Procedural institutionalisation of the evaluation through legal basis: A new typology of evaluation clauses in Switzerland

Damien Wirths

University of Lausanne, Swiss Graduate School in Public Administration (IDHEAP)
damien.wirths@unil.ch

Abstract

‘Evaluation clauses’ are a legal basis requiring mandatory evaluation of public policies’ impact. Evaluation becomes a procedural institutionalization but, although they are widespread, no phrasing standards have been yet established. Without clarification, the questions raised by origins and effects of mandatory evaluation remained impossible to treat. Past studies dealing with this topic dismissed the legal feature of the evaluation process to focus only on organisations. This paper aims to apply a methodology based on the numerical taxonomy to construct a conceptual scheme for classifying evaluation clauses. A hierarchical cluster analysis was run on 319 cases with eight variables related to their normative density. From such heterogeneous legal object, two families of clauses have emerged, and three significant templates are identified. This paper opens a research track which considers evaluation clauses to investigate both the reason why a specific type of phrasing is chosen and the consequences of these institution types.

I. Introduction

Evaluation is a crucial (but often forgotten) part of each policy-making process. However, during the past decades, scholars increasingly show (Bussmann, 2008; Jacob, 2005; Spinatsch, 2002) that, either public authorities are created (organisational institutionalisation) or prescriptions are enacted (procedural institutionalisation) to mandatory evaluate the impact of public policies. Both aspects (Mader, 1990) are part of a process of institutionalisation of evaluation which is not only related to the spreading of evaluation activities among States but also to its reinforcement during the policy-making process (Mader, 1985). Although evaluation institutionalisation has led to a growing literature, especially in comparative politics (Furubo et al., 2002; Jacob and Varone, 2003; Jacob, 2005), the procedural aspect and legal features have been systematically dismissed by scholar to focus on organisations (Varone and Jacob, 2004).

This paper investigates evaluation clauses, which are a way of institutionalizing evaluation legally. Evaluation clauses are a legal basis requiring a mandatory review of a public policy or of a measure (Bussmann, 1997; FOJ, 2007; Mader, 1985). Evaluation

*I would like to thank Prof. Eugene Horber for his valuable help regarding the methodology and PhD. Christian Roser and Johanna Schnabel for their helpful comments. This paper is part of SynEval project that analyses the relationship between different attributes of political systems and the practice and institutionalisation of policy evaluation (http://www.syneval.ch). It was supported by the Sinergia program of the Swiss National Science Foundation (grant CRSII1 141893/1).
clauses are considered here as institutions in the sense of explicit formal rules that are "developed and applied by a group of actors so as to structure repetitive activities that produce (if possible predictable) results involving these actors and potentially other actors" (Larrue et al., 2011). A recent study showed that 20.8% of parliamentarians have proposed such a kind of evaluation request at least once (Bundi et al., 2014). This mandatory evaluation can be realized with several types of information instruments related to the implementation of public policies. Since evaluation is not a fixed term, it has to be distinguished, on the one hand, the evaluation clauses considered as a means to investigate the policy effectiveness (Evaluation in a strict sense) and, on the other hand, different kinds of information instruments highlighting process efficiency (controlling) or providing statistics (monitoring). For the purpose of this paper, only the first category of instruments is considered since the other ones are not evaluation but rather reporting (FAO, 2011; FOJ, 2004, 2005).

Although mandatory evaluations are widespread in post-modern States, several governmental studies have highlighted that no standards have been established in evaluation clauses phrasing and they represent a very heterogeneous legal basis (Bussmann, 1997; FAO, 2011; FOJ, 2005). Evaluation of law effects needs a rigorous implementation and several aspects need to be clarified: who has to evaluate what, when, in which purpose, under which criteria and which product does he has to produce for who. In this respect, how evaluation is institutionalized (expectation, rules of the game, object to evaluate, etc.) is a crucial factor that requires special attention (Jacob and Varone, 2002). Unless the edges of evaluation clauses are clarified, in spite of pioneering works on the evaluation of law effects (Mader, 1985, 1990, 1994; Morand, 1993, 1999, 2001), the questions surrounding origins and effects of mandatory evaluation remained impossible to treat. The paper aims to fill this knowledge gap proposing an empirical classification of 319 evaluation clauses existing in Switzerland. Applying numerical taxonomy to the field of legal studies, three templates can be identified through a hierarchical cluster analysis applied to a recent census of evaluation clauses.

