

Motivation to Adopt Evaluation Practice as a Determinant of Evaluation Use

Tomasz Kupiec¹  and Zuzanna Wrońska¹ 

American Journal of Evaluation
1-17

© The Author(s) 2024



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10982140241290737
journals.sagepub.com/home/aje



Abstract

The literature contains several models that link types of evaluation use with organizational factors. However, until now, none of them have been thoroughly verified. This study aims to empirically verify the hypotheses proposed by Steven Højlund, who suggests that the type of evaluation use in organizations is determined by the adoption mode of evaluation practice (coercive, mimetic, normative, or voluntary). Qualitative comparative analysis was conducted on 23 cases from departments of Polish municipal administration, revealing the necessary conditions for most types of evaluation use. Instrumental use resulted from voluntary adoption, symbolic use from coercive adoption, and normative adoption—or the absence of coercive factors—led to legitimizing use. The situation was least clear for conceptual use, where we identified three possible combinations of necessary conditions. Although we found some significant relationships between adoption modes and types of evaluation use, the overall explanatory potential of Højlund's model appears to be quite limited.

Keywords

evaluation use, organizational theory, organizational institutionalism, local government, Poland, qualitative comparative analysis

Introduction

Evaluation is a key tool for facilitating learning in public policy and public administration (Picciotto, 2016; Van der Knaap, 1995). It has been institutionalized in the administrations of most developed countries as part of several waves of public management reforms under different labels, including new public management (Furubo et al., 2002; Norris & Kushner, 2007) and evidence-based policy-making (Davies, 2012).

For decades, evaluation use has dominated the concerns of practitioners and emerged as a central theme in academic research (Christie, 2007; King & Alkin, 2019). More recently, scholars have explored this issue through the lens of organizational theory (Hanberger, 2011; Raimondo, 2018).

¹ University of Warsaw, Warsaw, Poland

Corresponding Author:

Tomasz Kupiec, Centre for European Regional and Local Studies, EUROREG, University of Warsaw, Krakowskie Przedmieście 30, 00-927 Warsaw, Poland.

Email: tomasz.kupiec@uw.edu.pl

Several researchers have drawn on principal-agent theory and organizational institutionalism to argue that external pressures often shape how organizations use evaluation (e.g., Carman, 2005; Eckerd & Moulton, 2011; Raimondo, 2018).

Højlund (2014) suggested that the way organizations adopt evaluation practice determines the type of evaluation use. His model identifies four adoption modes: coercive, mimetic, normative, and voluntary, which correspond to four types of evaluation use: instrumental, conceptual, legitimizing, and symbolic.¹ While promising, the model has not undergone thorough testing. Additionally, some empirical studies (e.g., Kupiec et al., 2020; Preskill & Boyle, 2008) raise questions about certain assumptions in the model.

In this study, we examine the explanatory potential of Højlund's model within the context of municipal administration in Poland to verify the relationship between the motivation to adopt evaluation practice and the resulting types of evaluation use. The reason for the case selection was a high diversity of evaluation practices between different city halls, allowing us to verify the relationships between different adoption modes and types of evaluation use. The research utilizes qualitative comparative analysis (QCA).

The remainder of this paper is structured as follows: the next section outlines the theoretical background of the study, followed by a detailed explanation of the research design. The subsequent section presents the findings, and the paper concludes with a discussion.

Theory of Organizational Factors of Evaluation Use

Weiss (1972) laid the foundation for research on evaluation use factors, marking the start of an ongoing and complex investigation into how and why evaluations are used or not used. Subsequent reviews (Alkin, 1985; Cousins & Leithwood, 1986; Johnson et al., 2009; Leviton & Hughes, 1981; Shulha & Cousins, 1997) have analyzed the growing body of research, serving as key references in discussions on this topic. As argued by Alkin and King (2017) determinants of evaluation use fall into four major categories: user factors, evaluator factors, evaluation factors, and organizational/social context factors.

Since the early twenty-first century, particularly in Europe, systems thinking has been increasingly used as a complementary approach in the field of evaluation and evaluation use. This approach views evaluation as a system interconnected with other organizational systems (Leeuw & Furubo, 2008). Influenced by this approach, the focus has shifted from isolated studies to streams of studies flowing within evaluation systems (Rist & Stame, 2006) and toward understanding evaluation use in organizational contexts (e.g., Hanberger, 2011). This shift addresses the lack of contextual explanations in earlier research (Højlund, 2014) and responds to critiques that traditional evaluation theories rely too heavily on assumptions of rationality and causality (Raimondo, 2018). It also reflects the need to complement core assumptions about causality and rationality with organizational theory when analyzing evaluation use (Dahler-Larsen, 2012; Sanderson, 2000).

From an organizational theory perspective, traditional explanations of evaluation use often align with rational choice theory. This theory suggests that evaluations support informed decision-making to achieve goals and maximize efficiency (Sanderson, 2000). While this approach plausibly explains instrumental and conceptual use, it fails to address the prevalence of nonuse and symbolic use (Højlund, 2014; Raimondo, 2018). The latter refers to situations where the mere existence of an evaluation, rather than its actual results, is used for persuasion (Pelz, 1978).

Other organizational theories, such as agency theory and organizational institutionalism, aim to address this gap. Agency theory argues that evaluation primarily meets the information needs of funders, supervising bodies, and commissioners. For instance, in development aid context, evaluation serves to fulfill external accountability requirements (Carden, 2013). International organizations establish evaluation systems to ensure that agents (staff) meet the objectives set by principals

(member states) (Weaver, 2007). This function of evaluation—managing transaction costs, information asymmetries, and principal-agent problems—is especially common in large or multiorganizational contexts (Picciotto, 2016). In contrast, organizational institutionalism views evaluation's initial goal as securing legitimacy from external actors (Ahonen, 2015). Evaluation practice is shaped by environmental pressures (Carman, 2011), which may vary depending on organizational characteristics (Brunsson, 2002) or the policy area in which the organization operates (Boswell, 2008). These pressures can influence the evaluation process, affecting methods used (Eckerd & Moulton, 2011) and types of evaluation use (Højlund, 2014).

Among various theoretical frameworks on organizational factors influencing evaluation use,² this study focuses on a promising model proposed by Højlund (2014). His framework suggests that the dominant type of evaluation use in an organization results from the interaction of two factors: external pressure to adopt evaluation, which may come from regulations, cultural norms, uncertainty, or environmental expectations, and the organization's internal propensity to evaluate, which depends on its characteristics or role. Organizations focused on action, which rely on producing tangible outputs, tend to have a high propensity to evaluate, while political organizations, which gain legitimacy through discussions and decisions, typically have a low propensity (Brunsson, 2002). These factors form a two-dimensional matrix, leading to four possible modes of evaluation adoption.

