



Contents lists available at ScienceDirect

## World Development

journal homepage: [www.elsevier.com/locate/worlddev](http://www.elsevier.com/locate/worlddev)

# Does accountability enhance service delivery? Assessment of a local scorecard initiative in Uganda

Wil Hout<sup>a,\*</sup>, Natascha Wagner<sup>b</sup>, Binyam A. Demena<sup>a</sup>

<sup>a</sup> International Institute of Social Studies, Erasmus University Rotterdam, the Netherlands

<sup>b</sup> Department of Economics, Nijmegen School of Management, Radboud University Nijmegen, the Netherlands



## ARTICLE INFO

## Article history:

Accepted 15 June 2022

## ABSTRACT

This article assesses whether the Local Government Council's Scorecard Initiative, implemented in Uganda since 2009, achieved its intended impact of enhancing service delivery by providing information on the performance of local government. We analyse a district-level panel dataset (2005–2016) with administrative data, as well as Afrobarometer data on citizen perceptions (2005–2017). Empirically, we exploit the phasing in of the scorecard for a meticulous difference-in-difference framework with district-specific trends. The results show some small measurable impacts of the scorecard along the so-called 'long route of accountability' on public service delivery. Scorecard districts appear to spend less of their budgets in comparison with non-scorecard districts. This points to greater budgetary restraint of local government councils in scorecard districts. Although no direct impacts on service delivery can be detected, districts with more electoral competition in their constituencies perform better on one service-delivery indicator, the primary school leaving exam pass rate. Concomitantly, the scorecard impacts on perceptions of corruption, as citizens of scorecard districts perceive the local councillors as less corrupt compared to citizens of non-scorecard districts. This result can be interpreted as an indication of the trust-enhancing effect of government scorecards and civic engagement. Overall, our results provide a quantitative contribution to the literature on accountability by demonstrating that civil society reporting mechanisms about the performance of political representatives only trickle down slowly to improved services. The findings suggest that the sustained implementation of instruments to provide citizens with more information about their political representatives may have a positive impact on civil society perceptions as well as relevant political and policy outcomes. Like earlier research, we find that impacts also depend on political competitiveness, thus highlighting the positive role of democracy.

© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

There is broad agreement that representative democracy depends on mechanisms of accountability. Accountability, defined as 'a social relationship in which an actor feels an obligation to explain and to justify his or her conduct to some significant other' (Bovens, 2005: 184), is a crucial component of democratic rule. Representative democracy consists of a chain of principal-agent relations, in which political influence is shifted upwards, first from citizens to members of representative institutions and next to holders of executive offices (Bovens, 2005: 192). Mechanisms of accountability perform a variety of functions: they serve as mechanisms of democratic control, help to maintain integrity of public officeholders, and contribute to improving the performance of

political institutions (Bovens, 2005: 192). Whereas elections are key among accountability mechanisms in representative democracy, many observers argue that electoral mechanisms are deficient, because they do not allow for direct citizen influence on representatives in the period between elections (cf. Van Reybrouck, 2016: 24).

The awareness that elections are insufficient tools for enhancing the performance of political institutions has led to a search for new accountability mechanisms. The involvement of principals in directly assessing the performance of agents is a tool that has been used in a variety of cases across the world. In the business environment, the so-called 'balanced scorecard' was developed to provide managers with feedback about a variety of performance indicators, including customer satisfaction. Kaplan and Norton (1992: 74), the initiators of the balanced scorecard, argued that '[d]epending on customers' evaluations to define some of a company's performance measures forces that company to view its performance through customers' eyes'. In a similar vein as the private

\* Corresponding author at: International Institute of Social Studies, Erasmus University Rotterdam, Kortenaerkade 12, 2518AX The Hague, The Netherlands.

E-mail address: [hout@iss.nl](mailto:hout@iss.nl) (W. Hout).

sector, public sector actors started to reflect on ways to gauge citizens' assessment of public service provision and to use citizen inputs as a way of enhancing accountability (Wray and Hauer, 1997; Epstein et al., 2006).

The literature on policy making, public service delivery and accountability contains a good number of analyses of how information provision (for instance, on the performance of incumbent politicians or public service providers) impacts on political behaviour of voters/citizens or representatives (e.g., Dunning et al., 2019; see further section 2). This literature focuses on the so-called 'long route of accountability', where voters/citizens try to enhance the quality of service delivery by applying pressure on political decision makers. Next to this route, the literature distinguishes the 'short route' that relates to 'client power', where citizens exert direct influence on service providers, for instance by giving inputs on the quality of services or by changing providers (World Bank, 2003: 47-51). The 'long route' is relevant, in particular, in situations where there is limited choice of service providers, as in most instances of public service provision.<sup>1</sup> Our approach in this paper differs from other studies in that we do not concentrate on (social) accountability mechanisms that provide information about service delivery providers, which is the dominant focus in the literature as evidenced in reviews by Joshi (2013), Kosack and Fung (2014), Fox (2015) and Börang and Grimes (2021). Rather, we focus on information provision about the *performance of politicians* as important intermediaries along the 'long route of accountability'. As Waddington et al. (2019: 6) have phrased it, information provision on politicians' performance can be seen as an attempt to "shorten the long route" of citizen-state accountability.

In this article, we present the findings of a study on the extent to which the introduction of a new accountability instrument – a local government scorecard – impacted the outputs of local government in Uganda. This study addresses two gaps in the literature on 'transparency for accountability' interventions (Kosack and Fung, 2014: 68). First, most studies focus on the input side as they analyse the effects of enhanced information about political performance on the *perceptions* of individual voters or representatives, but leave the *outcome* of improved information provision, for instance on service delivery, undiscussed (see the literature review in section 2). Secondly, despite the preference among researchers to conduct experimental studies (cf. section 2), there is still major controversy over whether information provision has actually produced better accountability of representatives (Dunning et al., 2019). By focusing on the impact of the Ugandan local government scorecard on policy making and service provision, we contribute to the knowledge on the outcome of enhanced accountability of elected representatives by moving away from analysing perceptions to studying material effects (cf. Deiniger and Mpuga, 2005).

This article presents the results of a longitudinal analysis of the impact of the Local Government Council's Scorecard Initiative (LGCSCI, hereafter referred to as the 'scorecard initiative'), which the Ugandan NGO Advocates Coalition for Development and Environment (ACODE) has implemented, with external support, since 2009. The goal of the scorecard initiative is to deepen democratic governance by means of yearly, evidence-based performance assessments of local government councils and councillors. The scorecard initiative does not, however, operate exclusively at the level of councils and individual council members. Because ACODE organises annual information campaigns about the scorecard, broader awareness of the initiative is stimulated. The scorecard initiative thus has features of what Fox (2015: 352) has termed a 'strategic' intervention, which is 'an approach with a theory of

change that takes into account the relationship between pro-change actions and eventual goals by specifying the multiple links in the causal chain'. By contrast, tactical approaches are typically emphasising 'local-level dissemination of information on service-delivery outcomes and resource allocation to under-represented stakeholders' (Fox, 2015: 352). According to Fox (2015: 350), strategic interventions 'are more promising than tactical approaches for leading to tangible development impacts'. This additional characteristic of the scorecard initiative makes an analysis of the instrument's impact on the quality of service delivery especially relevant.

Our focus on this scorecard is informed by three considerations. First, the Ugandan scorecard initiative is currently one of the most elaborate schemes aimed at enhancing accountability of representatives at the local level. Secondly, the scorecard has been used for over a decade, which facilitates analysis over a longer period. Thirdly, the scorecard is only implemented in about 30 % of all districts in Uganda – a feature that allows a comparison between scorecard and non-scorecard districts.

We study the impacts of the scorecard initiative on two sets of outcomes. First, we analyse 12 years of district-level administrative data to identify differences in budgetary outcomes and service delivery in districts that implemented the scorecard, compared to a control group that did not use the instrument. This choice of outcomes is motivated by the theory of change used by the scorecard implementers. They argue that the scorecard was introduced with the explicit objective to bring about improvements in service delivery. Secondly, we study the impact of the scorecard initiative on citizens' perceptions of local councillor performance, using seven rounds of the Afrobarometer survey. Since the intervention is civil society-led it is expected that it affects citizens' perceptions. Our research indicates that the sustained implementation of instruments to provide citizens with more information about the performance of their political representatives has rather small impacts on service delivery, mainly by affecting budgetary discipline. Concomitantly, we show that citizen perceptions are more favourable towards councillors in scorecard districts, as councillors are perceived as being less corrupt.<sup>2</sup> In line with earlier results reported by Grossman and Michelitch (2018), we find that impacts depend on political competitiveness at the district level.