The paper is structured as follow. The first section provides a more thorough definition and delimitation of the research object. Then, the second section operationalizes the characteristics of the evaluation clauses and deduces the expectations related to a classification attempt. The third section presents the applied method and is followed by the validation and the interpretation of the results. Finally, findings are discussed in terms of research agenda and new research track for the fields of legal studies, institutionalism and evaluation.

II. THE EVALUATION CLAUSES AS A RESEARCH OBJECT

Although this form of evaluation institutionalisation has been well established in the United Stated for a long time (widespread practice since the 80’s), it can be observed only a rudimentary use in Europe (Mader, 1985). However, in Switzerland, such kind of legal basis has grown dramatically since the first clause introduced in 1984 (Bussmann, 2008) up to the constitutional level (Aubert, 2003; Bussmann, 2007; Mader, 2005; Morand, 2001). At the federal level, the Swiss federal office for justice (FOJ) has

1From Federal office for Justice (FOJ) and Federal audit office (FAO)
inventoried 130 clauses.

Evaluation clauses are formal institutions impacting the final step of a policy making process. They are a key features in the study of legislative evaluation (both for the field of law and of evaluation) since they determinate which stakeholder will be concerned by the evaluation (as evaluated or evaluator), their roles and how they have to interact to implement the evaluation. They are the framework of the evaluation and, as "rules of the game" (Steinmo, 2006: 293), their study allows for addressing the three seminal questions of new-institutionalism: how do actors behave, what do institutions do, and why do institutions persist over time? (Hall & Taylor, 1997: 472). Consequently, their theoretical and practical significance is twofold.

First, the study of the institutions effects, taking institutions as fixed and exogenous. Here, "institutions are modelled via their effects on the set of action available to each individual, on the sequence of actions and on the structure of information available to each decision-maker" (Weingast, 1996: 169). The introduction of an evaluation clause is a variable to explain in investigations focused on the political generation of law and on the reason why public authorities have specifically introduced efficiency evaluation obligations. Several factors can influence the lawmaking process to determinate a certain kind of rule (more or less restrictive for stakeholders). Evaluation clauses are considered as an explicative variable for researches focused on policy implementation and on whether and how evaluation obligations are correctly implemented. Can be highlighted a better implementation of the evaluation (and a better impact of the entire policy) with a certain kind of rule?

Second, analysts can also study why institutions take particular forms, instead of another, thus allowing institutions to be endogenous (Weingast, 1996, p. 175). For instance, it can be argued that responsible federal authorities would have an interest in proposing very imprecise rules in order to not critically question their activities and either "using an evaluation to try and legitimise" or "interpreting evaluation clauses in a way that is biased towards their own interests and to use the results of the evaluation to their advantage in an opportunistic manner in the political process (FAO, 2011).

Even though, in both cases, the evaluation clauses are key variable to understand the role of the evaluation in the policymaking process, their great heterogeneity has been an obstacle for research. As a result, researchers have often giving up on treating evaluation clause when institutionalisation of the evaluation is conceptualized (Varone and Jacob, 2004). Nevertheless, evaluation clauses are very widespread in Europe and in the United States and they represent a major form of evaluation institutionalisation which impact the entire evaluation process. Moreover, the increasing number of research projects related to institutionalisation of public policy evaluation (especially in Switzerland) triggers the need for solving the "heterogeneity problem".

Bussmann (2005) attempted to define evaluation clauses and he proposed a complex typology which has raised two problems. The first methodological problem being that the classification is based on very few cases (55) inventoried by the FOJ and the data are inadequate on the federal as well as on the cantonal level. The second problem is more theoretical and there is no evidence about the significance of the variables between groups and the similarity within group. For example, a first look on recent data shows that only 4% of the inventoried evaluation clauses are ex ante and it
can be suspected that this dichotomous dimension (ex ante/ex post) is probably not significantly related to clause classification. It means that the Bussmann’s classification is thus unidimensional (Collier et al., 2008) and reduces the evaluation clauses to only one categorical variable organized around a single dimension. A monothetic classification is obtained with as many classes as the variable “evaluation object” attributes and, consequently, a usable classification of evaluation clauses is still missing.