- **Coercive:** Evaluation occurs due to formal or informal pressure from other organizations, regulations, or cultural expectations.
- **Mimetic:** Evaluation takes place to emulate successful or leading organizations, aiming to enhance the chances of success.
- **Normative:** Evaluation is recommended by consultants, new management, or staff who have learned this practice from previous workplaces.
- **Voluntary:** Evaluation is conducted based on the belief that it facilitates learning and improvement.

The first three modes align with DiMaggio and Powell's (1983) concept of "institutional isomorphism," which describes how organizations adapt to resemble others in similar environmental conditions. The fourth mode reflects the rationalist perspective, which views "social betterment" as the ultimate goal, with evaluation aiding in its achievement. Each adoption mode is expected to result in different dominant types of evaluation use. This framework considers four types of evaluation use, along with nonuse.

- **Instrumental:** Evaluation findings directly influence decisions and lead to changes in the evaluand.
- **Conceptual:** Evaluation findings alter decision-makers' awareness and attitudes, contributing to new knowledge.
- **Legitimizing:** Evaluation findings serve to justify or support the organization, the evaluated intervention, or specific decisions.
- **Symbolic:** The evaluation process serves to affirm the organization's or the policy/program's validity or importance.
- **Nonuse:** Despite conducting the evaluation, none of the above types of use occur.

Table 1 presents the predicted links between evaluation adoption modes and primary evaluation use types. Voluntary adoption of evaluation typically leads to either instrumental or conceptual use. However, mimetic adoption may also result in instrumental use, while normative adoption can lead to conceptual use. Our study aims to test these assumptions by identifying significant relationships between evaluation adoption modes and evaluation use types. We outline specific hypotheses in the next section.

Table 1. Overview of Four Ideal Evaluation Adoption Modes and the Associated Expected Evaluation Uses.

Adoption mode	Coercive	Mimetic	Normative	Voluntary
Primary use types	No use or symbolic use	Legitimizing and instrumental use	No use or conceptual use	Instrumental and conceptual use

Source: Adapted from Højlund (2014)

Empirical verification of organizational factors is essential because proponents argue that these factors are the true determinants of evaluation use, influencing other “immediate” factors explored in previous studies. We selected Højlund’s framework (2014) for its universality, as it applies to organizations of various sizes, reach, and types.

Research Design

To provide a comprehensive overview of our research approach, this section is divided into three key parts outlining the methodology, case selection criteria, and data collection techniques.

Method

The relationship between evaluation adoption modes and types of use is complex and set-theoretic.³ Each adoption mode can lead to multiple types of evaluation use, and each type of use can arise from various adoption modes. This complexity, involving equifinality, conjunctural causation, and asymmetric causality, justifies using QCA to identify the necessary and sufficient conditions (adoption modes) for each type of evaluation use.

Qualitative comparative analysis examines logical implications and set relations regarding necessity and sufficiency (Thomann & Maggetti, 2020). This case-oriented method, based on set theory, employs Boolean logic to simplify data complexity and uncover complex interactions (Meuer & Rupiotta, 2017). For a detailed discussion of the ontological and technical aspects of QCA, see Rihoux and Ragin (2009) or Schneider and Wagemann (2012).

Qualitative comparative analysis has successfully addressed evaluation use factors in previous studies (e.g., Balthasar, 2006; Ledermann, 2012; Pattyn & Bouterse, 2020). Although our sample of 23 cases is small compared to typical QCA studies (Greckhamer et al., 2013) or on the lower end of medium-sized samples (Berg-Schlosser et al., 2008; Marx & Dusa, 2011; Mello, 2012), it is not small when compared to previous research on evaluation use factors.

Using QCA, we analyzed the relationship between four adoption modes (conditions)—coercive (Cor),⁴ mimetic (Mim), normative (Nor), and voluntary (Vol)—and five types of evaluation use (outcomes): instrumental (Ins), conceptual (Con), legitimizing (Leg), symbolic (Sym), and nonuse (Non).⁵ We examined conditions for each outcome separately, conducting five distinct analyses of necessary and sufficient conditions, with four adoption modes and one type of use per analysis. Testing four conditions is manageable with a 23-case sample (Marx & Dusa, 2011).

To test Højlund’s model (Table 1), we translated it into a set of hypotheses based on set relations. Each hypothesis addresses the necessary conditions for evaluation use.

H1: The disjunction of voluntary and mimetic adoptions is a necessary condition for instrumental use (Vol + Mim \leftarrow Ins).

H2: The disjunction of voluntary and normative adoptions is a necessary condition for conceptual use (Vol + Nor \leftarrow Con).

H3: Mimetic adoption is a necessary condition for legitimizing use (Mim \leftarrow Leg).

H4: Coercive adoption is a necessary condition for symbolic use (Cor \leftarrow Sym).

H5: The disjunction of coercive and normative adoptions is a necessary condition for nonuse (Cor + Nor \leftarrow Non).

Additionally, we conducted an analysis of sufficiency to identify which adoption modes are sufficient for each type of evaluation use.

Case Selection

This study focuses on cases from Polish municipal administration. Poland serves as a representative example of the 13 Central and Eastern European countries that joined the European Union in 2004 or later. These countries share a common administrative tradition (Meyer-Sahling, 2009) and have comparable trajectories of recent administrative development (Ágh, 2016), including a low maturity of evaluation culture (Kupiec, 2022).

Municipal evaluation is not the most developed area of evaluation within Polish administration but that distinction belongs to cohesion policy evaluation conducted by central and regional administrations (Kupiec et al., 2020). However, evaluation at the municipal level has become more common in recent years, with over 400 studies conducted in the 30 largest cities over the past decade (Kupiec & Celińska-Janowicz, 2024). This field offers ideal study material due to its diversity in terms of evaluands (reflecting varied municipal activities), organization of evaluation processes (both external and internal studies), community involvement, and, most importantly, the motivations behind the establishment of evaluation practice.

We selected 23 cases for analysis. Each case represents an evaluation practice within a specific department of a city hall, involving a series of evaluation studies initiated at a particular time and focused on one evaluand—a specific type of municipal activity. The selection process followed a maximum variation strategy (Flyvbjerg, 2006), using the following criteria for variation:

- Municipality: No more than two cases from the same municipality, with cases drawn from different departments.
- Evaluand: Topics such as participatory budgeting, social policy, development strategy, culture, and cooperation with nonprofit organizations.
- Evaluation Subject: Differentiating between evaluations focused on processes versus outcomes.
- Organization of the Evaluation Process: Differentiating between evaluations conducted internally and those commissioned to external contractors.