This remainder of the paper is structured as follows. Section 2 aims to take stock of the recent empirical literature on accountability, information provision and scorecards. The third section describes the rationale and objectives of the scorecard initiative and focuses on the way the instrument has been used in Uganda since 2009. Section 4 introduces our data and presents the empirical identification strategy used in this study, i.e., a quasi-experimental difference-in-difference approach. Section 5 reports the main findings, an assessment of parallel trends and placebo interventions. In section 6, we formulate our main conclusions and discuss the implications of our findings for the literature on accountability.

## 2. Accountability, information provision and scorecards: Overview of recent research

Our literature search on the impact of different forms of information provision and performance assessment (including score-

<sup>1</sup> Here we leave aside Fox's (2015: 347) valid criticism that the 'short route of accountability' may not at all be short in the case of 'government failure'.

<sup>2</sup> Note that this finding is not driven by social desirability bias as the data to assess councillor corruption has been collected independently of the scorecard initiative by Afrobarometer. Only in the unlikely event that the scorecard initiative influenced everyday life and perceptions, would we expect biased responses. Yet, based on our own observations and following Grossman and Michelitch (2018: 280) we understand the activities as part of the scorecard initiative as 'a relatively "weak dissemination", because performance information inappreciably reached constituents, if at all.' Similarly, Tumushabe et al. (2011: 52-53) indicate that level of civic awareness of the scorecard is generally low.

cards) as mechanisms of public accountability identified 35 studies that incorporate political functionaries or institutions, but further differ significantly in terms of research design, methods and units of analysis. Most of the reviewed studies attempt to determine whether (enhanced) information provision impacts on voters' electoral choices or on the behaviour of political representatives, such as local council members, mayors and community leaders. A clear minority of the studies focuses on the *effect* that information provision has on the activities of political functionaries or institutions in relation to service delivery.

With regard to research methods, about half of the reviewed studies use field experiments to determine the impact of information provision; all experimental studies focus on elections or actions of representatives, almost exclusively related to local governments. Other quantitative studies apply surveys or use panel data extracted from existing data sets on government performance. Qualitative studies are either based on interviews, use a case-study design or apply policy analysis. A final group of studies has performed secondary analysis of existing studies, either through a (qualitative) literature review or (quantitative) meta-analysis.

The studies in the review have produced considerably different results, with a sizeable number reporting a positive impact of information provision about the performance of representatives on voters/citizens. A smaller number of studies are either undecided about the size or nature of the impact or find that performance information has no impact at all on voter behaviour or citizen perceptions.

The majority of field experiments report a positive impact of information provision about incumbents or candidates on voter behaviour (Banerjee et al., 2011; Chong et al., 2015; Gottlieb, 2016; Adida et al., 2017; Arias, Balan et al., 2019; Bidwell et al., 2020; Platas and Raffler, 2019; Adida et al., 2020). Pande's (2011) literature review of 13 experimental studies underscores the generally positive impact of information on voters in low-income settings. Focusing on different units of analysis (respectively, mayors and local councillors), Ferraz and Finan (2008), Grossman and Hanlon (2014), Grossman and Michelitch (2018) and Raffler (2019) also find positive effects of information provision. Buntaine et al. (2018), Buntaine et al. (2019) and Boas et al. (2019b) report different effects of information for different groups of representatives and voters, respectively.

Other experimental analyses of voter behaviour find no effect of performance information, or are undecided about their impact (Humphreys and Weinstein, 2012; James, 2011; Adida et al., 2019; Arias, Larreguy et al., 2019; Boas et al., 2019a). Further, a rigorous meta-analysis of five studies in the so-called Metaketa project (Dunning et al., 2019) reports that there are few indications that information provision about incumbents shapes voters' evaluation of candidates or electoral behaviour. Wantchekon and Vermeersch's (2011) field experiment in Benin finds a negative impact of exposure of voters to electoral platforms that emphasize the provision of public services, while clientelistic platforms appear to reward candidates.

Analyses of electoral behaviour and citizen perception, based on surveys, interviews, administrative data, policy analyses and literature reviews, produce similar findings as the experimental studies. Although the majority of the surveyed studies concludes that information provision impacts positively on citizen and voter positions (Ravindra, 2004; Gainer, 2015; Harding, 2015; Thinyane and Siebörger, 2017), some studies fail to find an impact (Brixi, 2010; Ashworth, 2012).

Finally, there is a smaller number of studies that focus on service delivery, policy outputs and incumbent behaviour; these, similarly, show diverse results. While Askim's (2007) analysis of survey data from Norway reports a positive effect of information on the behaviour of local councillors, the literature reviews by

Devas and Grant (2003) and Ashworth (2012) are sceptical about the impact of information provision on service delivery. At the same time, various studies conclude that the introduction of new accountability mechanisms, including variants of scorecards, have positive effects on the quality or quantity of service delivery across contexts (Besley and Burgess, 2002; Thampi, 2011; Raffler, 2019). Kosack and Fung's (2014) review of 16 experimental studies finds that two-thirds of reviewed transparency initiatives, aimed at providing information (referred to as 'transparency for accountability'), had a positive effect on service delivery, while one-third showed no impact. Overall, they conclude that initiatives that provide information about inputs are more successful than those that focus on outputs delivered by service providers. In line with this, an experimental study of the Ugandan local government scorecard by Grossman and Michelitch (2018) reports that random dissemination of the results to constituents improved local politicians' performance of legally defined duties, related to their legislative role, contact with voters and attendance of representative council meetings. There seems to be evidence that politicians in competitive constituencies were more active in promoting development projects in their districts; yet, the authors did not find an impact of information provision on the quality of service delivery in education and health. Wild and Harris (2011) report that the introduction of a similar community scorecard in Malawi had no observable effect on service delivery.

The latter three studies are very relevant in light of our research objectives, but it is also clear that they have limited focus on the decisions of political institutions or representatives on service delivery in the 'long route of accountability' (World Bank, 2003: 49). While Kosack and Fung (2014) synthesize studies on the performance of a range of service providers across a variety of countries, their conclusions do not necessarily relate to political decisions on service delivery. Grossman and Michelitch's (2018) study of Uganda mainly concentrates on the performance of individual council members instead of the decisions of local councils. Wild and Harris' (2011) case study of two sectors in two Malawian districts leads to a comparison of certain district characteristics with outcomes in agriculture and education but does not present generalizable conclusions about decisions on service delivery in local councils.

Thus, the existing literature leads to two conclusions. First, the overview demonstrates that the majority of studies has focused on the effect of information provision on voter/citizen views or behaviour vis-à-vis elected representatives. Although the importance of this cannot be denied, it is clear that other dimensions of the accountability relationship between voters/citizens and their representatives have received relatively less attention. Secondly, the comparatively limited attention in previous studies to political decision making about service delivery indicates the relevance of a research focus on the impact of accountability tools on the output side. Our research on the Ugandan local government scorecard is intended to make a contribution to the literature by linking the implementation of the scorecard, as a tool of information provision about local councillor performance, to policy outputs (budgetary outcomes and service delivery).

### 3. The Ugandan context: ACODE and the local government council's score card initiative

The Local Government Councils' Score Card Initiative was set up in 2009 with support from the Deepening Democracy Programme (DDP), a basket fund for supporting initiatives to improve democratic governance in Uganda, established by Denmark, Ireland, The Netherlands, Norway, Sweden and the UK. The scorecard initiative has been implemented by ACODE, a Kampala-based NGO

focused on public policy research and advocacy, in collaboration with the Uganda Local Governments Association (ULGA). The initiative is a long-term programme to strengthen citizens' demand for effective public service delivery and accountability. The longevity of the initiative and the manifold reports of the implementing NGO show that the scorecard has been implemented without interruption, was expanded across the country, is based on a reliable and replicable tool, and shows a high level of rigour.<sup>3</sup> In this section, we pay attention to the theory behind the scorecard initiative, to the scorecard instrument itself, and to the way in which the scorecard has been used.