How the paper fills this knowledge gap is twofold. First, more data are currently available thanks to a recent census of the evaluation clauses in Switzerland (Wirths and Horber-Papazian, 2015). A new typology can now be established on 319 cases. Second, from a methodological point of view, a more empirical classification can be based on the normative density of the evaluation clauses through the FOJ standards (FOJ, 2012). The goal is to produce a clear and operational classification of evaluation clauses that could support further researches in the field of institutionalization of the evaluation.

III. Characteristics of evaluation clauses

Independently to the legal context, necessary elements that should be clear in every evaluation demand can be identified to avoid ambiguities during evaluation implementation. These elements represent an identifiable structure in the clause phrasing that I consider here as the expected normative density\(^2\). In this respect, it proposed recently a new ‘unité de doctrine’\(^3\) which is twofold. On the one hand, the necessity for introducing such kind of obligation in a bill has to be demonstrated and that the scope of the expected evaluation has to be clarified (FOJ, 2012). On the other hand, this doctrine allows for formulating evaluation clauses according to eight dimension (six mandatory and two optional) related to their phrasing and which are consistent with the literature on managing the design of policy evaluation (Walker & Wiseman, 2006:367) and with aspects of the evaluation plan that must be tailored (Rossi et al., 2003, 33). There are six mandatory elements:

- **Target group of the evaluation results**: Which authority has to receive the evaluation results?
- **Evaluation period**: When the evaluation has to be realized?
- **Authority in charge to present the report**: Which authority has to present the evaluation results?
- **Form of the final product**: How do the evaluation results have to be presented?
- **Criteria to evaluate**: Under which criteria do the object has to be investigated?
- **Evaluation object**: Which aspect has to be examined?

Two optional elements are also proposed:

- **Authority in charge to do the evaluation**: Which authority has to implement the evaluation?
- **Evaluation goals**: What are the goals of the evaluation?

\(^2\) i.e. the degree of detail of an act (FOJ, 2007, 270) and which is operationalized here as the number of required components (according to the FOJ) contained into a clause.

\(^3\) It is recommended that the mandatory items have to be specified in every future evaluation clause and that past clauses that do not satisfy the criteria have to be modified when the related laws are debated again.
These eight elements related to the clause phrasing can be categorized as belonging either to a formal dimension of the clause (how do it have to be implemented\textsuperscript{4} i.e. the nature of the evaluator-stakeholder relationship and eventually the methods and procedures the evaluation will use to answer the question(Rossi et al., 2003, 33) or to a substantial dimension (why do it have to be implemented\textsuperscript{5}, i.e. the question the evaluation is to answer(Rossi et al., 2003, 33)).

Although the clause contexts are not taken into account in these characteristics, the classification has to be consistent with this very important aspect of the legal framework. Consequently, legal studies allow to define three contextual features of the clauses that the classification has to have and which represent a secondary validity of the typology (the primary validity being how well classification achieve the aim of providing a clear and significant distinction between ideal-types). Thus, to determine the worth of our classification, the validating criteria will be significance tests on external variables (not used to generate the cluster solution) related to the following 'Expectations'\textsuperscript{6}.

First, several Swiss authors have highlighted that, although the lower levels of government have important participatory powers in constitutional revision and law making, most of the evaluation activities have been concentrated at the federal level (Spinatsch, 2002; Bussmann, 2007). These authors suggest that the small size of cantons raises a question of a critical mass for evaluation capacities - and the evaluation clauses at the cantonal level are expected having a weaker normative density than at the federal level.

**Expectation 1:** It would be a relationship between the level of government (cantonal or federal) and the normative density of clauses.

Second, comparative studies have suggested that the emergence of evaluation (and more specifically his institutionalisation) could be historically explained by three factors: the political constellation, the fiscal situation and the constitutional features (Derlien and Rist, 2002). Since, in Switzerland, these factors are the responsibility of regional governments, one can expect differences between cantons.