Data Collection

The primary source of information for this study was semistructured interviews. Since the evaluation practice, as we have defined it, is associated with a specific department of a city hall, our respondents were either heads of department or chief specialists they designated, typically with over 10 years of experience in both conducting evaluations and managing evaluated interventions.⁶

Polish municipalities do not have specialized, independent evaluation units, so those managing specific interventions are typically responsible for their evaluation as well. Consequently, our respondents were familiar with the reasons for conducting the evaluations and, as the intended users, were aware of how the studies were used.⁷

We chose interviews over surveys for several reasons. First, face-to-face conversations allowed us to clarify key concepts such as adoption mode and evaluation use, ensuring respondents understood them. This was particularly important in the context of Polish administration, where the evaluation culture is still developing, and awareness of essential evaluation concepts is often limited (Kupiec & Celińska-Janowicz, 2024). Second, interviews provided a better opportunity to explore and

verify various types and examples of evaluation use. While surveys can effectively identify types of evaluation use and describe them through closed-ended questions (e.g., Altschuld, Yoon & Cullen, 1993; Eckerd & Moulton, 2011), only interviews allow for follow-up, asking for stories, examples of specific decisions, and probing for more detail. Additionally, interviews enable the use of prefacing and indirect questions, techniques that help reduce socially desirable responses (Bergen & Labonté, 2020). As a result, interview-based studies on evaluation use factors are not uncommon (e.g., Balthasar, 2006; Ledermann, 2012; Pattyn & Bouterse, 2020).

Our interview protocol, after the introductory questions, consisted of two main sections: the first focused on evaluation adoption mode, and the second on types of evaluation use. Both sections began with general questions like, “Who decided to conduct the evaluation and why?” and “What were the effects and consequences of the evaluation, and how was it used?” These were followed by more specific questions designed to explore and verify four possible adoption modes and four potential types of evaluation use.

As previously noted, our unit of analysis was a series of studies or ongoing evaluation practice within a specific intervention area, such as participatory budgeting or collaboration with nonprofit organizations. During interviews, we began by referencing the most recent evaluation study, but if it was part of a longer series, we encouraged respondents to explain the rationale for starting with the initial study. We applied the same approach to questions about the use of evaluation.

We manually coded the interview transcripts in MaxQDA to identify statements related to specific adoption modes and use types. We assigned weights to each statement based on its significance and credibility. Two researchers conducted this process independently.

Our second, complementary source of information was document analysis, aimed at verifying the claims made during interviews, especially regarding evaluation use. For instance, in verifying instrumental use, we looked for evidence of the reported actions, decisions, or modifications in relevant documents (e.g., strategies, programs, regulations, or action schedules) and connected them to the findings in the evaluation report. We could verify approximately 60% of the significant statements from the interviews through documents, with 95% of these being confirmed. We disregarded the few statements that were refuted when determining the occurrence of adoption modes and use types.

Based on the data from the interviews and documents, we assessed each adoption mode and each type of use on a 3-point scale to create a fuzzy set. The scale was as follows: 0: no evidence of this adoption mode or type of use; 1: Some indications of this adoption mode or type of use, but it is not the only or dominant adoption mode or type of use; 2: clear and convincing indications that this adoption mode or type of use was dominant. Nonuse of evaluation was the exception, as there were no specific questions about it in the interview. Instead, we inferred its presence by the lack of evidence for the other four types of evaluation use. Two researchers independently assigned each score and then collaborated to reach a consensus on the final score.

Findings

The presentation of our findings is structured into several parts. It begins with a description of the frequency and examples of each adoption mode of evaluation practice. Next, we present the types of evaluation use in the same manner. The following parts analyze the relationships between adoption modes and use types from a set-theoretic perspective, focusing first on necessary conditions and then sufficient conditions of evaluation use types.

Adoption Modes

To better understand the motivations behind municipal evaluations, we situate them within the broader context of Polish evaluation culture. Systematic and institutionalized evaluation in Poland

has a relatively short history, beginning in 2004 in response to European Union requirements. Formal requirements to conduct evaluations remain largely limited to EU cohesion policy and the institutions responsible for implementing it at central and regional government levels. Local governments are not bound by any general regulation requiring evaluation or evidence-based policymaking. The only explicit obligation, introduced in 2021, concerns ex-ante evaluations of municipal development strategies (Bienias et al., 2024). Some regulations in specific areas, such as urban regeneration and cooperation with nonprofit organizations, require assessments, but only a few municipalities interpret these as mandates for evaluation. Evidence-based policymaking is not sufficiently institutionalized in both central and local Polish administrations (Kupiec et al., 2020). As a result, municipal evaluation practice is shaped largely by the preferences of local politicians, bureaucrats, and, indirectly, by community expectations.

Our findings indicate that most municipal evaluation practices are voluntary (see Table 2). Respondents frequently cited motivations such as the desire to assess program effectiveness and the need for feedback from aid recipients regarding implementation challenges and potential improvements. Several respondents expressed that evaluation is a logical step for those aiming to foster development and improvement. Additionally, some viewed evaluation as a tool for facilitating dialogue with citizens.

The second most common reason for establishing evaluation practice is coercion, such as formal obligations in funding agreements, official recommendations following audits by the Supreme Audit Office, or explicit expectations from program stakeholders. In one case, stakeholders demanded evaluation as a solution to a crisis in their relationship with local authorities.

Normative sources for adopting evaluation practice were less common and generally not decisive. However, some respondents acknowledged becoming familiar with evaluation through university courses or exposure in previous workplaces. In several cases, municipal evaluation was inspired by collaboration with experts (e.g., from the Batory Foundation) or by observing and experiencing the compulsory evaluation of EU-funded projects.

Adopting evaluation practice through imitation was relatively rare. However, in several cases, observing the practices of other municipalities or regional administrations influenced the decision to initiate evaluation. The participation forum run by Stocznia, a foundation dedicated to fostering evaluation culture, provided a platform where municipalities could learn from the evaluation experiences of others.

It is important to note that evaluation practice is typically driven by a combination of two or three motivations, with voluntary and normative motivations being the most common combination. When multiple motivations are involved, voluntary or coercive factors usually take the lead, while normative or mimetic motivations play a secondary role.

Types of Evaluation Use

Instrumental use is the most common type of evaluation use, observed in over half of the cases studied (see Table 3). Most decisions influenced by evaluation findings involved fine-tuning existing

Table 2. Occurrence of Evaluation Adoption Modes.

Score	Coercive	Mimetic	Normative	Voluntary
0-absent	9	14	11	5
1-present	8	9	11	3
2-dominating	6	0	1	15

Notes: The sample includes 23 cases. Adoption modes could occur in combination (two or more modes could receive nonzero scores in one case). For the details of the scale, see the supplementary materials.

Table 3. Occurrence of Evaluation Use Types.

