dynamics between clients and evaluators in three respects upon which future research can build on:

First, the findings revealed that dissatisfaction with various aspects of an evaluation increases a client's urge to take control, thereby exercising a destructive influence. In turn, this implies that a reduction of dissatisfied clients would also help to solve the problem of destructive influences on evaluation results. Hence, this insight is of key importance to the quality of evaluations and the present study offers important insights into the reasons for dissatisfaction on the part of clients (e.g. lacking competence of evaluators; or frustration about the fact that the clients did not succeed in utilizing the results for strategy development). Future (especially qualitative) studies should build on this knowledge and could provide even more in-depth information on the dynamics within an evaluation process that explain why dissatisfaction arises.

Second, the article provides insights for the practical implementation of preventive measures: A better dialogue between clients and evaluators initiated early in the evaluation process fosters mutual understanding (of expectations), helps to improve relationships, and strengthens trust. Through this, potential dissatisfaction can be reduced, enhancing the independence of evaluations. In cases where the relationship between client and evaluator is still troubled, the intervention of an independent mediator seems to be an acceptable solution. Furthermore, to reduce conflict, an open culture of communication between the two parties is recommended. Disagreements, dissatisfaction, and difficulties should be discussed openly through a constructive dialogue before a serious conflict arises. These insights are strongly related to the well-known trade-off in evaluations of evaluators either

being too close to commissioners (i.e. clients) or too far away: On the one hand, being at risk to be influenced versus, on the other hand, not involving the client enough (e.g. not considering the clients' needs and context or not nurturing dialogue) (see, for example, Bourgeois and Whynot, 2018; Mohan et al., 2006; Mohan and Sullivan, 2006) and hence generating results that will not be utilized. Finding the right balance in this regard is very challenging for not only the commissioner but also evaluators as "most evaluators are aware that the line between scientific objectivity and political conviction is often very fine" (Balthasar, 2011: 222). Our findings suggest that creating more closeness (and ultimately an increased exchange) between evaluators and clients increases rather than decreases independence in some circumstances. Intensifying the dialogue might decrease the clients' perceived needs to destructively influence the evaluation process because they have higher trust in the competences of the evaluators and the usefulness of the evaluation results.

Third, both professional experience and familiarity with client evaluation standards benefit evaluation processes and outcomes by increasing the chances of constructive influences (or decreasing destructive influences). Finally, higher expectations regarding the independence of evaluations can be assumed with professional experience and familiarity with professional standards. Accordingly, the alarming information deficit on the part of clients concerning the US "Program Evaluation Standards" needs to be reduced.

While the AEA (2021a) attaches great importance to inclusion and diversity in its membership, evaluation clients should also be increasingly involved in such exchanges, so that the relevance and urgency of evaluation standards can be communicated to this important stakeholder group (e.g. in the form of broad-based information campaigns, conferences). In the long term, this strategy could contribute to a continuous reduction of the information deficit of US clients regarding evaluation standards. In Switzerland, the national evaluation society (SEVAL) continuously strives to increase awareness of the standards among evaluation clients, and we indeed find a higher level of familiarity with the standards compared to US clients. Future research should further explore the relationship between familiarity with the standards and destructive influence on evaluations, thereby investigating the role that national evaluation societies can play in promoting professionalization in this regard.

Furthermore, understanding the client perspective enables a complete description of the mutual evaluation relationship in the complex context of the evaluation process. Overall, this exploratory article illustrates that the independence of evaluations can only be improved if we understand whether and to what extent there is awareness of unethical behavior on both sides (Pleger and Hadorn, 2018: 4). Picciotto (2019) also concluded that the independence of evaluations goes hand in hand with the professionalization of the evaluation discipline (p. 95). The integrity of evaluation processes can be maintained if the necessary professional autonomy is achieved through accreditation systems and ethically sound evaluation guidelines known and respected by all stakeholders. Overall, investing in the generation of knowledge about the motives for destructive influence on the part of clients as well as building on this knowledge in practice is thus worthwhile and will ultimately raise the social impact of policy evaluations.

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.

