

Indexing watchdog accountability powers a framework for assessing the accountability capacity of independent oversight institutions

Mark Bovens

School of Governance, Faculty Law, Economics, Governance, Utrecht University, Utrecht, the Netherlands

Anchrit Wille

Institute of Public Administration, Faculty of Governance and Global Affairs, Leiden University, Leiden, the Netherlands

Abstract

Independent oversight institutions are critical components of the accountability landscape in modern democracies. This paper presents a framework for assessing the accountability powers of these watchdogs. This watchdog accountability index is an empirical tool to assess the key accountability powers of accountability forums that operate in a democratic constitutional context. The aim is to provide a richer evidence base to assess evolving external accountability arrangements and their effectiveness. Our approach breaks down the concept of watchdog accountability power into three distinct, conceptually coherent dimensions. We apply the accountability index to assess the strength of one of the main watchdog institutions in the EU, the European Court of Auditors in 2017. Data were collected by means of a study of secondary sources and by an expert survey.

Keywords: accountability, European Court of Auditors, index, oversight institutions, watchdog.

1. Toward measuring the accountability powers of watchdogs

Independent oversight institutions have become critical components of the accountability landscape in modern democracies. These “watchdogs” play a key role in promoting the transparency, accountability, and effectiveness of the executive. They investigate performance problems, uncover cases of corruption, and report those findings to Parliaments and the general public. With the growth of executive spheres at all levels of the government, the need for oversight and accountability has become more and more obvious. Terms, such as “the Audit Society” (Power 1997), “counter-democracy” (Rosanvallon 2008), and “monitory democracy” (Keane 2009) attest to the rapid growth of such extra-parliamentary, power-scrutinizing mechanisms as ombuds-institutions, auditors, and integrity offices. As Keane (2011, p. 213) indicates: “Within and outside states, independent monitors of power begin to have tangible effects. By putting politicians, parties, and elected governments permanently on their toes, these monitors complicate their lives, question their authority, and force them to change their agendas and sometimes smother them in disgrace.” Likewise, Rosanvallon (2008, p. 274) argues that “the inability of electoral/representative politics to keep its promises” (has) led to the development of the threesome oversight, prevention, and judgment as core mechanisms whereby citizens can hold their rulers accountable between elections with the support of watchdogs, such as ombudsmen, audit institutions, and integrity offices.

As watchdogs have gradually become more salient in modern governance, it has become increasingly urgent to better understand the institutional and operational elements that affect their effectiveness as accountability institutions. The research on the accountability effectiveness of these watchdogs is still rather limited (Van Acker & Bouckaert 2019, p. 55). In the words of Wilkins (2019, p. 87): “It is evident that the issues of wider impacts and improving accountability would benefit from more detailed consideration.” This paper, therefore, presents an outline for an index to assess the accountability powers of watchdogs. A variety of indices are

Correspondence: Mark Bovens, School of Governance, Faculty Law, Economics, Governance, Utrecht University, Bijlhouwerstraat 6, 3511ZC Utrecht, the Netherlands. Email: m.bovens@uu.nl

Accepted for publication 19 April 2020.

available that seek to assess the quality of government across the globe.¹ As yet, no widely accepted accountability index with a broad coverage has been constructed. Conceptual complexity and the lack of quantifiable measures have hampered our empirical understanding of the development of the accountability landscape. By constructing a watchdog accountability index, we aim to provide more insight into the forum part of the accountability landscape. And by constructing an accountability index, we aim to make easy comparisons possible in space and time, creating a richer evidence base to assess these evolving external accountability arrangements.

The first part of the paper provides a theoretical framework in which we distinguish three dimensions of watchdog power. This is essential as we currently have only a vague understanding of how watchdogs contribute to executive accountability. This is followed by a detailed operationalization, classifying the relevant qualities and activities needed to measure the accountability powers of watchdogs along these three dimensions. We use a mixed deductive-inductive approach. After our initial theoretical conceptualization, we use preliminary observations of the accountability powers of one, the well-established watchdog, in particular, the European Court of Auditors (ECA), to illustrate the operation of our watchdog accountability index. With this in-depth theoretical and empirical approach, we strive to explicitly address the problems of measurement. We explain how data were generated by means of a study of annual reports and secondary sources and a survey among experts. In the final section, we discuss how the evidence from the ECA can be used to develop a robust index that can be applied to other watchdog institutions as well.

2. Watchdogs as core accountability forums

2.1. Actor accountability and forum accountability

We define accountability as “a relationship between an actor and a forum in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgment, and the actor may face consequences” (Bovens 2007). Starting from this definition, we can distinguish at least two types of institutions that can be assessed from an accountability perspective: the various executive actors that are the account-givers and the forums that can hold them to account.

One place to start could be the assessment and indexing of the accountability behavior of the executive bodies. This is what we call actor accountability. The executive actor itself can contribute to accountability, for example, by acting in a transparent way, and by proactively providing information and justifications. In the literature, a number of frameworks and measurements have been developed to assess the accountability of a wide variety of executive actors. A successful attempt to operationalize actor accountability is the Global Accountability Framework (Blagescu et al. 2005), which was developed by the charity One World Trust. It identifies four core dimensions that make an organization more accountable: transparency, participation, evaluation, and complaint and response mechanisms. Each of these four dimensions is formulated as a standard for accountable behavior (Lloyd et al. 2007, p. 11) and measured on the basis of a series of indicators. Koop (2011) offers another example of how actor accountability can be measured, namely through the operationalization of the *de jure* accountability obligations of independent agencies. Brandsma (2014, pp. 147–151) provides an overview of a range of indicators of *de facto* actor accountability. More specifically, Welch and Wong (2001) scored government websites on a broad range of indicators, including contact details, organization charts, and the availability of policy documents. Coy and Dixon (2004) constructed a public accountability index in order to assess the quality of annual reports of universities in New Zealand. Amtenbrink (2012) explored the possibilities of an accountability index for informal international lawmakers.

Executive accountability cannot flourish if the forums, the accountholders, are weak. In this paper, we will confine ourselves to the latter part of the relation and concentrate on the accountability powers of accountability forums. This we call forum accountability. The opportunity for “holding executive power to account” is essential for democratic systems and can take different forms, such as political, administrative, legal, financial, or integrity accountability, depending on the nature of the forum.

2.2. Watchdogs as core accountability forums

Watchdogs are important accountability forums. However, not all accountability forums are watchdogs. Courts and parliaments, for example, are very important for legal and political accountability, yet they perform other

equally important roles as well, such as providing adjudication or drafting legislation. In this paper, we limit the term watchdog to accountability forums that are purely engaged in second-order governance tasks, in delivering accountability by monitoring executive actors in the implementation of their first order governance tasks (Tömmel 2016). We follow the definition of Wilkins (2015, p.19), who states that watchdogs are “statutory agencies that have a primary function of scrutinizing the actions of the public sector and providing reports independent of those scrutinized.” They can do so by providing recommendations, solutions, or by administering sanctions. Ombudsmen, auditors, and anti-corruption agencies are primarily watchdog institutions; courts and parliaments are not. None-state actors, such as media or NGO’s, often call themselves watchdogs too. In developing our index, we confine ourselves to statutory agencies and public institutions, in line with the definition of Wilkins. However, in the future, the index could be adapted to assess the powers of private watchdogs or other regulatory institutions.

Accountability relations may serve numerous purposes of which democratic control, curtailing executive abuse, and learning are the most important (Bovens 2007; Bovens et al. 2008). Watchdog institutions can help to achieve each of these goals. Their detailed reports can contain vital information for Parliamentary committees that monitor and control the performance of the executive. McCubbins and Schwartz (1984) indicated, for instance, that legislators generally prefer fire alarms oversight to police patrol oversight: a highly decentralized system of legislative oversight, which relies on outside actors to “sound an alarm.” The work of auditors, ombudsmen, and integrity offices increasingly complements the oversight functions of the established branches of government. Some qualify watchdogs as satellites of the parliaments (Wilkins 2015); others even see them as an indication of the rise of a “fourth branch” of government alongside the legislature, executive, and judiciary (Head 2012, p. 7). Watchdog institutions also have an independent role in curtailing executive abuse. Their investigations can bring to light executive deviance, corruption, and mismanagement (Gustavson & Sundström 2018). They are a form of intra-governmental accountability—a case of public institutions that hold other public institutions accountable.² But watchdogs do not just expose abuses and faults in the public sector they also seek to remedy them (Aucoin & Heintzman 2000; Van Looke & Put 2011; Feldman & Eichenenthal 2014; Reichborn-Kjennerud & Vabo 2017).