**Expectation 2:** It would be a relationship between governmental units and the normative density of clauses.

Third, the evaluation culture in a policy field is not homogeneous and some of them, like Health or Education, are more familiar with evaluation and this activity is well integrated\textsuperscript{7}. Hence:

\textsuperscript{4}target group of the evaluation results, evaluation period, authority in charge to present the report, authority in charge to do the evaluation and form of the final product

\textsuperscript{5}Criteria to evaluate, evaluation object and evaluation goals

\textsuperscript{6}In the sense of propositions and concrete observations that can be expected in the empirical world and that can be derived from paradigms and theories(Blatter and Haverland, 2012)

Expectation 3: It would be relationship between the policy field and the normative density of clauses.

Fourth, the principle of hierarchy of norms requires that a certain normative content (or abstraction level) has to correspond to an appropriate normative level: constitution, law or ordinance (FOJ, 2007, 252). Consequently, it can be expected that:

Expectation 4: It would be a relationship between the type of act (constitution, law or ordinance) and the normative density of clauses.

Figure 1: Dendrogram representing the Ward’s minimum variance hierarchical classification of evaluation clauses

IV. Method

Although the technique is marginally used in political science (Figueiredo et al., 2014) and legal studies, the cluster analysis is considered as a major descriptive and exploratory statistical method (not inferential) for finding clusters of cases in a sample. The method has been ranked as a leading technique used in the past ten years to
explore and analyse large complex database (Kettenring, 2006). Basically, the method is used to sort out cases (rather than variables) into groups of similar cases, whereby the cases within each cluster are more alike than those outside the cluster (Uprichard, 2009). Clustering methods are considered to be case-based methods because they focus on understanding the relationship between the cases and not on the relationship between the variables. It is a method of partitioning a sample into homogeneous classes to produce an operational classification. The method differs from other methods that assign cases to previously defined categories (like discriminant analysis) since in unsupervised classifications, previously unknown clusters emerge out. The grouping is based on combinations of independent variables, which maximize the similarity of cases within each cluster while maximizing the dissimilarity between groups that are initially unknown. A cluster is a group of relatively homogeneous cases and it describes to which of its members it belongs.

Recently, a Swiss census of legal bases has been established in all cantons and in all policy sectors (Wirths and Horber-Papazian, 2015). Here, I consider 319 evaluation clauses identified in Switzerland, at the cantonal and the federal level. For the purpose of this paper, I collected the eight variables proposed by the FOJ and they have been recoded with binary attributes: the presence of a characteristic (code 1) or the absence (code 0). To compute the substantial and formal dimension of clauses, for each observation, I summed the attributes (code 1 or 0) related to both feature and each of the two dimensions was given the same importance since they have been weighted by

![Figure 2: Frequency of cases according to cantons and groups (only cantons representing more than 5% of observations)](image)
Since I had no prior knowledge of which evaluation clause belongs to which cluster (and since I didn’t know the number of groups that will emerge of our sample), I run a two-stage sequence of analysis combining unsupervised and supervised methods. First, I carried out a hierarchical cluster analysis using Ward’s minimum variance clustering method (Ward Jr, 1963) which is the most adapted family of method when one have no hypotheses concerning the number of clusters (Lapointe and Legendre, 1994). A hierarchical tree diagram, called a dendrogram (Fig. 1), has been produced to shows the linkage points linking the clusters at an increasing level of dissimilarity. The longer is the distance between two linkage points, the bigger the dissimilarity between the cases at the left of the dendrogram. Figure 1 shows potentially several groups in the data set. Secondly, I rerun a cluster analysis with a k-means algorithm which enables to allocate every case in the sample to a particular cluster. I computed the clustering successively with 2, 3 and 5 groups, which are the clearer sub-sets. Beyond 5 groups, since the linkage points are too closes, the dendrogram shows that clusters are too much confused.