### 3.1. Theory behind the scorecard initiative

A fundamental element of the scorecard initiative is the 'central premise [...] that by monitoring the performance of LGCs [Local Government Councils] and providing information about their performance to the electorate, citizens will be empowered and encouraged to demand accountability from their local elected officials' (Muyomba-Tamale and Cunningham, 2017: 190). A key element in the causality chain assumed by the scorecard initiative is the notion that strengthening the accountability of local councillors by providing information would impact positively on those councillors' political decisions about service delivery.<sup>4</sup>

The problem definition underlying the scorecard initiative is that

the delivery of public services is less than desirable at best or has malfunctioned at worst. Improvements in key service delivery indicators in the areas of health, education, agriculture and roads are not considered proportionate with the levels of public investment in these areas (Tumushabe et al., 2013: 17).

ACODE's analysis of insufficient service delivery focuses on two interrelated factors. First, lacking state capacity is felt to lead to low-quality public policy making and poor service delivery (Tumushabe et al., 2013: 18). Secondly, incentives of policy makers are seen to lead to political clientelism and patronage (Tumushabe et al., 2013: 19–20).

ACODE directed its attention to the 'demand side', which implies that citizens 'are empowered to demand for better performance from governmental and other institutions and leaders' (Tumushabe et al., 2010: 8). Existing monitoring was felt to be dominated too much by the 'supply side', based on the assumption that public service delivery could be improved by strengthening the oversight of local government institutions over service providers such as schools and hospitals (Tumushabe et al., 2010: 5). According to the logic of the scorecard initiative, existing mechanisms for horizontal accountability are supplemented by vertical accountability instruments, which should lead to more influence of citizens, as they were expected to express their preferences in their voting on the basis of better information about councillors' performance. The scorecard should provide 'a combination of regular assessments of performance of elected leaders and provision of performance information to citizens' (Tumushabe et al., 2013: 2). An underlying assumption of the scorecard is that its use would lead to better awareness among political leaders of their roles, as well as more awareness among citizens about the responsibilities of those political leaders (Tumushabe et al., 2013: 68).

According to its initiators, the positive impact of the scorecard could be impeded by various factors, some of which operate internally to local governments, while others are external and operate

at the national level. Internal factors relate to conflicts deriving from the existence of multiple leadership positions at the local level, the low level of revenue collection and lack of financial autonomy, and the failure of multi-party politics at the local level (Tumushabe et al., 2013: 68–70). Factors concerning the embedding of local governments include the distortions inherent to Uganda's decentralization policy, related to the use of decentralization for clientelistic purposes, as well as the central control of the financial resources of local government (Tumushabe et al., 2013: 70–71). An initial assessment of the impact of the scorecard on service delivery, performed by ACODE with a focus only on scorecard districts, showed that higher scoring councils reported better results in terms of exam scores in primary education, allocations to the development budget for roads, and allocations to the development budget for education (Tumushabe et al., 2013: 33–39).<sup>5</sup>

The theory of change underlying the scorecard initiative holds that the introduction of the scorecard instrument would have a positive impact on public service delivery through five interconnected channels.<sup>6</sup> First, dissemination of the scorecard results would increase the available information on councillors' performance. Second, information dissemination, together with capacity building activities focused on increasing citizens' demand for accountability, would lead to strengthened civic consciousness about the role and performance of councillors. Third, and consequentially, these activities would lead to increased citizen demand for better services. Fourth, the use of the scorecard would create incentives for local councillors to deepen knowledge of their formal roles and responsibilities and work on strengthening their capacity, which would ultimately result in improved performance. Finally, the information obtained through the scorecard initiative would result in the greater ability of key stakeholders, such as civil society organizations, to lobby for better services at the sub-national level.

### 3.2. The scorecard instrument

The scorecard initiative has developed a range of separate scorecards to assess the performance of key actors at the local government level in Uganda: chairpersons of district councils, speakers and deputy speakers of those councils, district councillors and the district councils. The scorecards are arranged in a number of broad categories, including the legislative role, contact with the electorate, participation in communal and development activities, and service delivery on national priority programme areas. The centrality of service delivery to the scorecard initiative is reflected in the presence of questions related to this aspect in the scorecards for all office holders and the district council. In addition to the general categories, the scorecards also contain function-specific assessments, related to political leadership (for district chairpersons), presiding and preservation of the order in the district council (for speakers and deputy speakers), and the accountability role as well as planning and budgeting (for local councils). By allocating scores to a range of items under the different categories, scorecards result in ratings between 0 and 100 (see Tumushabe, 2010: 35–46 and the Supplementary Materials, Figure A1 for an example of a scorecard).

### 3.3. Implementation

Implementation of the scorecard initiative involves several main stages. After a preparation phase, during which the participa-

<sup>3</sup> An up-to-date list of publications on the scorecard by ACODE is available at [ACODE \(2021\)](#).

<sup>4</sup> The chain of causality as described here, along with its underlying assumptions, is presented in more detail in [Tumushabe et al. \(2010: 1–11\)](#).

<sup>5</sup> The analysis was based on a comparison of descriptive statistics and trends across scorecard districts. No counterfactual analysis was provided.

<sup>6</sup> This theory of change was largely implicit in the documents produced by ACODE. We reconstructed it on the basis of the various project documents and publications. A more elaborate discussion is provided in [Hout et al. \(2016: 78–80\)](#).

tion of key stakeholders is ensured, scorecard data are collected on the performance of district councils, chairpersons, (deputy) speakers and councillors. Next, the initiative leads to the publication of feedback reports on the assessments, both at the level of included districts and across Uganda. Finally, in the 'outreach and advocacy phase', ACODE organises capacity-building activities aimed at increasing the effectiveness of councils and councillors on the one hand and citizens' demand for accountability on the other (Muyomba-Tamale and Cunningham, 2017: 192-193; Bainomugisha et al., 2020: 21).

Data collection as part of the annual scorecard process starts with the filling of the scorecard by ACODE researchers on the basis of interviews with local councillors and key informants, civic engagement meetings, field visits and document reviews. First, the scorecard assesses how elected political leaders and the district council as an institution perform their tasks and responsibilities specified in the Constitution, the Local Governments Act and other legal provisions. Next, data are verified in field visits of two-district-based researchers and a lead researcher to service-delivery units, as well as in rounds of interviews with service consumers. ACODE and district-based researchers organize focus-group discussions with community groups consisting of youth, women and a mixed group to discuss councillors' performance in all sub-counties of districts involved in the scorecard (Tumushabe et al., 2010: 27-33; Bainomugisha et al., 2020: 22-29).<sup>7</sup>

Outreach and advocacy are important elements of the scorecard initiative. As we alluded to in the introduction, these elements, in particular, give the initiative the character of a strategic intervention (Fox, 2015: 352). The publication and dissemination of the annual scorecard reports and the ensuing advocacy activities (including media campaigns, public events and actions of civil society) aim to make participants in local government councils aware of the assessments made by citizens, both in their roles of voters and users of local services (Bainomugisha et al., 2020: 30). Thus, the initiative attempts to strengthen the incentives of local government actors to work on strengthening their capacity and improving their work. By attempting to influence participants in local government, the initiative aims to include what Fox (2015: 352) has called the 'multiple links in the causal chain'. Yet, ACODE acknowledges that most impact of the outreach and advocacy strategy seems to be at the level of NGOs, and that the initiative is still far from large-scale citizen engagement (Tumushabe et al., 2011: 52-53).

The scorecard has been implemented in a stepwise fashion. Table 1 presents the districts included in the scorecard initiative and the year of entry. There were five waves of entry: 2009, 2010, 2012, 2014 and 2016. The first nine districts were included in 2009, and their number was increased until the current number of 35 districts was reached, comprising almost 30 % of all districts.<sup>8</sup>

Selection of the districts for implementation of the scorecard initiative was not random, but aimed to reflect representation of districts along the following four criteria:

- 1) Uganda's division into four major regions has been considered to achieve geographical representation.
- 2) The history of decentralization has been accounted for as this has led to the breakup of districts into smaller ones. To get a representation of the temporal dimension of district

creation, three groups of districts were selected: (i) districts that existed at independence, (ii) districts created in the 1980s, and (iii) districts created since 2000.