2.3. Watchdog accountability powers: Three dimensions

When can these watchdog institutions be said to be able to perform as they should? How can we establish whether they have the ability to hold executive actors to account? This depends on the accountability powers that they possess. Accountability power is defined here as the ability of a forum to hold the executive actors to account. In other words, what needs to be assessed is the strength of the forum in its relation to the actor. This is not the same as assessing whether there is an accountability relationship at all between a specific actor and a specific forum that would require measuring the various key elements of accountability: does the actor feel obliged to provide information, can the forum ask questions, pass judgment, and can that have consequences? Nor is it the same as measuring whether account holding actually occurs in concrete cases. That would require measuring the outcome of the actor’s account giving behavior and the outcome of the actor-forum interactions in a series of cases.

What we are rather looking for is the general accountability strength of these core watchdog institutions. Like other governance phenomena, accountability power is difficult to measure and evaluate since it is characterized by a multiplicity of aspects or dimensions. In our conceptualization, it consists of three dimensions. A watchdog has strength if it has the authority to contribute to account holding, if it has resources to hold the executive to account, and if it employs account holding activities. Strength depends, thus, on may-do (and must-do), on can do, and on real-do components.

The first dimension we distinguish is the formal powers of the watchdog to hold actors to account. In terms of the watchdog metaphor: an analysis of its formal powers is relevant to establish whether it has the constitution of a watchdog at all. These powers are likely to be codified in laws, treaties, regulations, decrees, or statutes. The remit of the watchdog determines its ability to set its own agenda and to act accordingly. Is the watchdog autonomous in governing itself, and can it decide what cases and topics to pursue? Is there no interference of the executive with the appointment and reappointment of its leadership? Next, it is important to establish whether the

watchdog is able to force the actor to provide all information it deems necessary for a full examination of executive behavior. Can it demand information, can it summon witnesses without their consent, and can it hear witnesses under oath? Finally, it matters whether the watchdog has formal competences to sanction executive bodies or their representatives, in case it finds they have demonstrated inappropriate behavior. Can the watchdog not only bark but also “bite”?

Second, it is a matter of organizational strength. This dimension indicates how powerful the watchdog actually is: does it have good ears and a strong bark? A legal framework is not a sufficient condition for accountability power (Torres et al. 2019). A watchdog can have a large array of formal powers and yet may have no resources to exercise them. Examples of this can be found in illiberal democracies, where watchdogs may have strong accountability powers on paper, but are very weak in practice, due to limited budgets, insufficient or incompetent staff, or partisan and corrupt leadership. To assess organizational power, the following elements are relevant: do these forums have the necessary resources to implement their formal powers? Do they have sufficient expertise, have they developed professional standards, do they have sufficient budgets and qualified personnel, and can they rely on professional networks? An additional factor that determines a watchdog’s organizational capacity is the quality of its leadership.

The third component depicts to what extent accountability holding is realized in practice by the forum. For that, we look at accountability in action—does the watchdog actually engage in account holding (Van Acker & Bouckaert 2019, p. 67)? In terms of the watchdog metaphor: does it actually bark? What matters most of all, is the actual exercise of these powers. It is not enough to measure formal remit, and organizational capacity; these should be complemented by outcome-based indicators (Kaufmann & Kraay 2008, p. 9). How often do the watchdogs exercise their formal powers, how many accountability activities do they engage in each year? To a large extent, their accountability powers depend on how they interpret and exercise their mandate, on the initiatives they undertake, and their effectiveness in carrying out their powers. A further aspect is whether they go beyond their mandate and find new and innovative ways to hold the executive to account. The effectiveness of a watchdog depends, partly, on whether it can go beyond its traditional formal oversight role and link its work into the executive’s policymaking and policy decisions. By finding innovative ways to achieve their goals, watchdogs can improve their salience and credibility, and increase their impact on executive policies and programs. A case in point is ombudsman institutions, which often have very limited formal powers, and yet in practice do have substantial accountability power vis-à-vis executive actors because they operate in creative ways to enhance their leverage (Hertogh & Kirkham 2018).

These three dimensions constitute a forum’s accountability powers (Table 1). A comprehensive watchdog accountability index should measure each of these three dimensions.

These three dimensions are broader than the oft-used distinction between the *de jure* and *de facto* components of accountability (Mechkova et al. 2019). This is a useful distinction to explore the difference between the *de jure* institutions of accountability and their *de facto* implementation. Our aim is not to focus on the *de jure-de facto* gap, but to appraise the strength of watchdogs, and for this, we not only need the formal and the factual parts of accountability holding, but also the administrative, logistic, and organizational components.

In addition, the relationship among the separate dimensions is not necessarily a causal relationship. Giving more formal powers or resources as such does not automatically imply more exercise of power by the watchdog. A watchdog can have a large array of formal powers, and yet may have no resources to actually exercise them.

Table 1 Three dimensions of accountability powers

Formal powers	Organizational powers	Exercise of powers
Autonomy	Strategy	Productivity
Independence	Staff size	Effectiveness
Agenda setting	Staff quality	Salience
Information powers	Budget	Credibility
Questioning powers	Network	Creativity
Sanctioning powers	Board	Impact
	Leadership quality	

And even if it does have sufficient resources, it still may not exercise them and thus remain a paper tiger. And on the other hand, watchdogs with limited formal powers, may in practice have substantial accountability power vis-à-vis executive actors when they operate in creative ways to enhance their leverage.

3. Indicators of watchdog accountability powers

3.1. Measuring and indexing accountability power

The next step is to move from conceptualizing toward measuring and indexing accountability power. Indices, or composite indicators (CI), are increasingly used as tools in policy analysis (Cooley & Snyder 2015). According to Merry (2011, p. 83), their use as tools for assessing and promoting a variety of social justice and reform strategies around the world is rising fast.³ In the case of watchdog institutions, an index can be helpful to assess how the accountability powers of a specific institution develop over time, or to compare the performance of various similar organizations across time. Indices can also be helpful in identifying trends and drawing attention to particular issues, and in setting policy priorities (Alston 2000, p. 250).

Measurement implies that one “attaches conceptual relevance to some phenomena and ignores others” (Hooghe et al. 2016, p. 3). In the past years, various measurements of forum accountability have been developed. Brandsma (2013) has looked at the frequency and the intensity of the discussions between subordinates and superiors and whether superiors applied formal or informal disciplinary sanctions to subordinates. Van de Steeg (2009) coded whether the European Parliament adopted resolutions after the debates with presidents of the European Council. Brandsma and Schillemans (2013) devised an accountability cube to measure the levels of de facto accountability in a variety of principal-agent relations that aggregates all three phases. Using this, for each principal-agent relationship they measure whether the amount of information provided is high or low, whether the discussions are intensive or not, and whether many or few sanctions are applied. These measures are an effort to assess a specific dimension of accountability, but they are not specific enough for assessing watchdog powers, and they focus only on the exercise of these powers. By contrast, we have developed a specific watchdog accountability power index, on the basis of the three dimensions we identified. The index not only gauges the formal powers accountability in the books—but also looks at accountability in action, by focusing on the exercise of these powers. We will discuss how we operationalized each of the three dimensions.

We began by developing indicators for the three dimensions of accountability power, based on the literature and a series of exploratory interviews with practitioners. In general terms, an indicator “is a quantitative or a qualitative measure derived from a series of observed facts that can reveal relative positions (e.g. of a country) in a given area. When evaluated at regular intervals, an indicator can point out the direction of change across different units and through time” (Organisation for Economic Co-operation and Development 2008, p. 13). Indicators need to be able to distinguish the magnitude and direction of the relevant differences in accountability powers of the watchdogs so that the quality and quantity of accountability power can be assessed and compared. The resulting assessments should lead to a transparent and comprehensible procedure for assigning scores to the individual items.

3.2. Indicators of formal accountability powers

With regard to formal powers, we first look at the watchdog’s formal autonomy and independence—is it permitted to govern itself without the interference of the executive?⁴ There is consensus in the literature that independence is an essential attribute of watchdog institutions.⁵ To assess this, we look at the following indicators:

- **Formal autonomy:** The extent to which the institution is legally recognized. Here we look at how the mandate is established. What is the legal basis of the watchdog? Is the mandate guaranteed constitutionally, by treaty, by formal law, or only by an administrative act (Blume & Voigt 2011, p. 219)? The constitution, or in the case of the EU a treaty, is usually more difficult to change than ordinary law, let alone an administrative act. Hence, constitutional guarantees make the autonomy of watchdogs more credible.
- **Formal independence:** The extent to which the board is free from interference by the executive. Independence is an essential feature of the role of watchdogs (Wilkins 2019, p. 84, 87). Here we examine whether there is any involvement of the executive in the selection, appointment, and removal of the leadership of

the watchdog (Wonka & Rittberger 2010, p. 738–739; Koop & Hanretty 2018). If the government can appoint auditors or ombudsmen, they may be tempted to choose “old buddies” who will be less critical. The same goes for their reappointment or removal: “if government members who dislike the findings of the supreme audit institutions (SAIs) can easily get rid of them, this would be inimical to their independence and the effectiveness of their reports” (Blume & Voigt 2011, p. 219).