Although the clustering has a high methodological quality, the value of the classification will also depend on how well a cluster analysis correspond to certain feature that I would like the classification to have (Romesburg, 2004). To validate a clustering solution, the best way is to perform significance tests that compare the clusters and
validating

Cluster analysis will always produce a grouping, but these may or may not prove usefulness for classifying clauses. The choice of the number of clusters depends on the explanatory power that these clusters will have. To determine this specific number of relevant groups, the cluster solution has to divide the data set into sufficiently distinct sub-groups (primary validity) to be consistent with the expectation of the concerned fields of evaluation and legal studies (secondary). Both aspects have to lead to clear and meaningful separations of the data set. Moreover, cost/benefit analysis has to be taken into account because the higher is the cluster number the higher can be the distinction, but also the complex is the typology (until to obtain as much clusters as cases). I have to find equilibrium between the typology complexity and its explanatory power. Hence, I tested the k-means partitioning with 2, 3 and 5 groups up to obtain a worth clustering (according to my validation criteria) and the best classification that emerged was three distinct clusters whom two of them are part of the same family of clause.

Regarding the primary validity, looking only at the variance of variable used to create the clusters is not enough relevant (Aldenderfer and Blashfield, 1984) since the results of this procedure are invariably highly significant (Tab. 1), excepted for the two group clustering solution. Regarding the substantial or the formal features of clauses belonging to the groups (with is a fundamental distinction related to the field of public
<table>
<thead>
<tr>
<th>Policy fields</th>
<th>Clusters 1</th>
<th>Clusters 2</th>
<th>Clusters 3</th>
<th>Total (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social policy</td>
<td>45.3%</td>
<td>34%</td>
<td>20.8%</td>
<td>100% (N=53)</td>
</tr>
<tr>
<td>Administration and Justice</td>
<td>53.5%</td>
<td>25.6%</td>
<td>20.9%</td>
<td>100% (N=43)</td>
</tr>
<tr>
<td>Education</td>
<td>59.5%</td>
<td>23.8%</td>
<td>16.7%</td>
<td>100% (N=42)</td>
</tr>
<tr>
<td>Public finances</td>
<td>36.4%</td>
<td>12.1%</td>
<td>51.5%</td>
<td>100% (N=34)</td>
</tr>
<tr>
<td>Health</td>
<td>51.5%</td>
<td>33.3%</td>
<td>15.2%</td>
<td>100% (N=33)</td>
</tr>
<tr>
<td>Economy</td>
<td>35.5%</td>
<td>25.8%</td>
<td>38.7%</td>
<td>100% (N=31)</td>
</tr>
<tr>
<td>Cantons</td>
<td></td>
<td></td>
<td></td>
<td>N=188</td>
</tr>
<tr>
<td>CH</td>
<td>45.3%</td>
<td>15.1%</td>
<td>39.6%</td>
<td>100% (N=53)</td>
</tr>
<tr>
<td>GE</td>
<td>30.4%</td>
<td>56.5%</td>
<td>13%</td>
<td>100% (N=46)</td>
</tr>
<tr>
<td>VD</td>
<td>36.7%</td>
<td>46.7%</td>
<td>16.7%</td>
<td>100% (N=30)</td>
</tr>
<tr>
<td>BE</td>
<td>79.3%</td>
<td>10.3%</td>
<td>10.3%</td>
<td>100% (N=29)</td>
</tr>
<tr>
<td>FR</td>
<td>40.9%</td>
<td>18.2%</td>
<td>40.9%</td>
<td>100% (N=22)</td>
</tr>
<tr>
<td>ZH</td>
<td>42.1%</td>
<td>15.8%</td>
<td>42.1%</td>
<td>100% (N=19)</td>
</tr>
<tr>
<td>VS</td>
<td>62.5%</td>
<td>25%</td>
<td>12.5%</td>
<td>100% (N=16)</td>
</tr>
<tr>
<td>TI</td>
<td>78.6%</td>
<td>7.1%</td>
<td>14.3%</td>
<td>100% (N=14)</td>
</tr>
<tr>
<td>NE</td>
<td>83.3%</td>
<td>8.3%</td>
<td>8.3%</td>
<td>100% (N=12)</td>
</tr>
</tbody>
</table>

Table 2: Distribution of external variables according to the clusters

Policy evaluation), clustering with 2 groups is only significant on the formal dimension and the relationship with substantial dimension is not enough strong (coef. < 0.5) this means that this is not a sufficient meaningful distinction. In contrast, clustering with 3 groups is significantly different in both dimension (respectively .519** and .656**). In this classification, the two first clusters (cluster 1 and 2) are characterized by low substantial features and quite low formal elements. The second category of clauses (cluster 3) is both highly formal and substantial. Moreover, the two first clusters have a low normative density (in average only 3/8 elements) while the third cluster has a higher one (in average 6/8). Hence, the formers can be labelled in the family of weak clauses and the latter in the strong family. This terminology will be used in the following sections of the paper.