- 3) Both model districts and historically marginalized districts have been sampled to achieve representation across the performance-disadvantage divide.
- 4) To ensure the sustainability of the scorecard approach, only districts were selected where research teams could be recruited from local CSOs.<sup>9</sup>

Across districts the same scorecard was implemented and the same training modules were applied, assuring that all interviewers had the same standard. In addition, the overall implementation was monitored by ACODE. There is no indication of any systematic deviations in assessment.<sup>10</sup>

## 4. Data and identification strategy

### 4.1. Data sources

We use district-level administrative data from 2005 to 2016 to assess the impact of the scorecard initiative on service delivery. We study three sets of service delivery outcomes<sup>11</sup>: (a) budgetary decisions related to public spending (total and per capita), the share of the budget spent, the delay in reporting to the Ministry of Local Government and the local contribution to revenues, (b) the primary school leaving exam pass rate, and (c) the number of health centres and hospitals that are available (scaled to population size). In line with the theory of change, the indicators collected by the scorecards and based on earlier analyses within scorecard districts, we expect possible impacts on budgeting, education and health (Tumushabe et al., 2010, 2013).<sup>12</sup>

Further, we make use of seven rounds of *Afrobarometer* (2020) data to assess how the Ugandan citizens perceive the quality of democracy and governance at the local level: we have a repeated cross-section for two years prior to the introduction of the scorecard (2005 and 2008) and five survey rounds (2010, 2011, 2012, 2015 and 2017) during the intervention. While being representa-

<sup>9</sup> While the last criterion ensures the quality of the implementation of the scorecards, it implies that a systematic element was part of the district selection process, i.e., the scorecard was not randomly implemented. This is akin to the context of a study on the privatization of water (Galiani et al. 2005), where a similar difference-in-difference model was used. In comparison with Galiani et al. (2005), our analysis is even more demanding since we do not only account for district-specific characteristics and district fixed effects but also for district-specific time trends to rule out bias stemming from the fact that some districts were chosen to participate, while others were not.

<sup>10</sup> Thus, any resulting measurement error can be assumed to be classical.

<sup>11</sup> Across service delivery outcomes we focus on the provision of the service by the local government and not on demand side aspects, which are important in themselves but not part of the analysis at hand.

<sup>12</sup> For education outcomes the district scorecard explicitly assesses enrolment in primary schools, school completion, academic performance and the promotion of girls. Furthermore, the scorecard assesses teacher recruitment and retention, utilization of funds, and school reports. Concerning health care, the construction and functionality of health care units as well as staff recruitment and retention, availability of essential drugs, evidence of immunization/family planning services, availability of HIV/AIDS services and maternal/child health care services are covered by the scorecard. The rationale to choose education and health services as measurable outcomes stems from its detailed inclusion in the scorecard assessment, and the fact that the local council controls local revenues, transforms these into budget allocations, out-turns and finally education and health services. Effects in these two sectors are within the realm of local governments. From an analytical point of view we have chosen these service outcomes since systematic information was available across districts and over time allowing us to conduct the counterfactual difference-in-difference analysis. Although assessment of other indicators, such as teacher absenteeism or information about hospital staff, would have been preferable, such data were not readily available for all districts. We further note that the original theory of change does not make any assumptions on the size of the impact in the education and health sector; it merely expects improvements in both areas.

<sup>7</sup> We are not aware of any other local initiative in Uganda that systematically, across districts and over time, follows up on councillors and assesses their performance. Such type of intervention is expensive, logistically demanding and would be known in expert circles. Therefore, we are confident that there is no similarly scoped intervention ongoing.

<sup>8</sup> The total number of districts was steadily increasing over the study period due to the split of larger districts into smaller ones. Our final district sample consists of 111 districts. Note that the number of districts has increased again since the completion of our study.

**Table 1**  
Districts included in the scorecard initiative.

	2009	2010	2012	2014	2016
Agago			✓	✓	✓
Amuria	✓	✓	✓	✓	✓
Amuru	✓	✓	✓	✓	✓
Apac				✓	✓
Arua				✓	✓
Bududa		✓	✓	✓	✓
Buliisa		✓	✓	✓	✓
Gulu		✓	✓	✓	✓
Hoima	✓	✓	✓	✓	✓
Jinja		✓	✓	✓	✓
Kabale					✓
Kabarole			✓	✓	✓
Kaliro					✓
Kamuli	✓	✓	✓	✓	✓
Kanungu			✓	✓	✓
Kisoro					✓
Lira			✓	✓	✓
Luwero	✓	✓	✓	✓	✓
Lwengo					✓
Masindi				✓	✓
Mbale	✓	✓	✓	✓	✓
Mbarara			✓	✓	✓
Moroto	✓	✓	✓	✓	✓
Moyo		✓	✓	✓	✓
Mpigi		✓	✓	✓	✓
Mukono		✓	✓	✓	✓
Nakapiripiriti		✓	✓	✓	✓
Nebbi	✓	✓	✓	✓	✓
Ntungamo	✓	✓	✓	✓	✓
Nwoya				✓	✓
Rukungiri		✓	✓	✓	✓
Sheema					✓
Soroti		✓	✓	✓	✓
Tororo			✓	✓	✓
Wakiso			✓	✓	✓
Total	9	19	26	30	35

tive, the Afrobarometer dataset suffers from the limitation that not all variables are covered in each survey round and thus sample sizes per outcome indicator differ. Yet, the sample size of at least 9,000 observations allows us to identify relatively small changes in the perception of local democratic processes across indicators.

The outcome variables covered in most Afrobarometer survey rounds are: citizen assessment of the share of local councillors who are involved in corruption; citizen perception of the performance of local government and of the frequency that local councillors listen to what citizens have to say; citizen trust in their local government council and the Ugandan national parliament; and the perceived quality of road maintenance by the local government. The advantage of employing Afrobarometer data is that they have not been collected in the context of the scorecard intervention and are an independent resource of political perceptions of Ugandan citizens, and thus are free of confirmation or social desirability bias.

#### 4.2. Summary statistics

Table 2 presents descriptive statistics for the budgetary and service delivery outcome variables, as well as the control variables included in the analyses for the pre-intervention period, i.e. 2005 to 2008.<sup>13</sup> Statistically significant differences exist between treated units and control units for three of the seven outcome variables and five of the eight control variables, indicating that the impact

<sup>13</sup> The district-level data were collected by an independent researcher from the local districts. For the outcomes and control variables under study, the same measurements are applied across districts.

assessment needs to carefully control for confounding factors since the scorecards have not been randomly implemented.

For the pre-intervention sample we observe that the average district has a budget of 12.65 billion Ugandan Shillings (UGX), which corresponds to roughly €3.5 million (based on the exchange rate in March 2020). Average spending per citizen is UGX 36,98 (or €9). These figures highlight the financial limits of Ugandan local governments. The average share of budget returned to the central government – which is a constitutional requirement for unused budget – amounts to slightly below 4%. On average, districts run a delay of 4.1 months in sending their budgetary reports to the Ministry of Local Government, and local revenues contribute as little as UGX 0.53 billion (or around €130,000) to the budget. We include two indicators that are related directly to service delivery: the primary school leaving exam pass rate (60.98%) and the number of health centres and hospitals in the district per 100,000 inhabitants (2.4 on average).

Next, we turn to the control variables. With the exception of population growth, the poverty level and the tons of maize produced, differences between scorecard and non-scorecard districts are all significant. Data on population and population growth show that the average number of inhabitants per district was 362,016 and that population grows by 3.3% annually. On average, more than one quarter of the district population is considered poor (28.2%) and the district dependence on agriculture is covered by the average yearly production of maize (31,194 tons), sweet potatoes (25,873 tons), sorghum (4,918 tons) and millet (3,567 tons).

Information about the status of the district is another control variable, because some districts were created only in 2006 or 2009. Our district-level dataset contains information for all 111

**Table 2**  
Summary statistics for the pre-intervention period, the district-level administrative outcomes and control variables.