- **Agenda setting autonomy:** The extent to which the watchdog has the discretion to investigate anything, which falls within its mandate. Here we look at whether the watchdog is autonomous in setting its own agenda (Wonka & Rittberger 2010, p. 739). Can it decide itself what topics to investigate, or solely act in response to an external request?

Second, we examine whether the watchdog has been provided with the various statutory powers that are essential to operating as an accountability forum. Does it have the formal powers to obtain information, to hear witnesses, and to sanction officials or actors in the case of irregularities? To assess this, we look at the following indicators:

- **Information power:** The extent to which the institution has the statutory power to obtain the information necessary for executing its mandate. Informing is an essential part of the accountability process. Therefore, it is crucial that the watchdog has the right to obtain any information it deems relevant, even against the will of the executive agency. Here we determine whether the watchdog may inspect or obtain documents, files, and minutes, without permission of the executive, or whether it even has the authority to enter and search offices.
- **Questioning power:** The extent to which the institution has statutory powers to question officials and witnesses. Another important element of the accountability process is the forum’s ability to question the actor, and to debate its motives and justifications. Hence, we also ascertain whether the watchdog has the right to subpoena any witnesses it wants to question. The quality of the information will be enhanced if it also has the power to question them under oath.
- **Sanctioning power:** The extent to which the institution has formal powers to sanction actors when it finds irregularities. The ability to impose sanctions—or hand out rewards—will render an extra “bite” to the judgment of the watchdog institutions and may enhance the chances that their findings and recommendations will lead to improvement of the executive performance. Ideally, watchdogs themselves have the power to impose sanctions on executive actors, but they may also simply act as informants to external principals of executive bodies, such as parliamentary commissions or ministers. A more informal, but in some cases effective way of sanctioning is the use of naming and shaming (Kostadinova 2015, p. 11). This requires that its reports can be made public.

3.3. Indicators of organizational powers

With organizational powers, we look at the administrative capacity of the watchdog to exercise its formal and informal accountability powers. An examination of the human, financial, and infrastructure resources helps to provide insight into its ability to carry out its institutional role. We distinguish between administrative powers and the strength of leadership and use a series of indicators to measure these two elements. Indicators of organizational power are:

- **Strategy & mission:** The extent to which the institution has outlined its objectives through a clear strategy and mission. To assess its organizational power, we first check whether the organization has a sense of purpose for which a clear mission and an explicit strategy serve as a proxy. Accountability forums with a clear mission and a set of clearly defined roles and aims for their staff members are assumed to be more capable of operating as an institution. Here we look at whether the watchdog has produced documents that attest to a clear and specific strategy, which is renewed or adjusted from time to time.⁶
- **Staff numbers:** The number of staff the institution has at its disposition to execute its mandate. Organizational power will depend upon the number of staff that is available for investigations. It is assumed that large watchdogs will have more accountability power than small ones. Obviously, this is relative to the size

of the executive it has to control. At the national level, a proxy for this could be the number of staff per capita. This proxy is more difficult to use at the EU level, given its multi-level structure.

- **Quality of staff:** The extent to which the institution hires staff with the appropriate professional qualifications. Research by Gustavson and Sundström (2018, p. 8–9), for example, shows that audit professionalism has a very large effect on reducing corruption levels in the executive. In order to assess the quality of the staff, we look at its level of professional expertise (Van Acker & Bouckaert 2019, p. 59). Indicators are educational attainment of staff, experience levels of staff, and staff professional specialization. This requires a qualitative estimate by professional peers of the watchdog.
- **Size of budget:** The extent to which the budget of the institution matches the task at hand. Accountability forums require funds to set and implement their activities. How large is the operational budget that is available for accountability activities, given the size of the actor it has to monitor? At the national level, a proxy for this could be the ratio of the watchdog's budget, and the budget of the executive it is tasked with overseeing. Again, at the EU level, this is more difficult, given its multi-level executive structure.
- **Network Integration:** The extent to which the institution participates in established networks that enhance its capacities (O'Donnell 1998; Santiso 2006; Wilkins et al. 2017). Institutional and functional linkages between watchdogs and other forums in the broader system of checks and balances can improve accountability. For many types of watchdogs, close relations with parliamentary committees can provide their reports with more "bite." In this way, they can operate "in the shadow of hierarchy" (Scharpf 1997, p. 202; Schillemans 2008). And for EU watchdogs, close cooperation with counterparts in the member states is also very relevant for their effectiveness. To assess this, we look at the availability of accountability networks (Harlow & Rawlings 2007). Network integration describes the ability of the institution to organize its external relationships as "inter-institutional linkages." This may be an indication of its impact in the wider institutional structure and on executive activity.⁷

A separate dimension of organizational powers is leadership. Good leadership is essential for watchdog institutions to operate effectively. Here we distinguish two different indicators:

- **Board composition:** The extent to which the leadership structure enables effective leadership (Stephenson 2015). In the literature (Elsayed 2011, p. 418), large board size is argued to enhance performance if an organization has a greater need for external linkages. This is not the prime concern for watchdogs, which have to maintain a fair degree of autonomy and independence from external stakeholders. On the other hand, large boards tend to lead to more communication and coordination difficulties, less cohesiveness, and less openness for new ideas. Single-headed leadership allows for swift and proactive decision-making, but may also lead to tunnel vision and discontinuities. On the basis of the literature and interviews with experts, we assume that watchdogs operate more effectively with a small board, or college, that provides both effective leadership and some organizational checks and balances. Therefore, single-headed leadership is preferred over large boards, but concise collegiate bodies are preferred to monocratic constellations.
- **Background of leadership:** The extent to which the board members have the appropriate professional qualifications. As with staff, it is assumed that a high level of professionalism of the core executive of the institution is to be preferred. Relevant degrees and previous work experience in similar positions will enhance the quality and effectiveness of leadership. Political expertise may be relevant to some extent at the helm of a watchdog. But with political appointments, there is the risk that political loyalties may trump professional judgment.

3.4. Indicators of the exercise of accountability powers

Formal powers and qualified personnel are not enough. The watchdog also needs to actually exercise these accountability powers. Otherwise it will remain a paper tiger—or dog, that is. What matters most is accountability in action. This is measured on the basis of the following indicators:

- **Productivity:** The extent to which the watchdog makes use of its *de jure* powers at all. Does it actively hold the executive to account, or is it a passive, "sleeping" dog? The most extensive way to score this would be to document for a forum the actual exercise of each of the three sets of formal powers: information

requests, discussion, and sanctions (compare Brandsma 2014, p. 148–149). For pragmatic reasons, we have opted to use a proxy and to concentrate on the main accountability activity of the forum. This should be assessed by comparing it to the resources of the watchdog. Small watchdogs, with limited staff and budgets, cannot undertake large-scale action.

- **Effectiveness:** The extent to which its recommendations and reports lead to a follow-up by the executive actor. Ultimately, what matters are not outputs, but outcomes. Is the watchdog able to lead the audited actor to correct deficiencies (Mizrahi & Ness-Weisman 2007, p. 191)? Establishing effectiveness is fraught with difficulties and a topic in its own right. In the literature, various highly sophisticated methods for operationalizing and measuring auditing effectiveness are available (Mizrahi & Ness-Weisman 2007; Pelizzo 2011). Many authors focus on a relatively simple proxy: the compliance rates, defined as the proportion of recommendations that are implemented (Pollitt et al. 1999; Kostadinova 2015; Van Acker & Bouckaert 2019, p. 61; Torres et al. 2019). However, an actor's favorable response to the watchdog's recommendations is not the same as the actual implementation of these recommendations, nor is such implementation a guarantee that the deficiencies in question have been resolved. Moreover, the implementation of recommendations or the absence thereof, is a limited indicator of impact (Desmedt et al. 2017). Effectiveness might also be gauged by examining whether the watchdog's reports have influenced the political debate, which is usually measured by the amount of attention gained in Parliament (Torres et al. 2019, p. 433). Others measure the impact of performance audits on the basis of the opinions of the auditees and their contribution to the public debate (Morin 2001; Reichborn-Kjennerud 2014; Torres et al. 2019, p. 433). Here we will mostly rely on a soft indicator—the opinions of a range of stakeholders on whether “there is a clear follow up to the reports and recommendations of the watchdog.”
- **Salience:** The extent to which the watchdog plays a vital role in keeping executive agencies accountable. Is it a marginal or a central player in the accountability landscape? This rather soft indicator will be measured on the basis of expert opinion.
- **Credibility:** The extent to which the forum is recognized as credible. This looks at how external stakeholders, such as the general public, the media, or civil society watchdogs rate the performance of the institution. This is a rather “soft” indicator, which can be measured by expert opinion, or on the basis of survey data that measure reputation (Carpenter & Krause 2012; Busuioc & Lodge 2017).
- **Creativity:** The extent to which the institution makes creative use of informal powers. Watchdogs typically exercise a variety of informal mandates. Particularly in international politics, informal governance is the norm, not the exception (Moravcsik 2010). This may be embodied in the standards they set for accountability, and the initiatives these institutions develop to show that their operations do indeed create accountability. To assess these informal powers, we look at whether the watchdog is creative in devising other ways to enhance its accountability powers, for example through the implementation of an active communication policy, publication and dissemination of reports, user-friendly result summaries, etc.
- **Impact:** The extent to which these formal and informal activities lead to the improvement of accountability practices (Van Acker and Bouckaert 2019). Accountability is not only about control, but also about learning and improvement. In the past decades, the emphasis of watchdogs has moved more in the direction of performance management. Ombudsmen, integrity offices, and SAIs all “contribute to public management by advocating learning, best practices, accountability, and innovation” (Pierre & de Fine Licht 2019, p. 226). We have, therefore, looked at the impact and effectiveness, and whether these activities help in setting new, and higher standards for transparency, proactive informing etcetera. In the expert survey, we have asked whether the accountability activities lead to improvement of accountability practices and whether there is a clear follow up to the reports and recommendations of the watchdog.