Score	Instrumental	Conceptual	Legitimizing	Symbolic	Nonuse
0-absent	5	6	9	14	16
1-present	7	12	12	6	5
2-dominating	11	5	2	3	2

Notes: The sample includes 23 cases. Use types could occur in combination (two or more types could receive nonzero scores in one case). For the details of the scale, see the supplementary materials.

interventions and their future iterations, rather than strategic changes. For example, the Cyberbullying Prevention Program initially focused too narrowly on school pupils. In response to the evaluators' recommendation, subsequent editions expanded to include training and information campaigns for senior citizens and parents of preschool children. In another case, the separate Alcohol Problem Solving and Drug Prevention programs were merged into a single long-term program, following evaluation findings that called for better integration and coordination of these efforts.

Conceptual use is also common, though it rarely is the dominant type of use of a particular evaluation. Respondents shared many examples of how evaluation shifted their views on implemented interventions. For instance, two respondents working on the Local Revitalization Program acknowledged that evaluation helped their team realize that revitalization goes beyond infrastructure and must include social activities aimed at promoting social inclusion and engaging residents in the targeted area.

Using evaluation to legitimize municipal programs or specific decisions was also fairly common, though it rarely was the dominant type of use. Legitimizing use often appeared in interactions between intervention managers, the city council, or the media. A typical example involves the evaluation of local initiative centers. The head of the department presented the evaluation findings to city council members on several occasions, helping to build awareness and foster positive attitudes toward the centers. She believed this helped convince decision-makers that the centers needed stable funding from the city budget to thrive.

Symbolic use was the rarest type of use observed. However, we did encounter cases where evaluation was seen as merely an administrative burden or used to give citizens the impression that their opinion was valued, regardless of the findings. A clear example is the cyclical evaluation of the local development strategy, which, according to one respondent, felt unnecessary because a new strategy was already being developed. The evaluation had to be conducted because it is compulsory, so the respondent expressed relief that at least its frequency had been reduced from annual to biennial.

Most evaluations were used in more than one way. The most common combinations were instrumental and conceptual, or instrumental and legitimizing, with instrumental usually being the dominant form.

Necessity Analysis

Atomic necessary conditions—single motivations or adoption modes necessary for a particular type of evaluation use to occur—were identified for only two types of use: instrumental and symbolic (Table 4). For other types of use, combinations (disjunctions) of two or more adoption modes were necessary. The following sections present the results of the necessity analysis in relation to the hypotheses formulated at the start of the study.

Voluntary adoption of evaluation emerged as the only necessary condition for instrumental use. Although a disjunction of voluntary and mimetic adoption showed a slightly higher consistency (0.94, compared to 0.9), its relevance was lower. Mimetic adoption alone was not a necessary condition for instrumental use. Therefore, H1 is only partially supported.

No single adoption mode qualifies as a necessary condition for conceptual use. When conceptual use occurs, it typically indicates that evaluation was adopted through coercion, mimetic behavior, or normative pressures. The disjunction of voluntary and normative adoption, as proposed in Højlund's model, is empirically trivial and undermined by several cases that were inconsistent in kind.⁸ H2 is, therefore, rejected.

Legitimizing use occurs when evaluation practice is adopted normatively or in the absence of coercion. No significant relationship exists between this type of use and mimetic adoption, resulting in the rejection of H3.

The findings suggest that symbolic use occurs when evaluation practice is adopted through coercion. Coercive adoption is the only consistent necessary condition for symbolic use, despite a few cases showing inconsistencies in kind. Therefore, H4 is confirmed.

No necessary condition for the nonuse of evaluation with satisfactory parameters of fit was observed in the analysis. The closest candidate was the normative adoption mode but only in disjunction with the lack of voluntary adoption, and even then, this combination is not empirically relevant. Therefore, H5 was not supported.

Sufficiency Analysis

We identified sufficient solutions with meaningful consistency for all types of evaluation use, except nonuse (Table 5). However, some solutions are quite complex, involving combinations of different adoption modes.

Voluntary adoption of evaluation practice serves as a sufficient condition for instrumental use, as it consistently leads to instrumental use across all cases. Another sufficient condition for instrumental use is mimetic motivation, but only when normative motivation is absent. However, this solution is uncommon, explaining only four cases, including one unique case.⁹

No single adoption mode is sufficient for conceptual use to occur; however, three combinations of conditions can lead to this type of evaluation use. The most empirically relevant combination involves voluntary and normative motivations. Additionally, the combinations of voluntary with mimetic and coercive with mimetic also result in conceptual use.

The explanation for legitimizing use is even more complex. Similar to conceptual use, legitimizing use is likely when both voluntary and normative motivations precede the evaluation. However,

Table 4. Analysis of Necessary Conditions.

Necessary condition(s)*	Outcome**	inclN***	RoN***
Vol	Ins	0.900	0.719
Vol + Mim		0.943	0.696
Mim + Nor	Con	0.838	0.941
Cor + Mim		0.838	0.666
Cor + Nor		0.903	0.637
~Cor + Nor	Leg	0.878	0.680
Cor	Sym	0.828	0.670
—	Non	—	—

Notes: *Cor = coercive; Mim = mimetic; Nor = normative; Vol = voluntary; "+" operator means disjunction, i.e., condition x or condition y must be met for the outcome to occur; "~" operator means negation, i.e., the absence of a particular condition; disjunctions with the absence of condition (~) are not included unless it is the only consistent and relevant combination for a particular type of use. **Outcomes: Ins = instrumental; Con = conceptual; Leg = legitimizing; Sym = symbolic; Non = nonuse. ***Only conditions (and disjunctions of conditions) with consistency (inclN) > 0.8 and relevance of necessity (RoN) > 0.6 are included.

for this solution to be sufficiently consistent, coercive motivation must also be present, or mimetic motivation must be absent (Con*Nor*Vol + ~Mim*Nor*Vol).

Table 5. Analysis of Sufficient Conditions.

Sufficient term*	Outcome**	inclS***	PRI***	covS
Vol	Ins	0.853	0.827	0.900
Mim*~Nor		0.954	0.925	0.287
Cor*Mim	Con	1.000	1.000	0.280
Mim*Vol		0.897	0.843	0.428
Nor*Vol		0.929	0.901	0.641
Cor*Nor*Vol	Leg	0.870	0.787	0.410
~Mim*Nor*Vol		0.840	0.741	0.530
Cor*~Nor*~Vol	Sym	0.892	0.846	0.685
—	Non	—	—	—

Notes: *Cor = coercive; Mim = mimetic; Nor = normative; Vol = voluntary; “*” operator means conjunction or intersection, i.e., presence of both x and y ensure that the outcome occurs; “~” operator means negation, i.e., the absence of a particular condition. **Outcomes: Ins = instrumental; Con = conceptual; Leg = legitimizing; Sym = symbolic; Non = nonuse. ***Only solutions with inclusion score (inclS) and proportional reduction in inconsistency (PRI) > 0.8 are included.