Secondary validity is related to the legal study concerns. The typology has to take into account the context of the clauses and I expect relationships (i.e. a contingency coefficient sufficiently high and significant) with (1) the level of government (.188**), (2) the policy fields related to the concerned laws (.409**), (3) the canton where they have been adopted (.561**) and (4) the type of legal act (.281**). Consequences of these relationships (Tab. 1) are developed bellow (fig. 2 and 3).
From a such heterogeneous legal object, two families of clauses have emerged, either weak (77,1% of clauses) or strong (22,9% of clauses). Three significant templates can be identified among these families with a very high level of dissimilarity (two weak and one strong). Regarding their properties, it can be emphasized that clusters are mainly specified by three aspects.

Firstly, all the clauses related to Constitutional acts are in the weak categories. This is a normal and very important feature that has to be stressed since Constitutions are legal act with the higher level of abstraction. Otherwise, the typology would have been inconsistent. Secondly, the clauses coming from the federal level can be slightly distinguished and tend to belong to the strong category (39,9% of them, while only 19,9% of cantonal clauses are strong). Finally, regarding the characteristics with very low relationships, the Authority in charge to do the evaluation is never a relevant criteria to distinguish the groups (very rarely specified, only in 23,8% of cases) and it confirms that the FOJ was right to consider this item as optional. In contrast, the Evaluation object and the Authority in charge to present the report are the most common items, present respectively in 69,9% and 75,2% of cases.

With these introductory remarks I turn to the presentation of the following terminology (summarized Tab. 3) which captures the meaning of each clusters:

On the one hand, there are weak clauses divided into two templates. Both of them specify most of the time the two common items but they are associated with a different element in each template:

**Weak and Criteria Focused (WCF) - 157 clauses (49,2% of total cases)**: Authority in charge to present the report and Evaluation object are associated with the Criteria to evaluate - which gives his name to this group. These characteristics are respectively present in 66,2%, 54,8% and 51% of cases belonging to this group.

In contrast, the Target group of the results, the Evaluation product and the Evaluation period are rarely mentioned (absent respectively in 97,5%, 96,2% and 80,3% of cases). In other words, this archetypical clause asks the Government (in 15,9% of cases) for investigating the effectiveness (in 22% of cases) of policy measures (in 36,5% of cases).
Clauses from this group tend to come from the cantons of Bern and the federal state (fig. 2). 79,3% of the clauses coming from the former and 45,3% of clauses from the latter are in this group. This type is also ordinary in cantons of Ticino, Neuchâtel and Valais there they represent respectively 83,3%, 78,6% and 62,5% of their evaluation clauses. Regarding the policy field, majority of clauses related to Education, Administration and Justice, Health and Social policy are of this type (see fig. 3 and tab. 2). Finally, it can be emphasized that, logically, most of the Constitutions containing an evaluation clause are in this category (with a high level of abstraction) but it is also found here, most of time, ordinances instead of laws. This is quite surprising since ordinances would be expected as being more concrete and precise.

Example: Art. 170 of the Federal Constitution of the Swiss Confederation:

The Federal Assembly shall ensure that federal measures are evaluated with regard to their effectiveness.

Example: Art. 7 al. 1, of the Ordinance of the canton of Bern on the University of Applied Sciences and Arts:

The Bern University of Applied Sciences and Arts evaluates the effect of its services and its work processes.