4. From indicators to observations

4.1. Application

We selected the ECA as an exemplary case from the broader population of watchdogs to develop the indicators. While audit institutions have been part of the state's institutional setup for a very long time—in many member states for centuries (Soll 2014)—“they recently left the periphery of the administrative landscape and moved

closer to center-stage” (Pierre and De Fine Licht 2019, p. 1). Several developments have contributed to this increased role and salience of audit institutions. First, the complexity of governance. Audit institutions “have become more important as the programs and commitments of government have become more complex and specialized.” Scrutiny of government finances requires “the input of sophisticated audit and other accountability professionals” (Posner & Shahan 2014, p. 488). Second, the growing importance of “value for money” movement (Posner & Shahan 2014; Pierre & De Fine Licht 2019). The increasing focus on efficiency, outputs, and outcomes heightened the interest in verification, control, and performance measurement. Finally, in Europe, the enhanced role of SAIs is linked to the emergence of an ever-denser institutional accountability landscape. The EU’s complex multi-level network governance structures, cutting across decisional levels, have contributed to a *mélange* of horizontal and vertical accountability regimes and practices at distinct levels. Most of the EU budget is spent under what is known as “shared management,” with individual EU member states actually distributing the funds and SAIs controlling the expenditure (Cipriani 2010; Aden 2015).

The ECA is a well established, major watchdog institution in this EU accountability landscape (Stephenson 2015, 2016; Wille 2015). It operates in the field of financial accountability, which is a well-developed professional field with a long history. The ECA is well known and part of an intergovernmental network of auditors, which made it easier to find international experts for our survey. We use the ECA as an example to illustrate the index and to explore the relevance and fit of the indicators, the methods for data collection, and the presentation of the data.

4.2. Methods of observation

We aimed to assess different dimensions of accountability power. Formal power depends on formal rules, which are laid down in laws, constitutions, protocols, and rules of procedure. These can be measured by using unobtrusive information sources, such as, in the case of the ECA, treaties, and protocols. To assess the exercise of the power dimension of the index, on the other hand, we mainly depended on the informed perceptions and observations of those familiar with and knowledgeable of the actual operation of these watchdogs. The exercise of accountability dimension was measured with the help of questionnaires and interviews with various stakeholders and experts. The organizational dimension was determined both on the basis of documents, such as annual reports and internal protocols and on the basis of an expert survey.

4.3. Unobtrusive measures

A plausible measurement is transparent, specific, and should tap the core meaning of a concept (Hooghe et al. 2017, p. 4). In order to measure the formal and organizational dimensions, we developed a coding sheet with instructions for completing it. We used a categorical approach, with clear labels for each of the categories to reduce intercoder disagreement. Coding was based on official webpages, fact sheets, rules of procedure, and annual reports alongside secondary sources to gain information on the formal and organizational features of the ECA.

4.4. Expert surveys

We then proceeded with a small-scale, qualitative preliminary pilot study, which aimed to investigate the feasibility of crucial components of the accountability index. To that end, we applied the various aspects to the ECA. We developed an online questionnaire based on the Index’s conceptual framework. For all dimensions and indicators, we developed questions with Likert scale and categorical answers (ordinal scale). We used the same categorical measures for the questions in the expert survey as for the unobtrusive methods. We then invited a group of experts from the ECA to participate in an online survey. This group joined us in a workshop at the ECA in Luxembourg in January 2018. This consultation with internal experts was key to validating the items in the survey. During this step, the indicators were reviewed as part of the process to improve and adapt these. In addition, the first drafts of the index, its dimensions and indicators, were presented at various conferences and scrutinized by academic peers.⁸ All tests were primarily intended to increase the validity of our measurements in order to see whether the indicators really captured the meaning of the concept and to pick up the variation on each dimension. The set of items for the expert survey is listed in Table 2.

Table 2 Measuring the Accountability Powers in the Expert Survey

Formal powers	
1.1	The watchdog is legally recognized as independent
1.2	The board is independent from the executive
1.3	The watchdog has the discretion to investigate anything, which falls within its mandate
1.4	The watchdog has the statutory power to obtain the information necessary for executing its mandate
1.5	The watchdog has the statutory powers to question officials
1.6	The watchdog has the formal powers to sanction actors when it finds irregularities
2.1	How important is legal independence?
2.2	How important is the independence of the board?
2.1	How important is the discretion to investigate anything that falls within its mandate?
2.2	How important is the statutory power to obtain information?
2.1	How important is the statutory power to question officials and witnesses?
2.2	How important is the power to sanction officials when it finds irregularities?
Organizational powers	
3.1	The watchdog has outlined its objectives through a clear mission and a specific strategy
3.2	The watchdog has the appropriate staff numbers to achieve its mission
3.3	The staff has the appropriate professional qualifications and expertise
3.4	The budget of the watchdog is sufficient to execute its mandate
3.5	The watchdog participates in established networks that enhance its capacities
3.6	The leadership structure enables effective leadership
3.7	The board members have the appropriate professional qualifications
4.1	How important is a clear mission and a strategy?
4.2	How important is the number of staff?
4.3	How important is the professional qualifications and expertise of the staff?
4.4	How important is the budget of the institution?
4.5	How important is the participation in relevant networks?
4.6	How important is the leadership structure of the institution?
4.7	How important are the professional qualifications of the board members?
Exercise of powers	
5.1	How productive is the watchdog?
5.2	Is the watchdog recognized as credible by the general public, media, and civil society watchdogs?
5.3	Does the watchdog make creative use of its powers?
5.4	Do these activities lead to the improvement of accountability practices?
6.1	The watchdog is very active and creative beyond its formal mandate
6.2	The watchdog challenges executive bodies to make information available
6.3	The watchdog builds coalitions on specific issues to oversee financial management in the EU
6.4	The watchdog generates standards that provide a base for improved financial management
6.5	The activities of the watchdog lead to improved accountability in the EU
6.6	The watchdog performs adequate given its resources
6.7	There is a clear follow up to the reports and recommendations of the watchdog
6.8	The watchdog plays a vital role as guardian of government agency accountability
6.9	The watchdog is recognized as credible by the general public, media, and civil society watchdogs
7.1	How important is the productivity of the watchdog?
7.2	How important is the effectiveness of the watchdog?
7.3	How important is the credibility of the watchdog?
7.4	How important is the creative use of its mandate?

Eventually, we used Likert five-point scales instead of categorical scales in the expert survey. We experimented with both formats and, while the differences in scoring proved to be minimal, respondents found the Likert scale easier and quicker to answer. Moreover, Likert scales are easier to treat as continuous variables measured on an interval scale.

In addition to the more substantive questions, we also asked our respondents to rate on a Likert scale the importance of each of the various indicators for the accountability practices of the ECA, as we felt that in practice some indicators might very well be more important than others for the accountability power of the watchdog.⁹ This allowed us to gain an idea of the salience of each of the indicators, which could ultimately lead to our assigning different weights to the indicators, or to construct a limited “core” version of the index. We will discuss this further in our concluding paragraph.

4.5. Data collection

The data collection process involved different methods. In the pilot in early 2018, the first online version of the questionnaire was completed by seven internal ECA experts. In the fall of 2018, a refined online questionnaire was sent to the ECA-liaisons at the SAI of all 28 EU member states, as well as to academic experts that have published on financial accountability and the ECA, civil servants at the Internal Audit Office in the European Commission and members of the Budgetary Control committee in the European parliament. Some of those in the latter two categories were also interviewed. All our experts have comprehensive experience with the ECA. This second round also included the items on the salience of each of the indicators. We received 26 completed surveys, which we have used in our analyses. In addition, we collected data on the basis of a desk review of public documents, such as treaties, legislation, annual reports, and existing analyses of the institutional performance of the ECA. The set of items for the indicators for each of the three dimensions is listed in the table in Appendix I.