Outcome variables	Overall sample		Treated		Control		Difference in means (p-value)
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	
Overall district budget (in billions of UGX)	12.65	6.85	15.55	6.81	10.71	6.18	0.000***
Per capita district budget (UGX)	36,977	11,834	37,708	11,366	36,486	12,147	0.401
Share of the budget spent (%)	96.283	6.736	96.499	5.617	96.138	7.407	0.663
Delay in reporting to the central government (months)	4.120	1.067	4.126	0.854	4.115	1.191	0.933
Local contribution to revenues (billions of UGX)	0.525	0.818	0.888	1.110	0.281	0.376	0.000***
Primary school leaving exam pass rate (%)	60.983	5.228	61.941	4.446	60.339	5.616	0.012**
Health centres and hospitals per 100,000 inhabitants	2.368	2.636	2.307	2.496	2.409	2.733	0.754
<b>Control variables</b>							
Population (number)	362,016	193,047	426,884	200,952	318,377	175,047	0.000***
Population growth (%)	3.33	1.16	3.26	1.26	3.39	1.09	0.358
Poverty level (%)	28.19	21.93	29.73	22.59	27.15	21.49	0.338
Production of maize (tons)	31,194	45,527	30,305	32,896	31,793	52,429	0.791
Production of sweet potatoes (tons)	25,873	42,349	34,204	39,480	20,269	43,397	0.007***
Production of sorghum (tons)	4,918	9,037	7,380	12,365	3,262	5,246	0.000***
Production of millet (tons)	3,567	5,528	5,337	7,718	2,377	2,784	0.000*** <sup>o</sup>
District was split (proportion)	0.03	0.16	0.05	0.23	0.01	0.08	0.013**
<b>Treatment</b>							
Scorecard as of 2009 and later	0.402						

Note: The descriptive statistics are derived for the period prior to the implementation of the scorecard, i.e. for the years 2005 to 2008. The total number of observations is 276, the number of observations for the eventually to be treated districts is 111 and for the control districts there are 165 observations. \*\*\*/\*\*/\* indicates statistical significance at the 1/5/10% level, respectively. <sup>o</sup> The comparison of means in logs (as included in the empirical model) is statistically insignificant.

Ugandan districts since 2009 (except for the capital district of Kampala<sup>14</sup>); since some districts were created after 2006, we have an unbalanced panel. The indicator on district reorganization (with a value of 1 indicating the event of the split) shows that only around 3 % of the observations in our pre-intervention dataset result from the event of a district split. Of this pre-intervention sample, 40.2 % of the district observations are from districts that later participated in the scorecard intervention.

Table 3 presents descriptive statistics for Afrobarometer data on citizens’ perceptions of local councillors’ performance. Again, we present pre-intervention data. All but one outcome variables (Councillors are perceived as being corrupt) show significant differences between scorecard and non-scorecard districts. On average, citizens perceive some of their local councillors to be corrupt (score of 1.41 on a 0–3 scale), while they indicate that they have a fairly good impression of the councillors’ overall performance (score of 2.77 on a 1–4 scale). Still, respondents are critical, as indicated by their perception that local councillors listen only sometimes to their constituency (score of 1.39 on a 0–3 scale) and their low level of trust in the local government council (score of 1.42 on a 0–3). The trust in local government matches with the low level of trust in the Ugandan national parliament (score of 1.37 on a 0–3 scale). Finally, citizens perceive the quality of road maintenance by local government as moderately poor (score of 2.35 on a 1–4 scale). We use a number of control variables in the analysis of citizen perceptions of local councillors’ performance, relating to the respondents’ age, gender, home language, education level, religion, consumption needs, frequency of media usage and attendance of public meetings. For the sake of brevity, the summary statistics on the controls are presented in the Supplementary Materials (Table A1).

### 4.3. Identification strategy

Since we have to rely on observational data for the analysis, we implement a difference-in-difference model for the empirical anal-

ysis. Our model compares the change in outcomes in the scorecard districts (treatment group) before and after the scorecard initiative to the change in outcomes in the non-scorecard control group. By comparing *changes*, we control for observed and unobserved time-invariant characteristics that might be correlated with treatment status as well as the outcome.

We estimate the following equation for the service delivery outcomes at the district level:

$$Y_{dt} = X_{dt}\alpha + D_{dt}\beta + \mu_d + t_t + Dist_d trend + \varepsilon_{dt}, \tag{1}$$

where  $Y_{dt}$  is one of the outcome variables for district  $d$  at time  $t$  and  $X_{dt}$  collects district-level control variables (logged population size, logged population growth, poverty level, logged agricultural production of maize, sweet potatoes, sorghum and millet, and a dummy variable for the event of district reorganization to capture structural change). The treatment, i.e. the scorecard intervention, at the district level is denoted by  $D_{dt}$ . The district-specific fixed effect is captured by  $\mu_d$  while the time effect is denoted by  $t_t$ . We include eleven year dummies to control for possible annual trends. Finally, we control for district-specific time trends to allow for the possibility that different trends operate across districts ( $Dist_d trend$ ) and to accommodate the lack of balance across districts. Standard errors ( $\varepsilon_{dt}$ ) are clustered at the district level. The average treatment effect is captured by the coefficient  $\hat{\beta}$ .

We compare the outcomes of the difference-in-difference model with a simple comparison of means estimation to assess whether it is indeed necessary to control for the annual trends, district effects and district-specific trends. The validity of the findings is further addressed by testing for parallel trends and different placebo interventions prior to the actual scorecard intervention.

We apply a similar difference-in-difference model for the analysis of citizens’ perceptions. Here, the unit of observation is the individual citizen nested within a district. We measure changes in the perceived political atmosphere and in public attitudes. By comparing citizens from treated and control districts over time we can assess the impact of the scorecard initiative on political perceptions. In addition to the district-level control variables introduced above, we include the individual level control variables that were mentioned in section 4.2. Similarly, we test for the parallel trends and one placebo intervention since we have only two obser-

<sup>14</sup> We exclude Kampala for two reasons. First, Kampala is the capital district and we expect that the dynamics related to public services may be distinct from that in other districts. Second, Kampala was part of the scorecard initiative in 2009, but the intervention stopped in 2012.

**Table 3**  
Summary statistics, Afrobarometer citizen perceptions of local councillors' performance.

Outcome variables	Overall sample			Treated		Control		Difference in means (p-value)
	Obs.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	
Councillors are perceived as being corrupt (ordinal, 0–3, 'none' to 'all of them')	3,999	1.406	0.789	1.402	0.779	1.413	0.805	0.682
Perceived performance of local government councillors (ordinal, 1–4, 'strongly disapprove' to 'strongly approve')	4,277	2.769	0.808	2.792	0.808	2.734	0.807	0.020**
Councillors listen to what people have to say (ordinal, 0–3, 'never' to 'always')	4,305	1.392	0.931	1.454	0.934	1.296	0.920	0.000***
Trust in the local government council (ordinal, 0–3, 'not at all' to 'a lot')	4,322	1.421	0.568	1.441	0.555	1.389	0.585	0.004***
Trust in parliament (ordinal, 0–3, 'not at all' to 'a lot')	4,234	1.372	0.568	1.390	0.548	1.343	0.597	0.009***
Quality of maintenance of local roads by local government (ordinal, 1–4, 'very bad' to 'very well')	4,363	2.347	0.889	2.410	0.880	2.251	0.895	0.000***

Note: The descriptive statistics are derived for the period prior to the implementation of the scorecard, i.e. for the years 2005 and 2008. \*\*\*/\*\*/\* indicates statistical significance at the 1/5/10% level, respectively.

valuations available prior to the introduction of the scorecard. We use a linear regression model to analyse the Afrobarometer data despite their ordinal character. We opt for the linear model as it accommodates the different fixed effects and time trends most readily and coefficient estimates can be directly interpreted as marginal effects. In addition, we estimate an ordered probit model as a robustness check.