ORCID iD

Fabienne Helen Schmidli  <https://orcid.org/0000-0002-4352-247X>

Notes

1. The terminological distinction between commissioner and client implies the aforementioned problematic dependencies as “depending on whom—which client, which evaluation commissioner—the evaluator works for, the power relationships between evaluator and the program being evaluated may well change” (De Laat, 2014: 20).
2. Qualtrics allows sending emails to a maximum of 25,000 email addresses each week. For this reason, data had to be collected in stages.
3. Excluded respondents were either not clients or had never commissioned an evaluation alone or in a team.
4. Four frequency categories: No, I have never done that before (0); Yes, I did that once (1); Yes, I did that more than once (2); Yes, I did that many times (3).

References

- American Evaluation Association (AEA) (2021a) About us. Available at: <https://www.eval.org/About/About-AEA> (accessed 23 March 2021).
- American Evaluation Association (AEA) (2021b) Guiding principles. Available at: <https://www.eval.org/About/Guiding-Principles> (accessed 23 March 2021).
- Australian Evaluation Society (AES) (2022) Ethical guidelines. Available at: <https://www.aes.asn.au/ethical-guidelines> (accessed 4 January 2022).
- Azzam T, Wanzer DL, Knight C, et al. (2021) The manifestations of politics in evaluation: An exploratory study across the evaluation process. *Evaluation and Program Planning* 88: 101947.
- Balthasar A (2011) Critical friend approach: Policy evaluation between methodological soundness, practical relevance, and transparency of the evaluation process. *German Policy Studies* 7(3): 187–231.
- Barnett C and Camfield L (2016) Ethics in evaluation. *Journal of Development Effectiveness* 8(4): 528–34.
- Berriet-Sollic M, Labarthe P and Laurent C (2014) Goals of evaluation and types of evidence. *Evaluation* 20(2): 195–213.
- Bourgeois I and Whynot J (2018) Strategic evaluation utilization in the Canadian federal government. *Canadian Journal of Program Evaluation* 32(3): 327–46.
- Canadian Evaluation Society (CES) (2022) Program evaluation standards. Available at: <https://evaluationcanada.ca/program-evaluation-standards> (accessed 4 January 2022).
- De Laat B (2014) Evaluator, evaluand, evaluation commissioner: A tricky triangle. In: Loud ML and Mayne J (eds) *Enhancing Evaluation Use: Insights from Internal Evaluation Units*. Thousand Oaks, CA: SAGE, 15–36.
- DeGEval (2022) Standards für Evaluationen [Standards for evaluations]. Available at: <https://www.degeval.org/en/degeval-standards/download/> (accessed 4 January 2022).
- Desautels G and Jacob SA (2012) The ethical sensitivity of evaluators: A qualitative study using a vignette design. *Evaluation* 18(4): 437–50.
- Eckhard S and Jankauskas V (2019) The politics of evaluation in international organizations: A comparative study of stakeholder influence potential. *Evaluation* 25(1): 62–79.