5. Assessing the accountability powers of the ECA

5.1. Assessing the formal powers of the ECA

We developed six indicators to measure formal powers. We scored the ECA on these indicators in our pilot in two different ways. We first constructed scores for each indicator based on our own coding of the relevant treaties, legal documents, and literature. The results of our coding on the basis of this unobtrusive method are shown in the first column in Table 3. Second, we scored the formal powers on the basis of the expert surveys. We have drawn a distinction between the initial pilot survey among the internal ECA experts, and the later, more extended survey to the external experts. These results are visible in the second and third columns of Table 3. The fourth column gives the aggregate results of the various experts. The final column displays the importance of each of the indicators according to the external experts.

Table 3 shows that the internal experts tend to be slightly more positive than the external experts, except with regard to questioning and sanctioning powers. The latter may be caused by the fact that the internal experts have taken the notion of formal powers very literally—the ECA has no formal right to hear witnesses under oath and it cannot impose sanctions on individual officials or institutions. There are somewhat larger differences between the results of the unobtrusive method and the expert survey. This is mainly due to the fact that we coded the ECA on the basis of substantive, categorical indicators, whereas the experts were asked to rate these indicators on the basis of a general Likert scale. For example, we were rather critical of the ECA’s institutional independence, because the members of the ECA are appointed by the Council, after consultation of the European Parliament, in accordance with a proposal made by the executive of each member state. This implies that the executive, both at the national and the EU level, has a strong and even final say in who is running this watchdog. A major

Table 3 Mean standardized scores on the dimension of formal power (0–100)

	Unobtrusive	Internal experts	External experts	All experts	Importance
Formal autonomy	100	96	88	92	93
Formal independence	25	85	78	82	89
Agenda setting autonomy	100	92	79	86	88
Information powers	88	78	72	75	92
Questioning powers	33	50	68	59	71
Sanctioning powers	75	7	27	17	43

difference can also be observed with regard to sanctioning powers. Formally, the ECA has no sanctioning powers, but it can inform the European Parliament or the Council, who do have formal sanctioning powers or resort to naming and shaming via the media.

5.2. Assessing the organizational powers of the ECA

We have used seven indicators to assess organizational powers. The results are displayed in Table 4.

As is evident from Table 4, a considerable degree of convergence was shown between the scores of the internal and the external experts, and our own coding. The ECA scored reasonably well, except on the leadership indicators—particularly among the internal experts. The ECA is led by a collegial body of 28 Court Members. In its institutional design, full national representation within a college of auditors was opted for, to ensure that different national auditing approaches would be incorporated into the ECA's work (Groenendijk 2004, p. 705). However, this also created decision and management problems because of differences in approaches, different national interests, and the sheer size of the college. It has created top-heavy management, hindering the ECA from becoming more flexible and responsive as an organization. Moreover, at least one-third of its present members, including the president, are former politicians at the national or European level.¹⁰

5.3. Assessing the exercise of accountability powers by the ECA

We assessed the exercise of powers on the basis of six indicators that were all based on expert surveys (see Table 5).

With regard to this dimension, the internal experts tended to be more positive than the external experts. The exception is impact, which they gave remarkably lower scores in comparison to the external experts. According to the annual report of the ECA, implementation rates run from 35% in the first year after a report is issued, to 97% after 4 years. However, an ongoing issue dating back two decades is the absence of a positive assurance statement, mainly in the areas where the management of the EU budget is shared with the member states, covering the payments underlying the EU's accounts. The fact that the absence of a DAS remained without consequences, suggests a limited impact of the ECA.

Table 4 Mean standardized scores on the dimension of organizational power (0–100)

	Unobtrusive	Internal experts	External experts	All experts	Saliency
Strategy	83	86	79	83	78
Size staff	80	79	64	72	70
Quality staff	n.a.	71	75	73	87
Size budget	80	79	63	71	72
Network integration	83	75	73	74	68
Size board	40	29	48	39	78
Composition board	58	43	51	47	72

Table 5 Mean standardized scores on the dimension exercise of power (0–100)

	Internal experts	External experts	All experts	Saliency
Productivity	86	73	80	70
Effectiveness	79	60	70	81
Saliency	71	68	70	
Credibility	79	70	75	90
Creativity	75	57	66	64
Impact	47	69	49	

5.4. The accountability power of the ECA

Figure 1 shows the scores for each dimension, based on the 19 indicators. Radar charts such as these are well suited for showing the accountability gaps and visualizing the features on which watchdogs perform well. The results reveal that the ECA is doing less than well in the areas of questioning and sanctioning powers, leadership, and achieving creativity and impact. The ECA’s low scores in terms of questioning and sanctioning powers may indicate that its strength on these points hinges on the existence of their inter-institutional cooperation with the EP. Further, the scores show that the ECA has developed organizational resources that match its work in a highly complex institutional context of the EU’s multi-level governance. Yet the low leadership scores raise questions about its top-heavy management. In terms of the exercise of power, the ECA scores low regarding impact and creativity. The ECA has an important role in the “verticalized” multi-level accountability system together with the SAI from the member states. The limited impact of the accountability power of the ECA maybe relates to the fragmented structure in which the control of European public expenditure functions under shared management. The low creativity score may indicate that the ECA in performing and disseminating its audits could become more innovative.

6. Indexing accountability powers: Discussion and agenda

6.1. Conceptualizing watchdog accountability powers

The three dimensions that we have distinguished provide a clearer conceptual understanding of watchdog accountability power. The index helps to measure the accountability powers of watchdog institutions and provides a more sophisticated insight into their strengths and weaknesses.

The index can have broader coverage than audit offices only. It can be used to measure the accountability powers of a variety of independent watchdogs, such as integrity offices, SAIs, and ombudspersons, operating at the national or international level. It is a supplement to the existing set of indices that is currently available to assess good governance, such as the Rule of Law Index, Transparency International’s Corruption Index, the Index of Legislative Oversight, the Quality of Government data, the V-Dem project, and the Open Budget Index.

6.2. Measurement issues

This article has defined the concept, its dimensions, and the indicators of watchdog power, but this is only the first level of measurement building. Establishing and enhancing both the validity and reliability of the

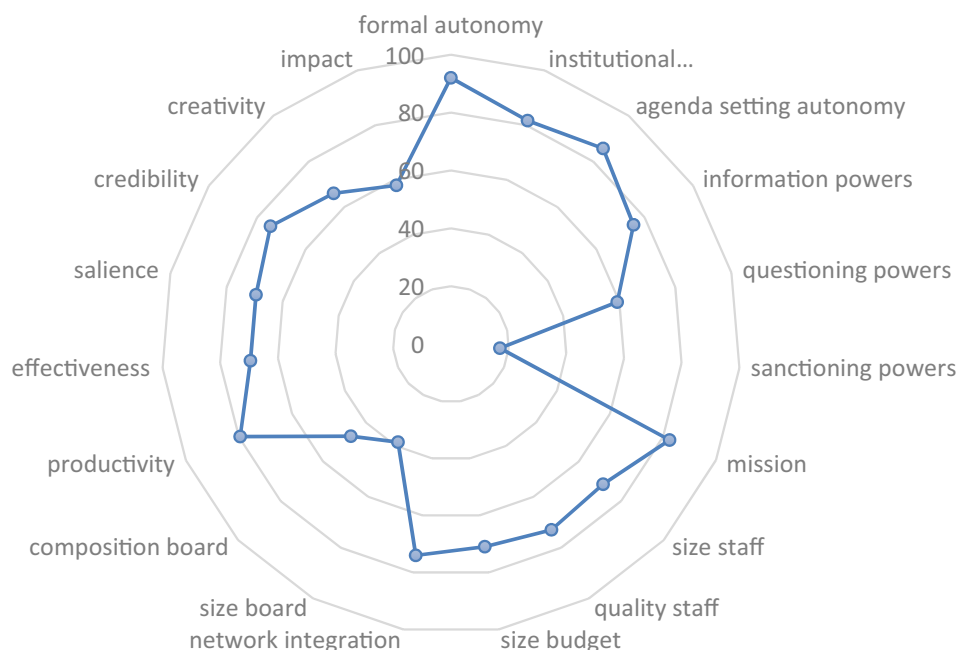


Figure 1 Accountability powers of the European Court of Auditors in 2017.

measurements are crucial next steps in the development of new measures (Adcock & Collier 2001). How do we know whether the accountability indicators generate valid and reliable measures? Applying the index to one single organization does not provide any answers to this question. For this, we need to apply these indicators to a variety of watchdogs in order to obtain enough observations that allow us to test and enhance their validity and reliability. More pilots with different watchdogs are needed to expand the number of cases in order to improve the measurement of the index. More observations from the population of watchdog institutions will enable us to extend tests on the validity and reliability of the items (e.g. with factor analysis) and offer a more solid basis for checking the quality of the measurement and for constructing the indices.