Coercive adoption is present in all cases of symbolic evaluation use. However, because evaluation practice often results from a combination of motivations, coercive adoption alone is not sufficient unless both normative and voluntary motivations are absent. No sufficient conditions were identified for the nonuse of evaluation.

Discussion

Based on the findings, several observations can be made about the relevance of the tested model in the context of evaluation practice in local government. Højlund (2014) posits that each adoption mode may lead to multiple types of evaluation use, and that nearly every type of evaluation use can stem from various adoption modes. This assumption aligns with the results of this study. Additionally, evaluation practice often arises from a combination of motivations, and evaluations may be used in several ways. However, only some of our observations match Højlund’s specific predictions (see Table 6). While voluntary adoption indeed leads to instrumental use, we found no evidence that mimetic adoption results in the same. As expected, normative adoption is linked to conceptual use, but coercive and mimetic adoption modes also contribute (often in combination). Contrary to Højlund’s prediction, voluntary adoption does not result in conceptual use. Similarly, the predicted relationship between mimetic adoption and legitimizing use was not confirmed. The clearest case is symbolic use—both the model and our empirical findings show that it results from coercive adoption of evaluation practice.

Table 6. Relationships Between Evaluation Adoption Modes and Evaluation Use Types.

Type of evaluation use	Instrumental	Conceptual	Legitimizing	Symbolic	Nonuse
Predictions from Højlund’s model (2014)	Vol or Mim	Vol or Nor	Mim	Cor	Cor or Nor
Necessary conditions identified in this study*	Vol Vol + Mim	Mim + Nor Cor + Mim Cor + Nor	~Cor + Nor	Cor	—

Notes: Cor = coercive; Mim = mimetic; Nor = normative; Vol = voluntary; “+” operator means disjunction, i.e., condition x or condition y must be met for the outcome to occur; “~” operator means negation, i.e., the absence of a particular condition.

An intriguing observation is the scarcity of cases where mimetic or normative adoption modes are dominant. Although these motivations are not uncommon, based on our respondents' accounts, they almost never serve as the direct reason for initiating evaluation. Mimetic factors, such as observing the practices of other municipalities, and normative influences, like feedback from consultants or experience from a previous workplace, can raise awareness of evaluation and create a supportive environment. However, starting an evaluation process typically requires a direct trigger, which is almost always coercive or voluntary in nature. In this respect, contrary to the theoretical model, mimetic and normative adoption modes do not seem to function as equal alternatives to voluntary and coercive motivations.

Although this study focuses on local administration in a country with a less mature evaluation culture, we believe this conclusion could be relevant to public administration more generally, though further evidence is needed. One possible reason public sector organizations do not simply imitate evaluation practice without additional motivation could be that they do not view uncertainty as a threat to their survival in the same way private sector organizations do. DiMaggio and Powell (1983) introduced the concept of mimetic isomorphism with private sector organizations in mind.

When operationalizing outcomes, it was not clear whether we should distinguish between the symbolic and nonuse of evaluation. Although the theoretical model separates these concepts, every evaluation is conducted for a reason and is meant to address organizational needs. From an organizational institutionalist perspective, which underpins Højlund's model, this reason often relates to external requirements or expectations. The mere act of conducting an evaluation, whether intentional or not, legitimizes the organization or the intervention being evaluated. It creates or maintains the appearance of appropriateness, alignment, or compliance with expectations. In this sense, the distinction between symbolic use and nonuse is minimal.

While nonuse may seem like a convenient concept when contrasted with instrumental or conceptual use—where a lack of decisions based on evaluation findings or no change in perception clearly constitutes nonuse—the line between symbolic use and nonuse appears blurred. It is worth noting that other evaluation use frameworks grounded in organizational theory (such as Carman, 2005 and Eckerd & Moulton, 2011) do not treat nonuse as a separate category. The unclear relationship between nonuse and other types of use becomes evident in how nonuse is defined by absence, while all other types are defined by presence. Unlike other types of use, which can coexist, nonuse stands alone. The distinct nature of nonuse likely contributed to the inability to identify either necessary or sufficient conditions for it.

Højlund's model has an appealing simplicity, but it limits its explanatory potential. The specific limitations we identified include a narrow view of evaluation use, treating the organization as a uniform unit of analysis, a lack of consideration for dynamics, and a focus on nonexistent ideal types. We briefly discuss these issues and their implications below.

The categorization of use types in the model, as potential outcomes of the evaluation process, raises concerns about the concept of use itself, which does not fully capture the broader and intangible effects of evaluation (e.g., Alkin & Taut, 2002; Henry & Mark, 2003). Kirkhart's integrated theory of influence (2000) suggests that by focusing on use, the model addresses only part of the potential evaluation effects. Højlund acknowledges the broader concept of influence but argues that a simple typology of evaluation use served a specific purpose in his model, although this has consequences. What we captured and linked to motivations were mostly examples of intended use at the end of an evaluation cycle. Although our aim was to analyze streams of evaluation studies over time rather than single studies, most identified cases of use occurred shortly after the conclusion of a specific study. The fact that intended use is easier to identify may explain why instrumental use was the most frequently observed in our study. It also suggests that the relationships between voluntary motivation and instrumental use, as well as coercive motivation and symbolic use, may not be as clear in practice as they appeared in our findings. While the simplified understanding of evaluation

use might be seen as a limitation of the tested model, it is also typical of frameworks that examine organizational factors influencing evaluation use (see the overview in Kupiec et al., 2023).

Another simplification of Højlund's model is its treatment of the organization as a uniform entity, a single unit of analysis. While disproving this was not the goal of the study, such an observation was made. Respondents who described their evaluations as voluntary gave different answers when asked who specifically made the decision. Their responses ranged from city councils, mayors, and heads of department to simply "we." Even if the approval procedure was similar in these cases, the differences in perception could lead to variations in evaluation use, especially since it was the intended users' perceptions that varied.

Following Thompson's (1967) concept of three levels of responsibility within organizations, when the voluntary decision to adopt an evaluation practice is believed to come from the institutional level (city council, mayor), it might be perceived as coerced at the managerial and technical levels (department staff or supervised units). Moreover, regardless of the adoption mode, different types of evaluation use are more likely to occur at different organizational levels. At the technical, rational, and closed level, only instrumental use of operational recommendations is expected, while at managerial and institutional levels, other types of use should also emerge.

The need to "unpack" an organization to understand the change mechanisms triggered by evaluation aligns with the idea of three levels of evaluation influence (Henry & Mark, 2003). This concept suggests that evaluation can influence thoughts or actions at three levels: the individual, interactions between individuals, and the organization's overall decisions and practices.