**Weak and Time Focused (WTF) - 89 clauses (27,9% of total cases)**: The clauses of this type specify a same number of characteristics than the first group but is more focused on formal items. The same common items are presented in cases belonging to this group (respectively 73% for the Authority in charge to present the report and 94,4% for the Evaluation object) but they are rather associated with the Evaluation period (in 93,3% of cases) - instead of the Evaluation criteria, absent in 94,4% of group members. This is the reason why the group is called "time focused".

This authority can be either a dedicated instance internal to the administration (26 of cases), the front line administration (25% of cases) or the executive (22% of cases). Like in the first group, the Evaluation product and the Target group of the results are also rarely mentioned (absent respectively in 97,6% and 95,2% of cases). Basically, these clauses say that the Government (31,5% of cases) has to evaluate "Periodically" (39,3% of cases) either the entire law (39,3% of cases) or just policy measures (37,1% of cases).

This group is fairly characterized by clauses formulated in cantonal laws and coming from the cantons of Geneva and Vaud (fig. 2). In each of them, this type of clauses represents respectively 56,5% and 46,7% of the repertoried cases. The main concerned policy fields are the same than in the first group: Social policy, Health, Administration and Justice and Education (fig. 3).

Example: Art. 18, al. 3 of the Swiss Federal Act on the Elimination of Discrimination against people with disabilities (LHand):

---

8 Official translation
9 Own translation
10 Official translation
The Confederation evaluates periodically the impact of the measures taken for the integration of disabled people. She can also evaluate the impact of measures taken in this field by other authorities or by individuals.

Example: Art. 4, al. 2 of the Act of the canton of Geneva on Health policy\textsuperscript{11}.

Whether a legislative project is likely to have negative health consequences, the State Council may decide to go with an evaluation of its potential impact on health.

On the other hand, there is the template for strong clauses:

**Strong - 73 clauses (22.9\% of total cases)**: The clauses belonging to this cluster have a very high normative density because they define a lot of elements (on average 6). Every formal and substantial elements are specified including the *Evaluation goal* which is considered as optional by the FOJ, excepted the *Authority in charge to do the evaluation* (only presents in 39.7\% of cases). The most common items are the *Authority in charge to present the evaluation results* (97.3\%), the *Evaluation period* (84.9\%) and the *Target group of the evaluation results* (83.6\%).

Hence, the archetypical clause demands reports (74\%) of the effectiveness (24.7\%) of policy measures (47.9\% of cases), written "periodically" (42.4\% of cases) to the Legislative (39.7\% of cases) by the Government (35.6\%) and if possible accompanied by recommendations (34.2\%).

51.5\% of clauses related to Public finance and 38.7\% of clauses related to Economy are gathered here. Clauses are most of time formulated into federal laws (as emphasized above) but 42.1\% of clauses coming from Zürich and 40.9\% of clauses coming from Fribourg are in this group.

Example: Art. 19 of the Swiss Federal Act on Freedom of Information in the Administration\textsuperscript{12}

1. The Commissioner shall review the execution and effectiveness of this Act and, in particular, the costs incurred in its implementation, and shall report on a regular basis to the Federal Council.
2. The Commissioner shall submit the first report on the implementation costs of this Act to the Federal Council within three years of its entry into force.
3. The reports of the Commissioner shall be published.

Example: Art. 31 of the Act of the canton of Zürich on fiscal equalization\textsuperscript{13}.

1. The State Council shall submit to the Parliament at least every four years, a report on the implementation and effectiveness of this Act.
2. The report provides information on

\textsuperscript{11}Own translation
\textsuperscript{12}Official translation
\textsuperscript{13}Own translation
(a) The achievement of the objectives of financial compensation in the past period,
(b) Changes in the distribution of public tasks between Canton and communities, and the resulting impact on the flexibility and the finances of municipalities,
(c) The development of the resources of the communities and their exposure to the completion of necessary tasks.