Future work should construct different indices containing combinations of highly rated items and assess them according to specified validation criteria. In doing so, it is possible to test the degree to which the included items represent all key watchdog accountability dimensions, the association of each item in the index with the total index score, the ability of the index to discriminate between poorly and well-performing watchdogs, and the cross-validation of the index scores with other well-established governance indicators. In that way, it can be demonstrated that the index is able to quantitatively capture real-world differences in account holding powers of organizations.

Another measurement issue is the use of unobtrusive measures and expert surveys. We have used the opinions of experts for stipulating the indicators of the watchdog index. The review of expert survey responses showed variation among different expert subgroups in the ratings of specific items to represent these dimensions. Further analysis based on extended surveys may help to examine whether this is the result of error or disagreement among the different categories of expert. The use of categorical scale among the coders and the Likert scale among the experts brought about different assessment of the items on the formal dimension. This has confounded our results and we need to take appropriate steps to avoid this.

The use of different methods to evaluate the quality of watchdog institutions can lead to more reliable assessments. Non-obtrusive measurements, on the basis of document analyses, are easy to perform and to repeat for a large number of institutions and data points. Expert surveys can be rather demanding when, for example, the aim is to compare watchdogs across a broad range of EU member states. It proved to be a somewhat difficult and time-consuming task to identify a broad range of experts and to persuade them into completing a survey on the ECA. Yet, for a valid institutional assessment, different institutional stakeholders need to be included in the survey, which requires a clear strategy for expert selection. Solving these measurement issues is relevant for institutionalizing systematic evaluation of the three dimensions among watchdogs. Selecting alternative proxy measures, that can seek to determine the extent of account holding activity of these watchdogs, can be another approach.

6.3. Are all indicators equally important?

Another question is whether all the indicators are equally important for watchdog accountability power. One way to find a balance between parsimony and sophistication could be to distinguish between core indicators, on which watchdogs need to achieve high scores in order to be relevant, and the more peripheral indicators, which add extra power. Therefore, we asked the experts in our survey to weigh the various indicators for each of the three dimensions. The results were shown in Tables 3–5. According to the external experts, autonomy, independence, information powers, strategy, quality of staff, size of board, effectiveness, and credibility are the most important, need-to-have indicators. Others, such as creativity, are merely nice to have. These expert assessments can provide a basis for developing several versions of the watchdog accountability index, such as a minimal, or core index, and more extended indexes.

Likewise, the index can also be made more sophisticated by giving different weights to the various indicators. For example, it could be argued that formal sanctioning powers are less important than information powers and therefore should count for less. Moreover, having a clear strategy and qualified personnel might be said to be more important than staff numbers, or impact to be more important than creativity. We can compare the differences between estimations of accountability power based on different component construction methods (cf. Mazziotta & Pareto 2016). This requires further structural assessment and more discussion and theory development about the power and effectiveness of watchdog institutions.

6.4. Toward longitudinal and comparative usage

We conducted a single assessment of the accountability powers of the ECA in 2017. A next step could be to track changes in its accountability powers over longer periods of time by repeating these assessments at regular intervals over a number of years. This would provide insight into the dynamics of the ECA's accountability powers: have they increased or not, and in what respect? Another step could be to compare these scores to those of similar institutions in the member states or in the EU. For example, how is the ECA doing, compared to SAI in the member states, or compared to its EU-peers, such as the European Ombudsman or OLAF?

By applying the index across time and across various watchdogs, we can acquire a better understanding of the position of watchdog institutions in the accountability landscape of modern constitutional states. It can also help to gain a more systematic insight into the effectiveness of watchdog strategies and institutional arrangements. After a series of observations, for example, it should be possible to relate the exercise of powers dimension to the various indicators of formal and organizational power. Which of these enhance the impact and the credibility of the watchdog as an accountability forum? This will help to improve our understanding of the workings of these vital institutions in modern, monitory democracies.

6.5. Finding a balance between parsimony and sophistication

In any index, there is tension between parsimony and sophistication. We realize that more sophistication is possible in developing adequate indicators and measurements. On the other hand, too much sophistication will render an index impracticable. Our basic aim in this paper was to develop a systematic index for measuring and comparing executive forum accountability that could be used to compare the power of watchdogs across time and space. A result is a basic tool that can be tailored according to the specific type of watchdog that is assessed. Watchdogs come in various sorts and shapes: auditors, ombudsmen, anti-corruption agencies etcetera. In order to compare watchdogs with the same pedigree, for example, anti-corruption agencies, one could add specific indicators, which are specifically relevant for these agencies.

Creating a valid accountability index is beset by debate and implies solving empirical difficulties and measurement dilemmas. The approach presented here will continue to search for and to apply more appropriate data that match our specific research aims and thus increase validity. As Kaufmann and Kraay (2008, p. 26) wisely observe "All governance indicators include measurement error and so should be thought of as imperfect proxies for the fundamentals of good governance."

ACKNOWLEDGMENTS

We are grateful to Michael Bauer, Luciana Cingolani, Christian Kreuder-Sonnen, Paul Stephenson, and Hartmut Aden for their very helpful comments on earlier versions of this paper.

Endnotes

¹Prominent examples are the Corruption Perception Index (CPI) of Transparency International, the Rule of Law Index of the World Justice Project, The Human Poverty Index, the Human Freedom Index (HFI), the Political Freedom Index of Freedom House, the State of Democracy Assessment of IDEA, and the Parliamentary Power Index (Cooley & Snyder 2015).

²In the literature this is also seen as a form of horizontal accountability (O'Donnell 1998, 2003).

³For a critical review of the rise and use of indicators see Davis et al. 2012.

⁴For a comparable, more extensive operationalization of the independence of agencies see Wonka & Rittberger (2009) and for international bureaucracies the work of Bauer and Ege (2014, 2016).

⁵See Gustavson and Sundström (2018: 8-9) for an overview of the literature.

⁶One could also look at whether the watchdog operates on the basis of a series of explicit and professional standards. Compare Brenninkmeijer et al. (2017).

⁷In general, one could argue that the features of the institutional environment of accountability forums also can play an important role. Accountability forums do not operate in isolation but in conjunction with other institutions and organizations. Is the state a federal state with more accountability institutions? Is there competition among the institutions? Are there NGOs

or voluntary organizations supporting or facilitating the institutional impact of the forums (Blume & Voigt 2011)? Given that this paper is limited to EU institutions, we will not take this factor into consideration here.

⁸See acknowledgments.

⁹A short explanation was provided in the introduction to the question to guide the respondents and to avoid misunderstandings of the salience question.

¹⁰Based on their resumes, retrieved from: <http://www.eca.europa.eu/en/Pages/OrganizationChartList.aspx>. Consulted 27 April 2017.