A further limitation of the model is its static nature. It does not account for the potential evolution of evaluation practice, evaluation use, or the perception of why evaluations are conducted. The model assumes that the primary type of evaluation use remains constant and is determined by the initial context of adoption. However, the literature suggests that evaluation practice and types of use can evolve as evaluation capacity grows, often shifting from symbolic compliance toward instrumental and conceptual uses (e.g., Gibbs et al., 2002; House et al., 1996). This lack of dynamism has also been noted as a feature of several other evaluation use frameworks based on organizational theory, as highlighted in a recent review (Kupiec et al., 2023).

Our findings also suggest that the evolution of evaluation practice should be accounted for, though the changes we observed moved in the opposite direction. In several cases, evaluation was initially adopted voluntarily but later became a formal routine. When asked why the most recent evaluation was conducted, some respondents explained that it was due to custom or even an obligation outlined in internal documents. In another case, a respondent noted that while they initially adopted evaluation because they believed in its utility, it has since become so widespread and established as a tool for dialogue with nongovernmental organizations that they could not make decisions without first consulting through evaluation. These examples suggest that a voluntarily adopted evaluation practice may, over time, become routine and perceived as obligatory, which could impact its use. More broadly, this indicates that evaluation practice, or at least perceptions of it, can evolve over time.

The static nature of the model stems from its reliance on a single variable to explain evaluation use, even though this variable reflects several characteristics of the organizational environment and the organization's role. However, the literature points to numerous other potential factors (see the review of Cousins & Leithwood [1986] as an example) related to users, evaluators, and evaluation as both process and product (Alkin & King, 2017). Højlund refers to these as immediate, microlevel factors and argues that they are less significant than the underlying, medium-level organizational context factors. This perspective, though not always explicitly stated, appears to be shared by other scholars studying the organizational context of evaluation and knowledge use (e.g., Eckerd & Moulton, 2011; Lall, 2015; Rimkutė, 2015).

It is reasonable to believe that certain immediate factors—such as timeliness, evaluator competence, evaluation quality, and the characteristics and receptiveness of users—are shaped by more

fundamental organizational factors. However, an equally significant set of immediate factors is tied to the specific context of each evaluation study, reflecting the current situation within the organization and its environment. These factors are not influenced by the historical context of adopting evaluation practice and are not addressed by the model. They include the nature of the findings, the importance and type of decision at hand, and competing information.¹⁰ From the model's perspective, these factors create noise and may be responsible for cases we labeled as inconsistent in our analysis.

Lastly, the model defines four ideal types, also referred to as "extreme positions." As expected, most of the cases we studied did not fit these ideal types. In fact, many were quite far from them, as evaluation practice often stemmed from a combination of two or more motives, and evaluation studies were used in multiple ways. Qualitative comparative analysis, designed to handle such complex relationships, still enabled us to identify conditions for each type of use. However, this raises questions about the practical utility of a model based on ideal types in a reality where, in many instances, it is neither possible to distinguish a single pure adoption mode nor identify one primary type of evaluation use.

The issues discussed above relate to the logic of the model being tested. Our research design may also raise concerns about social desirability bias. When relying on respondents' testimonies, we must acknowledge that, in some cases, they may provide answers they believe are appropriate, even if not entirely accurate. In our study, voluntary adoption and instrumental use likely represented the most "appropriate" responses, and indeed, these were the most common. To mitigate this bias, we assured respondents of anonymity and confidentiality; normalized honest "incorrect" responses; and encouraged contextualizing and providing examples. Interviewers did not observe any clear signs of social desirability bias.¹¹ Additionally, where possible, we verified responses through document analysis, which was most effective in confirming instrumental use, with nearly all cases validated.

Regarding adoption mode, we know that formal requirements for local government in Poland are minimal. Thus, the prevalence of voluntary adoption and instrumental use in our sample may be partly explained by self-selection. It is likely that individuals who take evaluation seriously and use it because they find it useful were more inclined to participate in our study. It may also have been easier for these individuals to demonstrate positive evaluation outcomes.

The interpretation of our findings must also consider the aforementioned concept of organizations as multilevel, complex entities. In this context, it is important to note that our respondents were typically heads of department.¹² While distinguishing between the technical and managerial levels within city hall departments may be challenging (using Thompson's [1967] terminology), it is clear that we did not capture the institutional-level perspective (i.e., city councils and mayors).¹³ This likely had little effect on the perception of the adoption mode—respondents generally viewed evaluations as their own and voluntary, even when the formal decision came from the mayor. However, the absence of the institutional level may have had a greater impact on the identification of use types, as more cases of legitimizing and symbolic use might have emerged if that level had been included. Given this, our results should be seen as contextual, reflecting the managerial and technical levels, but they should be verified at the institutional level in future studies.

Conclusions

This study explored the relationship between the types of evaluation use observed in local governments and the adoption modes of evaluation practice, using Højlund's (2014) theoretical model, which is based on the neoinstitutional concept of organizational isomorphism, to assess its validity.

While our empirical findings indicate that the model's explanatory potential is limited, we confirmed some significant relationships between adoption modes and types of evaluation. For example, evaluation practice adopted voluntarily typically leads to instrumental use, while coerced evaluations result in symbolic use. We also found that evaluation practices driven by a single motivation are rare; in most

cases, a combination of two or three motivations is present. Additionally, mimetic and normative factors, though present, are almost never the immediate triggers for evaluation.

Given that this study was conducted within the specific context of local government in a country with a relatively low maturity of evaluation culture, further research in different administrative settings is necessary to verify the relationship between the motivation for adopting evaluation practice and the dominant type of evaluation use in public sector organizations.

Acknowledgments

Wiktoria Abramczyk, Maria Antoniak, and Robert Bengsz, who were at the time studying at the University of Warsaw, EUROREG, helped the authors to conduct the interviews.


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) disclosed receipt of the following financial support for the research and/or authorship of this article: This work was supported by the National Science Centre (Narodowe Centrum Nauki), Poland, grant number 2019/33/B/HS5/01336.

ORCID iDs

Tomasz Kupiec  <https://orcid.org/0000-0002-6469-7746>

Zuzanna Wrońska  <https://orcid.org/0009-0004-9654-2409>

Supplemental Material

Supplemental material for this article is available online.

Notes

1. The next section provides definitions of all the listed adoption modes and use types.
2. An overview of existing frameworks can be found in Kupiec et al. (2023).
3. This means it is best explained in terms of set membership, following the principles of set theory (Ragin, 2009).
4. These abbreviations are used throughout the rest of the paper.
5. Doubts related to the concept of non-use as a type of evaluation use are described in “Discussion” section.
6. Earlier studies revealed that the head of department is the highest level in local government administration in Poland at which there is awareness of evaluation and how it is used (Kupiec, 2015).
7. See the supplementary materials for more details on respondent characteristics and recruitment procedure, document analysis, the 3-point scale used to assess adoption modes, use types, and the creation of a fuzzy set for QCA.
8. For an explanation of the difference between consistency in kind and consistency in degree, see Oana et al. (2021).
9. A unique case is one that is not covered by other sufficient conditions.
10. Factors listed here are adapted from the classification proposed by Cousins and Leithwood (1986) and then refined by Johnson et al. (2009).
11. Such as providing partial or vague answers, showing nervous facial expressions and body language, and inconsistent use of vocabulary (Bergen & Labonté, 2020).
12. Or chief specialists expressing the views of their direct superiors.
13. This choice was already explained in the research design section.