Further to assessing the impact of the scorecard intervention, we conduct an analysis controlling for the win margin<sup>15</sup> in the 2011 local elections to take account of local political dynamics, similar to the approach adopted by Grossman and Michelitch (2018: 289). We interact the 2011 win margin with all consecutive years to represent the political status quo after the elections.<sup>16</sup> In a second additional specification, we interact the win margin with the scorecard intervention to directly assess the sensitivity of the scorecard intervention to political competition.

## 5. Results

### 5.1. Analysis of local councils' budgetary policy and service delivery

Table 4 presents the regression results for the five budgetary and two service delivery outcomes. Panel A, which presents a simple comparison of means of scorecard versus non-scorecard districts, suggests that the scorecard intervention has had large and significant impacts. Yet, this naïve comparison is misleading since we do not have RCT data. The results of the difference-in-difference specification with the full set of control variables in Panel B indicate not only that the coefficient estimates decrease in magnitude (all but one), but also that two change sign and that only one remains statistically significant. These results demonstrate the need to control for confounding district and time factors in the empirical analysis to avoid that impacts are wrongly attributed to the scorecard initiative, as they result rather from omitted variables bias in the simple comparison of means.

Focusing on Panel B (Column 1), we see that scorecard districts spend on average 4.0 % less of their allocated budget than districts that are not part of the initiative. This finding is in stark contrast to the simple comparison in Panel A (Column 1) and results from a positive time trend, which reflects that districts receive larger bud-

gets year by year between 2009 and 2016 (captured by the positive coefficients associated with the time dummies, not reported). This means that studying the impact of the scorecard initiative on total public budget might result in a misleading outcome. Therefore, we also assess the impact on per capita public budget. In Panel B, Column 2 we find a very small positive and statistically insignificant effect.

Turning to the share of the budget spent (Column 3), we find that scorecard districts spend on average 3.8 % less of their budget than non-scorecard districts. In the Ugandan system, unused budget needs to be returned to the central government. This finding suggests that local government councils in scorecard districts appear to be more careful in their use of the local budget. The finding indicates that the scorecard's accountability aspect results in less waste of public money and possibly less spending for clientelistic purposes. The finding further points to the likely existence of an 'accountability-expenditure trade-off', which implies that increased public oversight, operationalized through the scorecard, leads to under-exhaustion of local budgets. It is not surprising that districts where the scorecard has not been implemented are not subject to the same restraint on expenditure of public funds. This suggests that top-down government accountability on its own is insufficient and the extra layer of accountability provided by the scorecard (even under the circumstances of limited civic involvement) provides an additional control mechanism. Note that this is already reflected in the simple comparison of means (Panel A), which shows that non-scorecard districts spend 2.6 % more of their budget than scorecard districts. The multivariate analysis further reinforces this finding, indicating that the simple comparison is likely an underestimation.

Delay in reporting to the central government (Column 4) is another performance indicator for local government councils. Although the simple comparison of means indicates that the use of the scorecard has a great impact on reporting delays, the difference-in-difference specification shows that there is no significant difference between scorecard and non-scorecard districts when it comes to their reporting to the Ministry of Local Government. Similarly, the introduction of the scorecard does not seem to impact significantly on the local collection of revenues (Column 5). Since raising local revenue requires adding (expensive) manpower within the local bureaucracy, it is not surprising that the scorecard initiative as an accountability instrument for local politicians did not have noticeable impact on this aspect of local public administration.

Finally, service delivery outcomes are analysed in Columns 6 and 7 of Table 4. While the simple comparison of means gives the impression that scorecard districts show better results on the

<sup>15</sup> The average win margin is calculated as the average margin of victory per district,  $[avgW - avgC]$ , where  $avgW$  is the average vote share of the incumbent politicians and  $avgC$  is the average vote share of main challengers (cf. Grossman and Michelitch 2018: 289). We used alternative indicators for political competition, such as the proportion of councillors belonging to the ruling National Resistance Movement (NRM) and the party affiliation (NRM or non-NRM) of the district chair.

<sup>16</sup> Information on earlier elections proved unavailable.



**Table 4**  
Analysis of local councils' budgetary policy and service delivery.

Dependent variable	[1] District budget (log)	[2] Per capita district budget (log)	[3] Share of the budget spent	[4] Delay in reporting to central government	[5] Local contribution to revenues (log)	[6] Primary school leaving exam pass rate	[7] Health centres and hospitals/ 100,000 inhabitant
<b>Panel A: Simple comparison of means</b>							
Scorecard intervention	0.421*** (0.000)	0.177*** (0.000)	-2.623*** (0.000)	-0.598*** (0.000)	2.146*** (0.000)	5.126*** (0.000)	1.045*** (0.000)
<b>Panel B: Main difference-in-difference results with the full set of controls</b>							
Scorecard intervention	-0.040 (0.361)	0.014 (0.653)	-3.810* (0.093)	0.041 (0.915)	1.084 (0.294)	0.018 (0.986)	0.236 (0.165)
<b>Panel C: Main specification with electoral results of 2011 elections</b>							
Scorecard intervention	-0.038 (0.373)	0.013 (0.665)	-3.719* (0.099)	0.052 (0.893)	1.086 (0.295)	-0.053 (0.958)	0.241 (0.168)
Avg. win margin in 2011 elections	0.206 (0.560)	-0.082 (0.753)	14.625 (0.243)	1.758 (0.219)	0.341 (0.940)	-12.068** (0.027)	0.903 (0.704)
<b>Panel D: Electoral results interacted with treatment</b>							
Scorecard intervention	-0.008 (0.857)	0.003 (0.938)	-1.448 (0.449)	-0.108 (0.762)	0.603 (0.624)	0.639 (0.614)	0.111 (0.551)
Avg. win margin in 2011 elections	0.228 (0.521)	-0.090 (0.733)	16.338 (0.185)	1.649 (0.253)	0.015 (0.997)	-11.583** (0.032)	0.815 (0.726)
Interaction (scorecard × average win margin)	-0.201 (0.344)	0.070 (0.650)	-15.244* (0.091)	1.076 (0.257)	3.242 (0.472)	-4.688 (0.265)	0.876 (0.317)
Observations	1,038	1,041	1,038	1,048	1,048	1,044	1,048

Notes: Panel A presents the simple comparison of means. The regressions presented in Panels B to D contain district fixed effects, time fixed effects and district-specific time trends that are jointly statistically significant within each group. In addition, the regressions control for the district poverty rate (log), the amount of millet produced (log), the amount of maize produced (log), the amount of sorghum produced (log), the amount of sweet potatoes produced (log), population size (log), population growth rate (log) and district reorganization as observable structural change. The panel dataset contains information for the years 2005 to 2016 and for 102 districts. All regressions use standard errors clustered at the district level; corresponding *p*-values are presented in parentheses. The number of observations presented at the bottom of each column is identical across the four estimation panels. \*\*\*/\*\*/\* indicates statistical significance at the 1/5/10% level, respectively.

**Table 5**  
Main results of the Afrobarometer data.

Dependent variable	[1] Councillors are perceived as being corrupt	[2] Performance of local government councillors	[3] Councillors listen to what people have to say	[4] Trust in the local government council	[5] Trust in the national parliament	[6] Maintenance of local roads by local government
<b>Panel A: Simple comparison of means</b>						
Scorecard intervention	-0.054*** (0.001)	-0.048*** (0.008)	-0.194*** (0.000)	-0.082*** (0.000)	-0.025** (0.036)	-0.067*** (0.003)
<b>Panel B: Main difference-in-difference results with the full set of controls</b>						
Scorecard intervention	-0.156** (0.046)	0.020 (0.859)	-0.060 (0.651)	0.019 (0.788)	0.087 (0.185)	0.001 (0.993)
<b>Panel C: Main specification with electoral results of 2011 elections</b>						
Scorecard intervention	-0.158** (0.043)	0.006 (0.953)	-0.074 (0.575)	0.011 (0.876)	0.081 (0.226)	-0.038 (0.805)
Average win margin in 2011 local elections	-0.254 (0.412)	-1.111*** (0.000)	-1.001 (0.219)	-0.654** (0.023)	-0.487* (0.076)	-2.772*** (0.020)
<b>Panel D: Electoral results interacted with treatment</b>						
Scorecard intervention	-0.207** (0.049)	-0.048 (0.709)	-0.264 (0.525)	0.022 (0.779)	0.115 (0.124)	0.512 (0.223)
Average win margin in 2011 local elections	-0.303 (0.319)	-1.163*** (0.000)	-1.265 (0.218)	-0.644** (0.024)	-0.455* (0.092)	-2.004 (0.159)
Interaction (scorecard x average win margin)	0.294 (0.292)	0.333 (0.296)	0.674 (0.618)	-0.064 (0.703)	-0.204 (0.209)	-1.955 (0.152)
Observations	12,055	12,485	9,146	12,589	12,524	9,282