References

- Adcock R, Collier D (2001) Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *American Political Science Review* 95(3), 529–546.
- Aden H (2015) The European Court of Auditors and its Relationship with National Independent Audit Institutions: The Evolving Audit Function in the EU Multilevel System. *The Palgrave Handbook of the European Administrative System*, pp. 313–326. Palgrave Macmillan, UK.
- Alston P (2000) Towards a Human Rights Accountability Index. *Journal of Human Development* 1(2), 249–271. <https://doi.org/10.1080/713678039>.
- Amttenbrink F (2012) Towards an Index of Accountability for Informal International Lawmakers? In: Pauwelyn Wessel JR, Wouters J (eds) *Informal International Lawmaking*, pp. 337–356. Oxford University Press, Oxford.
- Aucoin P, Heintzman R (2000) The Dialectics of Accountability for Performance in Public Management Reform. *International Review of Administrative Sciences* 66, 45–55.
- Bauer M, Ege J (2014) The Autonomy of International Bureaucracies. In: Soonhee K, Ashley S, Lambright W (eds) *Public Administration in the Context of Global Governance*, pp. 63–84. Edward Elgar, Cheltenham.
- Bauer M, Ege J (2016) Bureaucratic Autonomy of International organizations' Secretariats. *Journal of European Public Policy* 23(7), 1019–1037.
- Blagescu M, De Las Casas L, Lloyd R (2005) *Pathways to Accountability: The Global Accountability Framework*. One World Trust, London.
- Blume L, Voigt S (2011) Does Organizational Design of Supreme Audit Institutions Matter? A Cross-Country Assessment. *European Journal of Political Economy* 27, 215–229.
- Bovens M (2007) Analysing and Assessing Accountability: A Conceptual Framework. *European Law Journal* 13(4), 447–468.
- Bovens M, Hart P, Schillemans T (2008) Does Accountability Work? An Assessment Tool. *Public Administration* 86(1), 225–242.
- Brandsma GJ. (2013) *Controlling comitology: accountability in a multi-level system*. Palgrave Macmillan, Houndsmills.
- Brandsma GJ (2014) Quantitative Analysis. In: Bovens M, Goodin RE, Schillemans T (eds) *The Oxford Handbook of Public Accountability*, pp. 143–158. Oxford University Press, Oxford.
- Brandsma GJ, Schillemans T (2013) The Accountability Cube. Measuring Accountability. *Journal of Public Administration Research and Theory* 23(4), 953–975.
- Brennkmeijer A, Debets R, Hock B, Moonen G (2017) *The Application of Audit Standards in ECAs Work, TILEC Discussion Paper, DP 2017-013*. Tilburg University, Tilburg.
- Busuioc M, Lodge M (2017) Reputation and Accountability Relationships: Managing Accountability Expectations through Reputation. *Public Administration Review* 77(1), 91–100.
- Carpenter D, Krause G (2012) Reputation and Public Administration. *Public Administration Review* 72(1), 26–32.
- Cipriani G (2010) The EU Budget: Responsibility Without Accountability? Brussels: CEPS Paperbacks. Available at SSRN: <https://ssrn.com/abstract=1721515>.
- Cooley A, Snyder J (2015) *Ranking the World: Grading States as a Tool of Global Governance*. Cambridge University Press, Cambridge.
- Coy D, Dixon K (2004) The Public Accountability Index: Crafting a Parametric Disclosure Index for Annual Reports. *The British Accounting Review* 36(1), 79–106.
- Davis K, Fisher A, Kingsbury B, Merry SE (2012) *Governance by Indicators: Global Power through Classification and Rankings*. Oxford University Press, Oxford.
- Desmedt E, Morin D, Pattyn V, Brans M (2017) Impact of Performance Audit on the Administration: A Belgian Study (2005–2010). *Managerial Auditing Journal* 32, 251–275. <https://doi.org/10.1108/MAJ-04-2016-1368>.
- Elsayed K (2011) Board Size and Corporate Performance: The Missing Role of Board Leadership Structure. *Journal of Management & Governance* 15(3), 415–446.
- Feldman D, Eichenthal D (2014) *Art of the Watchdog: Fighting Fraud, Waste, Abuse, and Corruption in Government (Excelsior Editions)*. State University of New York Press, Albany.
- Groenendijk NS (2004) Assessing Member States' Management of EU Finances: An Empirical Analysis of the Annual Reports of the European Court of Auditors, 1996–2001. *Public Administration* 82(3), 701–725.
- Gustavson M, Sundström A (2018) Organizing the Audit Society: Does Good Auditing Generate less Public Sector Corruption? *Administration & Society*, 50(10), 1–25. <https://doi.org/10.1177/0095399716674306>.
- Harlow C, Rawlings R (2007) Promoting Accountability in Multilevel Governance: A Network Approach. *European Law Journal* 13(4), 542–562.

- Head BW (2012) The Contribution of Integrity Agencies to Good Governance. *Policy Studies* 33(1), 7–20.
- Hertogh M, Kirkham R (2018) The Ombudsman and Administrative Justice: From Promise to Performance. In: Hertogh M, Kirkham R (eds) *Research Handbook on the Ombudsman*, pp. 1–15. Edward Elgar Publishing, Cheltenham.
- Hooghe L, Marks G, Schakel AH (2016) Measuring Regional Authority. *A Postfunctionalist Theory of Governance, Vol I*. Oxford University Press, Oxford.
- Hooghe L, Marks G, Lenz T, Bezuijen J, Ceka B, Derderyan S (2017) *Measuring International Authority: A Postfunctionalist Theory of Governance, Vol. III*. Oxford University Press, Oxford.
- Kaufmann D, Kraay A (2008) Governance Indicators: Where Are We, Where Should We Be Going? *The World Bank Research Observer* 23(1), 1–30.
- Keane J (2009) *The Life and Death of Democracy*. Simon & Schuster, London.
- Keane J (2011) Monitory Democracy? In: Alonso S, Keane J, Merkel W (eds) *The Future of Representative Democracy*, pp. 212–235. Cambridge University Press, Cambridge.
- Koop C (2011) Explaining the Accountability of Independent Agencies: The Importance of Political Salience. *Journal of Public Policy* 31(2), 209–234.
- Koop C, Hanretty C (2018) Political Independence, Accountability, and the Quality of Regulatory Decision-Making. *Comparative Political Studies* 51(1), 38–75.
- Kostadinova P (2015) Improving the Transparency and Accountability of EU Institutions: The Impact of the Office of the European Ombudsman. *Journal of Common Market Studies* 53, 1–17. <https://doi.org/10.1111/jcms.12245>.
- Lloyd R, Oatham J, Hammer M (2007) *Global Accountability Report*. One World Trust, London.
- Mazziotta M, Pareto A (2016) On the Construction of Composite Indices by Principal Components Analysis. *Rivista Italiana di Economia Demografia e Statistica* 70(1), 103–109.
- McCubbins MD, Schwartz T (1984) Congressional Oversight Overlooked: Police Patrols Versus Fire Alarms. *American Journal of Political Science* 28(1), 165–179.
- Mechkova V, Lührmann A, Lindberg S (2019) The Accountability Sequence: From De-Jure to De-Facto Constraints on Governments. *Studies in Comparative International Development* 54(1), 40–70.
- Merry SE (2011) Measuring the World: Indicators, Human Rights, and Global Governance. *Current Anthropology* 52(S3), S83–S95.
- Mizrahi S, Ness-Weisman I (2007) Evaluating the Effectiveness of Auditing in Local Municipalities Using Analytic Hierarchy Process (AHP): A General Model and the Israeli Example. *International Journal of Auditing* 11, 187–210.
- Moravcsik A (2010) The Old Governance: Informal Institutions in the EU. In New York University Workshop ‘Rule-Making in the EU and Global Governance (Vol. 5) (Unpublished Paper).
- Morin D (2001) Influence of Value for Money Audit on Public Administrations: Looking beyond Appearances. *Financial Accountability & Management* 17, 99–117.
- O’Donnell G (1998) Horizontal Accountability in New Democracies. *Journal of Democracy* 9(3), 112–126.
- O’Donnell G (2003) Horizontal Accountability: The Legal Institutionalization of Mistrust. In: Mainwaring S, Welna C (eds) *Democratic Accountability in Latin America*. Oxford University Press, Oxford. <https://doi.org/10.1093/0199256373.003.0002>.
- Organisation for Economic Co-operation Development (OECD) (2008) *Handbook on Constructing Composite Indicators Methodology and User Guide*. OECD, Paris.
- Pelizzo R (2011) Public Accounts Committees in the Commonwealth: Oversight, Effectiveness, and Governance. *Commonwealth & Comparative Politics* 49(4), 528–546.
- Pierre J, de Fine Licht J (2019) How Do Supreme Audit Institutions Manage their Autonomy and Impact? A Comparative Analysis. *Journal of European Public Policy* 26(2), 226–245.
- Pollitt C, Girre X, Lonsdale J, Mul R, Summa H, Waerness M (1999) *Performance or Compliance?: Performance Audit and Public Management in Five Countries*. Oxford University Press, Oxford.
- Posner P, Shahan A (2014) Audit institutions. In: Bovens M, Goodin RE, Schillemans T (eds) *The Oxford Handbook of Public Accountability*, pp. 488–506. Oxford University Press, Oxford.
- Power M (1997) *The Audit Society. Rituals of Verification*. Oxford University Press, Oxford.
- Reichborn-Kjennerud K (2014) Performance Audit and the Importance of the Public Debate. *Evaluation* 20, 368–385.
- Reichborn-Kjennerud K, Vabo SI (2017) Performance Audit as a Contributor to Change and Improvement in Public Administration. *Evaluation* 23, 6–23. <https://doi.org/10.1177/1356389016683871>.
- Rosanvallon P (2008) *Counter-Democracy: Politics in an Age of Distrust*. Translated by A Goldhammer. Cambridge University Press, Cambridge.
- Santiso C (2006) Improving Fiscal Governance and Curbing Corruption: How Relevant Are Autonomous Audit Agencies? *International Public Management Review* 7(2), 97–108.
- Scharpf FW (1997) *Games Real Actors Play. Actor Centered Institutionalism in Policy Research*. Westview Press, Boulder.
- Schillemans T (2008) Accountability in the Shadow of Hierarchy: The Horizontal Accountability of Agencies. *Public Organization Review* 8(2), 175–194. <https://doi.org/10.1007/s11115-008-0053-8>.
- Soll J (2014) *The Reckoning: Financial Accountability and the Rise and Fall of Nations*. Penguin Books, London.
- Stephenson P (2015) Starting from Scratch? Analysing Early Institutionalization Processes: The Case of Audit Governance. *Journal of European Public Policy* 23(10), 1481–1501.
- Stephenson PJ (2016) Sixty-Five Years of Auditing Europe. *Journal of Contemporary European Research* 12(1), 467–485.
- Tömmel I (2016) EU Governance of Governance: Political Steering in a Non-Hierarchical Multilevel System. *Journal of Contemporary European Research* 12(1), 406–423.
- Torres L, Yetano A, Pina V (2019) Are Performance Audits Useful? A Comparison of EU Practices. *Administration & Society* 51(3), 431–462. <https://doi.org/10.1177/0095399716658500>.