References

- Ágh, A. (2016). The decline of democracy in East-Central Europe: Hungary as the worst-case scenario. *Problems of Post-Communism*, 63(5–6), 277–287. <https://doi.org/10.1080/10758216.2015.1113383>
- Ahonen, P. (2015). Aspects of the institutionalization of evaluation in Finland: Basic, agency, process and change. *Evaluation*, 21(3), 308–324. <https://doi.org/10.1177/1356389015592546>
- Alkin, M. C. (1985). *A guide for evaluation decision-makers*. Sage.
- Alkin, M. C., & King, J. A. (2017). Definitions of evaluation use and misuse, evaluation influence, and factors affecting use. *American Journal of Evaluation*, 38(3), 434–450. <https://doi.org/10.1177/1098214017717015>
- Alkin, M. C., & Taut, S. M. (2002). Unbundling evaluation use. *Studies in Educational Evaluation*, 29(1), 1–12. [https://doi.org/10.1016/S0191-491X\(03\)90001-0](https://doi.org/10.1016/S0191-491X(03)90001-0)
- Altschuld, J. W., Yoon, J. S., & Cullen, C. (1993). The utilization of needs assessment results. *Evaluation and Program Planning*, 16(4), 279–285. [https://doi.org/10.1016/0149-7189\(93\)90040-F](https://doi.org/10.1016/0149-7189(93)90040-F)
- Balthasar, A. (2006). The effects of institutional design on the utilization of evaluation: Evidenced using qualitative comparative analysis (QCA). *Evaluation*, 12(3), 353–371. <https://doi.org/10.1177/1356389006069139>
- Berg-Schlosser, D., De Meur, G., Rihoux, B., & Ragin, C. (2008). Qualitative comparative analysis (QCA) as an approach. In B. Rihoux & C. Ragin (Eds.), *Configurational comparative methods. Qualitative comparative analysis (QCA) and related techniques* (pp. 1–18). Sage. <https://doi.org/10.4135/9781452226569>
- Bergen, N., & Labonté, R. (2020). ‘Everything is perfect, and we have no problems’: Detecting and limiting social desirability bias in qualitative research. *Qualitative Health Research*, 30(5), 783–792. <https://doi.org/10.1177/1049732319889354>
- Bienias, S., Kupiec, T., & Strzęboszewski, P. (2024). *Poradnik ewaluacji polityki rozwoju*. Ministerstwo Funduszy i Polityki Regionalnej.
- Boswell, C. (2008). The political functions of expert knowledge: knowledge and legitimation in European Union immigration policy. *Journal of European Public Policy*, 15(4), 471–488. <http://dx.doi.org/10.1080/13501760801996634>
- Brunsson, N. (2002). *The organization of hypocrisy: Talk, decisions and actions in organizations*. Abstrakt & Liber.
- Carden, F. (2013). Evaluation, not development evaluation. *American Journal of Evaluation*, 34(4), 576–579. <https://doi.org/10.1177/1098214013495706>
- Carman, J. G. (2005). *Program evaluation use and practice: A study of nonprofit organizations in New York State*. State University of New York Press.
- Carman, J. G. (2011). Understanding evaluation in nonprofit organizations. *Public Performance & Management Review*, 34(3), 350–377. <https://doi.org/10.2307/41104065>
- Christie, C. A. (2007). Reported influence of evaluation data on decision makers’ actions: An empirical examination. *American Journal of Evaluation*, 28(1), 8–25. <https://doi.org/10.1177/1098214006298065>
- Cousins, J. B., & Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research*, 56(3), 331–364. <https://doi.org/10.3102/00346543056003331>
- Dahler-Larsen, P. (2012). *The evaluation society*. Stanford Press. <https://doi.org/10.11126/stanford/9780804776929.001.0001>
- Davies, P. (2012). The state of evidence-based policy evaluation and its role in policy formation. *National Institute Economic Review*, 219(1), R41–R52. <https://doi.org/10.1177/002795011221900105>
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147–160. <https://doi.org/10.2307/2095101>
- Eckerd, A., & Moulton, S. (2011). Heterogeneous roles and heterogeneous practices: Understanding the adoption and uses of nonprofit performance evaluations. *American Journal of Evaluation*, 32(1), 98–117. <https://doi.org/10.1177/1098214010381780>
- Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, 12(2), 219–245. <https://doi.org/10.1177/1077800405284363>