- Eliadis P, Furubo J-E and Jacob S (2011) *Evaluation: Seeking Truth or Power?* New Brunswick, NJ: Transaction Publishers.
- Gullickson AM (2020) The whole elephant. *Evaluation and Program Planning* 79: 101787.
- Head BW (2008) Three lenses of evidence-based policy. *Australian Journal of Public Administration* 67(1): 1–11.
- Hung-Baesecke C-JF and Chen Y-RR (2020) Explicating trust and its relation to dialogue at a time of divided societies. *Public Relations Review* 46(1): 101890.
- Jacob S, Speer S and Furubo JE (2015) The institutionalization of evaluation matters: Updating the *International Atlas of Evaluation* 10 years later. *Evaluation* 21(1): 6–31.
- Kelley HH (1973) The processes of causal attribution. *American Psychologist* 28(2): 107–28.
- Leeuw FL (2012) Linking theory-based evaluation and contribution analysis: Three problems and a few solutions. *Evaluation* 18(3): 348–63.
- LSE GV314 Group and Page EC (2014) Evaluation under contract: Government pressure and the production of policy research. *Public Administration* 92(1): 224–39.
- Marien H, Custers R and Aarts H (2018) Understanding the formation of human habits: An analysis of mechanisms of habitual behaviour. In: Verplanken B (ed.) *The Psychology of Habit: Theory, Mechanisms, Change, and Contexts*. Cham: Springer, 51–69.
- Mayne J (2017) Independence in evaluation and the role of culture. In: Barbier J and Hawkins P (eds) *Evaluation Cultures: Sense-Making in Complex Times*. London: Routledge, 111–44.
- Mohan R and Sullivan K (2006) Managing the politics of evaluation to achieve impact. *New Directions for Evaluation* 2006(112): 7–23.
- Mohan R, Tikoo M, Capela S, et al. (2006) Increasing evaluation use among policymakers through performance measurement. *New Directions for Evaluation* 2006(112): 89–97.
- Morris M (2007) Foundation officers, evaluation, and ethical problems: A pilot investigation. *Evaluation and Program Planning* 30(4): 410–5.
- Morris M (2015) Research on evaluation ethics: Reflections and an agenda. In: Brandon PR (ed.) *Research on Evaluation: New Directions for Evaluation*. Hoboken, NJ: Wiley Periodicals, 31–42.
- Morris M and Clark B (2012) You want me to do what? Evaluators and the pressure to misrepresent findings. *American Journal of Evaluation* 34(1): 57–70.
- Morris M and Cohn R (1993) Program evaluators and ethical challenges: A national survey. *Evaluation Review* 17(6): 621–42.
- Newman DL, Scheirer MA, Shadish WR, et al. (1995) Guiding principles for evaluators. *New Directions for Program Evaluation* 1995(66): 19–26.
- Organisation for Economic Cooperation and Development (OECD) (2010) *Quality Standards for Development Evaluation*. Paris: OECD Publishing.
- Patton MQ (1997) *Utilization-Focused Evaluation: The New Century Text*. Thousand Oaks, CA; London; New Delhi, India: SAGE.
- Perrin B (2018) How to manage pressure to change reports: Should evaluators be above criticism? *American Journal of Evaluation* 40: 354–375.
- Picciotto R (2019) Is evaluation obsolete in a post-truth world? *Evaluation and Program Planning* 73: 88–96.
- Pleger LE and Hadorn S (2018) The big bad wolf's view: The evaluation clients' perspectives on independence of evaluations. *Evaluation* 24(4): 456–74.
- Pleger LE and Sager F (2018) Betterment, undermining, support and distortion: A heuristic model for the analysis of pressure on evaluators. *Evaluation and Program Planning* 69: 166–72.
- Pleger LE, Sager F, Morris M, et al. (2016) Are some countries more prone to pressure evaluators than others? Comparing findings from the United States, United Kingdom, Germany, and Switzerland. *American Journal of Evaluation* 38(3): 315–328.
- Schoenefeld J and Jordan A (2017) Governing policy evaluation? Towards a new typology. *Evaluation* 23(3): 274–93.

- Smith NL (1998) Designing investigative evaluations: Problem solution versus client responsiveness. *Evaluation* 4(2): 117–29.
- Stockmann R, Meyer W and Schenke H (2011) Unabhängigkeit von Evaluationen [Independence of Evaluations]. *Zeitschrift Für Evaluation* 10(1): 39–67.
- Swiss Evaluation Society (2022) SEVAL standards. Available at: <https://www.seval.ch/en/standards-competences/standards/> (accessed 4 January 2022).
- Turner D (2003) Evaluation ethics and quality results of a survey of Australasian Evaluation Society members. *AES Ethics Committee*. Available at: https://www.aes.asn.au/images/ethics_survey_summary.pdf?type=file
- UK Evaluation Society (2022) Good practice guidelines. Available at: <https://www.evaluation.org.uk/professional-development/good-practice-guideline/> (accessed 4 January 2022).

Fabienne Helen Schmidli is Research Associate at the Zurich University of Applied Sciences (ZHAW), Center for Arts Management. Her research interests include evaluation, arts and nonprofit management, and audience development.

Lyn Ellen Pleger, PhD, is Research Associate at the Zurich University of Applied Sciences (ZHAW), Institute of Public Management. Her research interests include evaluation, public policies, e-government, and voting behavior.

Susanne Hadorn, PhD, is a Research Associate and Evaluator at the KPM Center for Public Management, University of Bern. Her research interests include evaluation, health policy, and policy implementation.