Notes: Panel A presents the simple comparison of means. The regressions presented in Panels B to D contain district fixed effects, time fixed effects and district-specific time trends that are jointly statistically significant within each group. In addition, the regressions control for the district poverty rate (log), the amount of millet produced (log), the amount of maize produced (log), the amount of sorghum produced (log), the amount of sweet potatoes produced (log), population size (log), population growth rate (log) and district reorganization as observable structural change. In addition, the regressions contain individual characteristics related to the respondents' age, gender, home language, education level, religion, consumption needs, frequency of media usage and attendance of public meetings (see Supplementary Materials, Table A1). The dataset contains information for 2005, 2008, 2010, 2011, 2012, 2015, 2017. All regressions use standard errors clustered at the district level; corresponding *p*-values are presented in parentheses. The number of observations presented at the bottom of each column is identical across the four estimation panels. \*\*\*/\*\*/\* indicates statistical significance at the 1/5/10% level, respectively.

primary school leaving exam and have significantly more health facilities per 100,000 inhabitants, these findings disappear when we employ the full difference-in-difference specification. The result in Column 7 of Panel B indicates a possible relationship between the scorecard initiative and the number of health facilities, but the effect is only significant at the 17 % level. Yet, how reliable are the results? It is important to note not only that the time fixed effects are jointly significant across outcome variables, but this applies also to the district fixed effects and district-specific time trends. Thus, clearly there are considerable structural differences across districts paired with differential trends, which are accommodated in our empirical specification. Moreover, accounting for multiple hypothesis testing employing *q*-values (Benjamini and Hochberg, 1995; Anderson, 2008) we find no significant impact of the scorecard across outcomes.

In Panel C of Table 4 we added the average win margin per constituency in the district elections of 2011 as a proxy for district-level political competition in 2011 and later years, similar to the indicator used by Grossman and Mitchelitch (2018: 289). In their analysis of the impact of the scorecard on individual councillors, Grossman and Michelitch (2018: 291-294) find that the instrument's effect on performance is limited to so-called competitive constituencies, that is to electoral areas where there is significant competition among candidates.<sup>17</sup> Our analysis does not identify any relationship between political competition and budgetary outcomes. This holds for scorecard as well as non-scorecard districts. The one effect that stands out relates to service delivery: the primary school leaving exam pass rate is lower in districts with a higher average win margin. At the mean win margin of 0.168 this implies a reduction of the pass rate by 2 percentage points (-12.068\*0.168, *p*-value < 5 %). We can only hypothesize where this effect comes from. We conjecture that districts with a higher average win margin

are less competitive and have therefore a lower pressure on performance resulting in less favourable decisions regarding education and ultimately lower pass rates.

Finally, we analyse the relationship between political competition and the scorecard initiative, i.e. the interaction term. Results are presented in Panel D. Only one results stand out across outcome variables. There is some indication that *competitive scorecard districts* show under-exhaustion of their budgets – this result supports the findings of Grossman and Michelitch (2018). Moreover, the earlier identified result concerning competitive districts with higher primary school leaving exam pass rates is reinforced, although no interaction effect could be established. Overall, our findings suggest that political competition (and its interaction with the scorecard initiative) is not a key determinant of differences in district outcomes.

We also tested for other measures of political competition, namely whether the district chair is a member of the ruling national party, the National Resistance Movement (NRM), and the vote share of NRM councillors. Results are presented in the Supplementary Materials (Table A3). We do not find any indication that local council dynamics are related to national party politics beyond the already established findings.

## 5.2. Analysis of citizen perceptions

Results of the analysis of Afrobarometer data on citizens' perceptions of the work of local councillors are presented in Table 5.

Again, although the simple comparison of means across treatment and control districts produces some statistically significant results, these effects disappear when we apply the full difference-in-difference model. The application of the latter model points to one significant finding: local councillors in scorecard districts are perceived as being less corrupt (Panel B, Column 1). The effect of -0.156 is sizeable compared to the average pre-

<sup>17</sup> Note that their analysis only studies scorecard districts.

intervention rating on this outcome variable of 1.406 (Table 3). It corresponds to an 11 % decrease in the original rating. This particular finding is in line with the claim of the scorecard initiative that the instrument represents an additional layer of monitoring representatives. However, the evaluation of the performance of local government councillors (Panel B, Column 2) does not appear to be affected in the same way. Although the scorecard is positively associated with the performance evaluation, the effect is very small and statistically insignificant. Similarly, the scorecard initiative does not seem to enhance citizens' perceptions of councillor responsiveness (Panel B, Column 3), nor does it increase trust in the local government council (Panel B, Column 4) or in the national parliament (Panel B, Column 5). The latter is not surprising since we expect that this local scorecard initiative would only affect perceptions of the work of local councillors. The service delivery variable that is coherently available across all rounds of the Afrobarometer survey concerns the maintenance of local roads by local government. Again, we do not find a significant impact of the scorecard initiative (Panel B, Column 6).

We observe from the analysis that it is important to control not only for changes over time and across districts but also for district-specific trends. The analysis of the Afrobarometer data reinforces the findings in section 5.1 that there are considerable structural differences across districts, paired with differential trends. When applying multiple hypothesis testing to the Afrobarometer results, the positive impact of the scorecard on perceptions of corruption is supported further.

Panel C presents the findings that bring in electoral competition in districts: again, we control for the average win margin in the 2011 elections and their possible impact in later years. Competitiveness proves to be a stronger predictor for citizens' perceptions than the scorecard intervention, as it impacts on four of the six outcome variables. Yet, competitiveness in elections does not seem to impact on the perceived level of corruption of local councillors (Panel C, Column 1) or perceived responsiveness of councillors (Panel C, Column 3). In turn, in districts with a higher average win margin, the performance of the local government councillors is rated far lower (-1.111, Panel C, Column 2), and trust in the local council (-0.654, Panel C, Column 4) and the national parliament (-0.487, Panel C, Column 5) is significantly lower and road maintenance is judged much more critically (-2.772, Panel C, Column 6). The findings indicate that power concentration at the local level

results in more critical evaluations of local and nationally elected representatives and of service delivery by Ugandan citizens. Multiple hypothesis testing supports the findings.

The results of possible interaction effects between the scorecard initiative and electoral competitiveness in 2011 are presented in Panel D. Only the service delivery variable (the perceived quality of maintenance of local roads, Panel D, Column 6) suggests the possibility of an interaction effect. The finding on road maintenance suggests that the negative effect of political power concentration on the perceived quality of service delivery is reinforced in scorecard districts but the finding is only significant at the 16 % level. The findings reported for the other three outcome variables that showed to be significantly affected by political power concentration in Panel C persist with comparable magnitudes. These three outcomes remain unaffected by the scorecard intervention both directly and indirectly in form of the interaction with political power concentration.

Contrary to the analysis of the administrative data in section 5.1, we could not include information about all districts in the analysis of citizen perceptions with Afrobarometer survey data. This has to be taken into account when interpreting the findings. Yet, on the basis of available information and the related analyses reported in Table 5, Panel D we conclude that the scorecard intervention has impacted on the perceptions of corruption of local councillors, and potentially as well on the quality of service delivery in scorecard districts with relatively less political competition. Moreover, we find evidence that perceptions of performance and political trust are affected negatively by the relative lack of opposition at the district level.

Similar to the analysis of the administrative data we assessed whether national party politics, i.e., NRM membership of the district chair and the vote share obtained by the NRM, affect citizens' perceptions about local council dynamics. Results are presented in the Supplementary Materials (Table A4). The results corroborate the findings established for the average win margin.