- Furubo, J. E., Rist, R. C., & Sandahl, R. (2002). *International atlas of evaluation*. Transaction Publishers.
- Gibbs, D., Napp, D., Jolly, D., Westover, B., & Uhl, G. (2002). Increasing evaluation capacity within community-based HIV prevention programs. *Evaluation and Program Planning*, 25(3), 261–269. [https://doi.org/10.1016/S0149-7189\(02\)00020-4](https://doi.org/10.1016/S0149-7189(02)00020-4)
- Greckhamer, T., Misangyi, V. F., & Fiss, P. C. (2013). The two QCAs: From a small-N to a large-N set theoretic approach. In *Configurational theory and methods in organizational research* (Vol. 38, pp. 49–75). Emerald Group Publishing Limited. [https://doi.org/10.1108/S0733-558X\(2013\)0000038007](https://doi.org/10.1108/S0733-558X(2013)0000038007)
- Hanberger, A. (2011). The real functions of evaluation and response systems. *Evaluation*, 17(4), 327–349. <https://doi.org/10.1177/1356389011421697>
- Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*, 24(3), 293–314. <https://doi.org/10.1177/109821400302400302>
- Højlund, S. (2014). Evaluation use in the organizational context—changing focus to improve theory. *Evaluation*, 20(1), 26–43. <https://doi.org/10.1177/1356389013516053>
- House, E. R., Haug, C., & Norris, N. (1996). Producing evaluations in a large bureaucracy. *Evaluation*, 2(2), 135–150. <https://doi.org/10.1177/135638909600200202>
- Johnson, K., Greenseid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation*, 30(3), 377–410. <https://doi.org/10.1177/1098214009341660>
- King, J. A., & Alkin, M. C. (2019). The centrality of use: Theories of evaluation use and influence and thoughts on the first 50 years of use research. *American Journal of Evaluation*, 40(3), 431–458. <https://doi.org/10.1177/1098214018796328>
- Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 2000(88), 5–23. <https://doi.org/10.1002/ev.1188>
- Kupiec, T. (2015). Program evaluation use and its mechanisms, the case of cohesion policy in Polish regional administration. *Zarządzanie Publiczne*, 33(3), 67–83. <https://doi.org/10.15678/ZP.2015.33.3.05>
- Kupiec, T. (2022). Does evaluation quality matter? Quantitative analysis of the use of evaluation findings in the field of cohesion policy in Poland. *Evaluation and Program Planning*. <https://doi.org/10.1016/j.evalprogplan.2022.102103>.
- Kupiec, T., & Celińska-Janowicz, D. (2024). Praktyka ewaluacji interwencji publicznych w miastach na prawach powiatu. *Studia Regionalne I Lokalne*, 94(4), 66–82. <https://doi.org/10.7366/1509499549405>
- Kupiec, T., Celińska-Janowicz, D., & Pattyn, V. (2023). Understanding evaluation use from an organisational perspective: A review of the literature and a research agenda. *Evaluation*, 29(3), 338–355. <https://doi.org/10.1177/13563890231185164>
- Kupiec, T., Wojtowicz, D., & Olejniczak, K. (2020). Between compliance and systemic change – evaluation practice in eight CEE countries. In S. Mazur (Ed.), *Public administration in central Europe. Ideas as causes of reforms* (pp. 129–146). Routledge. <https://doi.org/10.4324/9780429286452-11>
- Lall, S. A. (2015). Measuring to improve vs. measuring to prove: Understanding evaluation and performance measurement in social enterprise. [Unpublished PhD thesis. George Washington University]. <https://doi.org/10.1007/s11266-017-9898-1>
- Ledermann, S. (2012). Exploring the necessary conditions for evaluation use in program change. *American Journal of Evaluation*, 33(2), 159–178. <https://doi.org/10.1177/1098214011411573>
- Leeuw, F. L., & Furubo, J. E. (2008). Evaluation systems: What are they and why study them? *Evaluation*, 14(2), 157–169. <https://doi.org/10.1177/1356389007087537>
- Leviton, L. C., & Hughes, E. F. (1981). Research on the utilization of evaluations: A review and synthesis. *Evaluation Review*, 5(4), 525–548. <https://doi.org/10.1177/0193841X8100500405>
- Marx, A., & Dusa, A. (2011). Crisp-set qualitative comparative analysis (csQCA), contradictions and consistency benchmarks for model specification. *Methodological Innovations Online*, 6(2), 103–148. <https://doi.org/10.4256/mio.2010.0037>

- Mello, P. A. (2012). *A critical review of applications in QCA and fuzzy-set analysis and a "Toolbox" of proven solutions to frequently encountered problems*. APSA Annual Meeting.
- Meuer, J., & Rupiotta, C. (2017). A review of integrated QCA and statistical analyses. *Quality & Quantity*, 51, 2063–2083. <https://doi.org/10.1007/s11135-016-0397-z>
- Meyer-Sahling, J. H. (2009). Varieties of legacies: A critical review of legacy explanations of public administration reform in East Central Europe. *International Review of Administrative Sciences*, 75(3), 509–528. <https://doi.org/10.1177/0020852309337670>
- Norris, N., & Kushner, S. (2007). The new public management and evaluation. In N. Norris & S. Kushner (Eds.), *Dilemmas of Engagement: Evaluation and the New Public Management* (pp. 1–16). Elsevier. [https://doi.org/10.1016/S1474-7863\(07\)10001-6](https://doi.org/10.1016/S1474-7863(07)10001-6)
- Oana, I. E., Schneider, C. Q., & Thomann, E. (2021). *Qualitative comparative analysis using R: A beginner's guide*. Cambridge University Press. <https://doi.org/10.1017/9781009006781>
- Pattyn, V., & Bouterse, M. (2020). Explaining use and non-use of policy evaluations in a mature evaluation setting. *Humanities and Social Sciences Communications*, 7(1), 1–9. <https://doi.org/10.1057/s41599-020-00575-y>
- Pelz, D. C. (1978). Some expanded perspectives on use of social science in public policy. In J. M. Yinger & S. J. Cutler (Eds.), *Major social issues: A multidisciplinary view* (pp. 346–357). Palgrave Macmillan.
- Piccioletto, R. (2016). Evaluation and bureaucracy: The tricky rectangle. *Evaluation*, 22(4), 424–434. <https://doi.org/10.1177/1356389016657934>
- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29(4), 443–459. <https://doi.org/10.1177/1098214008324182>
- Ragin, C. C. (2009). Redesigning social inquiry: Fuzzy sets and beyond. *Social Forces*, 88(4), 1936–1938. <https://doi.org/10.7208/chicago/9780226702797.001.0001>
- Raimondo, E. (2018). The power and dysfunctions of evaluation systems in international organizations. *Evaluation*, 24(1), 26–41. <https://doi.org/10.1177/1356389017749068>
- Rihoux, B., & Ragin, C. C. (Eds) (2009). *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*. Sage. <https://doi.org/10.4135/9781452226569>
- Rimkutė, D. (2015). Explaining differences in scientific expertise use: The politics of pesticides. *Politics and Governance*, 3(1), 114–127. <https://doi.org/10.17645/pag.v3i1.82>
- Rist, R., & Stame, N. (Eds) (2006). *From studies to streams: Managing evaluative systems*. Transaction Publishers. <https://doi.org/10.4324/9780203791189>
- Sanderson, I. (2000). Evaluation in complex policy systems. *Evaluation*, 6(4), 433–454. <https://doi.org/10.1177/13563890022209415>
- Schneider, C. Q., & Wagemann, C. (2012). *Set-theoretic methods for the social sciences: a guide to qualitative comparative analysis*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139004244>
- Shulha, L. M., & Cousins, J. B. (1997). Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice*, 18(3), 195–208. [https://doi.org/10.1016/S0886-1633\(97\)90027-1](https://doi.org/10.1016/S0886-1633(97)90027-1)
- Thomann, E., & Maggetti, M. (2020). Designing research with qualitative comparative analysis (QCA): Approaches, challenges, and tools. *Sociological Methods & Research*, 49(2), 356–386. <https://doi.org/10.1177/0049124117729700>
- Thompson, J. D. (1967). *Organizations in action*. McGraw-Hill. <https://doi.org/10.4324/9781315125930>
- Van der Knaap, P. (1995). Policy evaluation and learning: Feedback, enlightenment or argumentation? *Evaluation*, 1(2), 189–216. <https://doi.org/10.1177/135638909500100205>
- Weaver, C. (2007). The world's bank and the bank's world. *Global Governance*, 13(4), 493–512. <https://doi.org/10.1163/19426720-01304005>
- Weiss, C. H. (1972). Utilization of evaluation: Toward comparative study. In C. H. Weiss (Ed.), *Evaluating action programs: Readings in social action and education* (pp. 318–326). Allyn and Bacon.