3. Whether the resource differences or the agglomerations have changed significantly, the State Council requested the Parliament a revision of the law, which takes into account the new conditions

VI. Discussion

Even if the accuracy and the significance of the typology has been demonstrated, now the most important questions future works will have to address are: (1) Research into the causes of types’ existence and (2) how they are implemented (and with which consequences). First, to our knowledge, no studies have been devoted to the reasons why public authorities on the federal and cantonal level have specifically introduced evaluation requirements through evaluation clauses - especially of a specific type. Does the normative density (strong/weak or formal/substantial clauses) is depending on the (un)competency of the lawmaker, or does it reflect a political willingness? Can an empirical link be established? The Federal Audit Office assumed that the imprecise phrasing of the evaluation clauses (too low normative density) tends to legitimise the authority activities and not to critically investigate them. However, the empirical link between an incomplete phrasing and a strategic goal has to be demonstrated in further studies and could open a fruitful research track on the instrumental or symbolic use of knowledge (Boswell, 2009) through the institutionalisation of the evaluation.

This research agenda could be twofold. Firstly, how fields and governmental contexts impact evaluation demands? On the one hand, I have emphasized that Education and Health are characterised by weaker evaluation clauses. It raises the question of the extent to which these fields would have intrinsic characteristics which could lead to request for evaluations without information on their implementation. In contrast, the policy fields of Economy and Public Finances have stronger and more substantial clauses, especially regarding the fiscal equalization. It can now be investigated whether it means that the more the topic is related to money, the more policymakers have precise expectations regarding how and why the evaluation has to be implemented. Moreover, fiscal equalization is mainly related to intercantonal contracts (not present in our dataset) and it would be interesting to verify if the high density of the clauses is linked to a contractual evaluation clause. On the other hand, the cantons of Fribourg, Zürich and Geneva have the strongest clauses. Which particular institutions or political culture related to these places can explain that? Following my second expectation, it can be argued that fiscal pressure (the need for curbing ineffective programs in order to cut back the federal/cantonal budget), the political constellation (politicians and government parties) and the constitutional feature (how the power is divided among instances) could have had an impact on how evaluation has emerged.
and has been institutionalised (Derlien and Rist, 2002).

Secondly, does the higher normative density provide a better implementation of the evaluation? Can be established a link between non-utilization or the imperfection of evaluation results and the weaker formulation of the clauses? Which type of clauses is more appropriated to reach the evaluation goals? Once again, no empirical studies have been able to show such kind of evaluation designs and the concrete evaluations. However, fields like Education and Health, characterized by weak clauses, are also fields with a very high evaluation activity\textsuperscript{14}. Finally, Christina Boswell (Boswell, 2008, 2009) highlighted that most of the evaluation results are never used and it can be asked if the stronger clauses increase the chance of evaluation utilization.

VII. Conclusion

This paper has constructed a typology of evaluation clauses by means of a cluster analysis. Two families of clauses have emerged with three distinct templates related to their normative density. First, regarding the practical significance of this paper for politicians and practitioners, it opens a research track for further researches which will consider evaluation clauses as institutions and that could now distinguish different kinds of rules, more or less strong, thanks to the new typology exposed above. As a research perspective, neoinstitutionalism remains a powerful approach which, combined with this typology, allows the researcher to investigate both the reason why a specific type of institution has been chosen, and the consequences of these institution types on the policymaking process and the administration.

Secondly, the theoretical significance of this paper can be stressed since the attributes proposed by the Federal office for justice are not politically or culturally linked to Switzerland. These items can be easily used outside the Swiss context. Changing only the first expectation (related to the Swiss cantonal government) to adapt it to other national context, the methodology exposed here could provide an innovative input to comparative studies related to the institutionalisation of evaluation (Furubo et al., 2002; Jacob, 2005). Until now, the comparative works only focused on organisational feature of the institutionalisation, and scholars admitted that it was impossible to take into account the procedural (legal) dimension of this topic.

Finally, from an epistemological perspective, this paper has applied to the field of law an atypical tool to construct a useful conceptual scheme for classifying entities. Even though the classification was here a goal in and of itself, the proposed typology represents a first multidisciplinary step within a broader research agenda related to the study of institutionalisation of evaluation and made possible by means of to the cross-over between political science (especially evaluation of public policy), law and numerical taxonomy. In this respect, I am convinced that the cluster analysis would have fruitful practical application for the legal studies through the conversion of heterogeneous normative acts into complex structured institutions.

\textsuperscript{14}SynEval Collaboration, op. cit.
References


FOJ (2012). Recommandations pour la formulation des clauses d’évaluation.