- Van Acker W, Bouckaert G (2019) The Impact of Supreme Audit Institutions and Ombudsmen in Belgium and The Netherlands. *Financial Accountability & Management* 35, 55–71. <https://doi.org/10.1111/faam.12182>.
- Van de Steeg M (2009) Public Accountability in the European Union: Is the European Parliament able to hold the European Council accountable? *European Integration online Papers* 13: 003a. http://eiop.or.at/eiop/index.php/eiop/article/view/2009_003a.
- Van Looke E, Put V (2011) The Impact of Performance Audits of SAIs: A Review of the Existing Evidence. In: Lonsdale J, Wilkins P, Ling T (eds) *Performance Auditing: Contributing to Accountability in Democratic Government*, pp. 175–208. Edward Elgar, Cheltenham.
- Welch EW, Wong W (2001) Effects of Global Pressures on Public Bureaucracy: Modelling a New Theoretical Framework. *Administration & Society* 33(4), 371–402.
- Wilkins P (2015) Watchdogs as Satellites of Parliament. *Australian Journal of Public Administration* 75(1), 18–27.
- Wilkins P (2019) Watching the Watchdogs: How Auditing is Contributing to Governance. In: Pilcher R, Gilchrist D (eds) *Public Sector Accounting, Accountability and Governance: Globalising the Experiences of Australia and New Zealand*, pp. 80–91. Routledge, London.
- Wilkins P, Phillimore J, Gilchrist D (2017) Benefits and Risks of Collaboration between Watchdogs: The Western Australian Experience. *Policy Studies* 38, 291–310. <https://doi.org/10.1080/01442872.2017.1301414>.
- Wille A (2015) The Evolving EU Accountability Landscape: Moving to an Ever Denser Union. *International Review of Administrative Sciences* 82(4), 694–717. <https://doi.org/10.1177/0020852315589697>.
- Wonka A, Rittberger B (2009) *How Independent Are EU Agencies? RECON Online Working Paper*. Oslo.
- Wonka A, Rittberger B (2010) Credibility, Complexity and Uncertainty: Explaining the Institutional Independence of 29 EU Agencies. *West European Politics* 33(4), 730–752.

APPENDIX I: TABLE WITH INDICATORS & DATA SOURCES

	Source data
Indicators formal powers	
Formal autonomy	Legal documents & Expert opinion
Formal independence	Legal documents & Expert opinion
Agenda setting autonomy:	Legal documents & Expert opinion
Information power	Legal documents & Expert opinion
Questioning power	Legal documents & Expert opinion
Sanctioning power	Legal documents & Expert opinion
Indicators Organizational powers	
Mission & strategy	Documents & Expert opinion
Size of staff	Documents & Expert opinion
Quality of staff	Expert opinion
Size of budget	Documents & Expert opinion
Network integration	Documents & Expert opinion
Board structure	Documents & Expert opinion
Board composition	Documents & Expert opinion
Indicators Exercise of powers	
Productivity	Expert opinion
Effectiveness	Expert opinion
Salience	Expert opinion
Credibility	Expert opinion
Creativity	Expert opinion
Impact	Expert opinion

APPENDIX II: LIST OF INDICATORS AND SCORES FOR DOCUMENT ANALYSES

Formal Powers

Indicator name	Formal autonomy
Question	What is the legal basis for the autonomy of the institution?
Clarification Q	
Score (1)	No legal guarantee for autonomy
Score (2)	Administrative act guarantees autonomy
Score (3)	Secondary legal act guarantees autonomy
Score (4)	Formal law guarantees autonomy
Score (5)	Constitution or treaty guarantees autonomy

Indicator name	Formal independence
Question	Has the executive (such as Cabinet, Commission, Council, member states) formal powers to (re)appoint the board?
Clarification Q	
Score (1)	Board is (re)appointed by the executive.
Score (2)	Board is (re)appointed by executive on the basis of a nomination by an independent committee.
Score (3)	Board is (re)appointed by Parliament on the basis of a nomination by the executive.
Score (4)	Board is (re)appointed by Parliament on the basis of a nomination by an independent committee.

Indicator name	Agenda setting autonomy
Question	Has the institution the discretion to investigate anything, which falls within its mandate?
Clarification Q	
Score (1)	Upon request from executive branch only.
Score (2)	Upon request from legislative only.
Score (3)	Upon request, but from a variety of external actors (legislative, civil society/general public.
Score (4)	Upon request, but internal decision.
Score (5)	Fully autonomous in setting its agenda.

Indicator name	Information power
Question	Has the institution the statutory power to obtain the information necessary for executing its mandate?
Clarification Q	
Score (1)	The institution has no formal powers to obtain information.
Score (2)	The institution has the right to access the information it needs to execute its mandate.
Score (3)	The institution may request or obtain any information it deems necessary to execute its mandate.
Score (4)	The institution may also access and search premises and ICT and retrieve any dossiers and data against the will of the executive body.

Indicator name	Questioning power
Question	Has the institution statutory powers to question officials and witnesses?
Clarification Q	
Score (1)	The institution has no formal power to subpoena and question witnesses.
Score (2)	The institution has the right to subpoena witnesses.
Score (3)	The institution has the right to subpoena and hear witnesses under oath.

Indicator name	Sanctioning power
Question	Has the institution formal powers to sanction actors when it finds irregularities?
Clarification Q	
Score (1)	The institution has no formal sanctioning power.
Score (2)	The institution has no formal sanctioning power but may engage in public naming and shaming.

(Continues)

Indicator name	Sanctioning power
Score (3)	The institution has no formal sanctioning power but can refer irregularities, deviations or unlawful practices to other authorities, which have sanctioning powers.
Score (4)	The institution itself can impose formal sanctions, such as fines, disapproval of budget, or disciplinary measures.

Organizational Powers

Indicator name	Mission & strategy
Question	Has the institution outlined its objectives through a clear mission and a specific strategy?
Clarification Q	
Score (1)	The institution has no mission or strategy.
Score (2)	The institution has no explicit mission or strategy, but some strategic objectives may be implicitly inferred.
Score (3)	The institution has issued some statements, which speak to a mission or strategy.
Score (4)	The institution has an explicit mission, but this has not been translated in strategic objectives.
Score (5)	The institution has an explicit mission and strategic objectives, but these are not operationalized.
Score (6)	The institution has a clear mission and specified and measurable strategic objectives that relate to a specific period.

Indicator name	Staff
Question	Is the number of staff the institution has at its disposition sufficient to execute its mandate?
Clarification Q	<i>A proxy for this can be: fte staff/fte public sector.</i>
Score (1)	The number of staff is very low given the size of the executive.
Score (2)	The number of staff is low given the size of the executive.
Score (3)	The number of staff is limited given the size of the executive.
Score (4)	The number of staff is adequate given the size of the executive.
Score (5)	The number of staff is very adequate given the size of the executive.

Indicator name	Budget
Question	Is the budget of the institution sufficient to execute its mandate?
Clarification Q	<i>A proxy for this can be: budget institution/budget public sector.</i>
Score (1)	The budget is very insufficient given the task.
Score (2)	The budget is very limited given the task.
Score (3)	The budget is limited given the task.
Score (4)	The budget is adequate given the task.
Score (5)	The budget is very adequate given the task.

Indicator name	Network integration
Question	Does the institution participate in established networks that enhance its capacities?
Clarification Q	<i>Examples are networks of similar institutions in other countries or at other levels of governance, cooperation with other watchdog institutions or parliamentary committees, or with professional associations and academic institutions.</i>
Score (1)	The institution does not participate in any relevant network.
Score (2)	The institution is a passive member of one or two relevant networks.
Score (3)	The institution is a passive member of several relevant networks.

(Continues)

Indicator name	Network integration
Score (4)	The institution actively makes use of one or two relevant networks.
Score (5)	The institution actively makes use of several relevant networks.
Score (6)	The institution actively makes use of a large variety of relevant networks, at various levels of governance (nationally and internationally).
Indicator name	Board structure
Question	Does the leadership structure enable effective leadership?
Clarification Q	
Score (1)	The institution has no clear leadership structure.
Score (2)	The institution is governed by a very large board (10 members and more).
Score (3)	The institution is governed by a medium size board (4 till 10 members).
Score (4)	The Institution is governed monocratically (President or Auditor General).
Score (5)	The board consists of two or three persons.
Indicator name	Board composition
Question	Have the board members the appropriate professional qualifications?
Clarification Q	
Score (1)	All board members are former politicians, who are appointed solely on the basis of their loyalty to the ruling political establishment.
Score (2)	Most board members are former politicians, with little relevant professional qualifications.
Score (3)	A substantial number of board members are former politicians, with little relevant professional qualifications.
Score (4)	A substantial number of board members are former politicians, but with a relevant degree or relevant work experience.
Score (5)	Most board members have a professional background with a relevant degree and relevant work experience.
Score (6)	All board members have a professional background with a relevant degree and extensive relevant work experience.