### 5.3. Parallel trends and placebo interventions

Table 6 presents the results of the identifying parallel trends assumption of our empirical model. For this analysis, the above-mentioned outcomes in scorecard and non-scorecard districts are compared prior to the onset of the intervention in 2009.

**Table 6**  
Parallel trends and placebo interventions for the administrative data.

Dependent variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]
	District budget (log)	Per capita district budget (log)	Share of the budget spent	Delay in reporting to central government	Local contribution to revenues (log)	Primary school leaving exam pass rate	Health centres and hospitals/ 100,000 inhabitant
<b>Panel A: Assessing the parallel trends between treatment and control districts prior to the intervention</b>							
Parallel trend	1.230 (0.647)	1.369 (0.436)	0.149 (0.999)	-4.708 (0.805)	76.807 (0.477)	-57.280** (0.023)	-108.013 (0.414)
<b>Panel B: Placebo treatment taking place in 2008</b>							
Placebo treatment	0.036 (0.734)	0.083 (0.305)	-4.135 (0.256)	0.062 (0.885)	1.075 (0.777)	0.851 (0.589)	-0.600 (0.518)
<b>Panel C: Placebo treatment taking place in 2007 and 2008</b>							
Placebo treatment	-0.032 (0.756)	-0.039 (0.580)	-2.128 (0.686)	-0.047 (0.926)	-1.292 (0.662)	-0.576 (0.699)	0.143 (0.927)
<b>Panel D: Placebo treatment taking place in the period 2006-2008</b>							
Placebo treatment	-0.032 (0.771)	0.564 (0.579)	3.808 (0.457)	0.103 (8.190)	-1.081 (0.756)	-0.857 (0.290)	-0.033 (0.980)
Observations	276	276	276	280	280	280	280

Notes: Only pre-intervention periods, i.e. the years 2005 to 2008, are considered. All regressions contain district fixed effects, time fixed effects and district-specific time trends. In addition, the regressions control for the district poverty rate (log), the amount of millet produced (log), the amount of maize produced (log), the amount of sorghum produced (log), the amount of sweet potatoes produced (log), population size (log), population growth rate (log) and district reorganization as one observable structural change. All regressions use standard errors clustered at the district level; corresponding p-values are presented in parentheses. The number of observations presented is identical across the four panels. \*\*\*/\*\*/\* indicates statistical significance at the 1/5/10% level, respectively.

**Table 7**  
Parallel trends and placebo intervention for the Afrobarometer data.

Dependent variable	[1]	[2]	[3]	[4]	[5]	[6]
	Councillors are perceived as being corrupt	Performance of local government councillors	Councillors listen to what people have to say	Trust in the local government council	Trust in the national parliament	Maintenance of local roads by local government
<b>Panel A: Assessing the parallel trends between treatment and control districts prior to the intervention</b>						
Parallel trend	-7.677 (0.135)	-4.615 (0.691)	-13.241* (0.087)	-1.283 (0.742)	3.968 (0.348)	-10.127 (0.317)
<b>Panel B: Placebo treatment taking place in 2008</b>						
Placebo treatment	-0.008 (0.935)	-0.039 (0.655)	-0.079 (0.506)	0.010 (0.787)	0.125** (0.031)	0.021 (0.851)
Observations	3,875	4,135	4,165	4,180	4,096	4,219

Notes: Only pre-intervention periods, i.e., the years 2005 and 2008, are considered. All regressions contain district fixed effects and time fixed effects along with the district and individual control variables as specified in the notes of Table 5. All regressions use standard errors clustered at the district level; corresponding *p*-values are presented in parentheses. The number of observations presented is identical across the two panels. \*\*\*/\*\*/\* indicates statistical significance at the 1/5/10% level, respectively.

Panel A of Table 6 does not show any difference in the coefficient estimates of the five budgetary outcome variables for the years 2005 to 2008, which reassures us that the districts were following similar patterns in budget allocation. However, for the primary school leaving exam pass rate the parallel trend could not be established (Column 6): scorecard districts had a significantly lower primary school leaving exam pass rate in the period 2005 to 2008 compared to non-scorecard districts. Since we do not identify any impact of the scorecard intervention in the main specification, we are not concerned about the lack of a parallel trend in the primary school leaving exam pass rate. In turn, for the hospitals and health centres the parallel trend holds (Column 7).

Next, we estimate a model where we wrongly assume that the scorecard initiative started in 2008 (Table 6, Panel B). This assumption appears to have no impact on any of the outcome variables, which suggests that the impact we reported above can credibly be attributed to the scorecard initiative. Panel C and D use 2007 and 2006 as further placebo interventions; these do not lead to changes to our earlier results. In sum, the analyses of parallel trends and placebo treatments reported in Table 6 indicate that the assumptions underlying the identification strategy hold and that our difference-in-difference analysis is valid.

Similar analyses are conducted for the Afrobarometer data (Table 7), for which only two pre-treatment observations (2005 and 2008) are available.

Panel A of Table 7 shows that parallel trends can be established for all except one outcome variable. This implies that our finding with regard to local councillors' responsiveness (Panel A, Column 3) needs to be treated with caution: prior to the scorecard intervention, councillors in the future scorecard districts were rated significantly more negatively than councillors in future non-scorecard districts.

Importantly, the parallel trends analysis highlights the necessity to include district-specific time trends along with time and district fixed effects to arrive at credible results, since district-specific trends seem to vary. Yet, the limitations of the Afrobarometer dataset also need to be taken into consideration. Because Afrobarometer does not include all districts in every survey round, our pre-treatment analysis contains only 55 out of over 100 districts, and as a consequence our parallel trends estimations are limited. This limitation is also reflected in the reported coefficient estimates, which tend to be large in absolute terms suggesting that the data do not fully support the analysis that was performed. By contrast, the analysis of the full sample includes 96 districts.

Panel B contains the results of the placebo intervention for 2008, with 2005 as baseline. The weakness of the pre-treatment dataset, which contains data for only 55 districts, carries over into this analysis. The analysis of the placebo intervention also reports one significant impact, namely for trust in the national parliament (Panel B, Column 5). Yet, compared to the parallel trends analysis,

coefficient estimates are smaller in absolute terms. The absence of an impact on perceived corruption gives us further confidence that our main result is credible within the limitations imposed by the data. Although it proves technically possible to carry out parallel trends and placebo analyses, these have to be interpreted with caution given the discussed limitations of the Afrobarometer data.

As a final robustness check, we conduct an ordered probit analysis for the Afrobarometer data to account for the ordinal nature of the data. Results are presented in the Supplementary Materials (Table A2). Our earlier findings are fully corroborated by this analysis. The ordered probit analysis only identifies an impact of the scorecard intervention on the perceived level of corruption. Consistent with the main results, the marginal effects show that, on average, the scorecard intervention makes it 3.3 percentage points more likely that no councillor is perceived as corrupt and 4.6 percentage points more likely that only some of them are seen to be corrupt. The scorecard reduces the likelihood that most of the local councillors, or all of them, are perceived as corrupt by 4.3 and 3.6 percentage points, respectively. With all coefficients having a *p*-value below 5 %, the impact of the scorecard intervention is reinforced.

## 6. Conclusions and discussion

This article set out to analyse the impact of a long-term accountability instrument, the Ugandan Local Government Council's Score Card Initiative, on budgetary outcomes, service delivery and citizen perception of representatives' performance at the district level. The longevity of the instrument, which was initiated in 2009, made the scorecard initiative a good candidate to analyse the effect of efforts to enhance representatives' accountability. The fact that the scorecard has been rolled out in approximately 30 % of all Ugandan districts in a staggered way enabled us to compare the use of the tool in a quasi-experimental design. Finally, the inclusion of service delivery improvements as one of the outcomes in the initiative's theory of change meant that the scorecard was a proper focal point for an analysis of the effects of information provision about politicians' performance beyond the narrow political domain where the representatives operate. Thus, we are able to contribute new empirical evidence about the so-called 'long route of accountability', in particular regarding the impact of information provision on public service delivery.

The first part of our analyses focused on administrative data related to budgetary policies and two service delivery variables, collected for scorecard and non-scorecard districts over more than a decade (2005–2016). The second part of the study analysed Afrobarometer survey data containing citizen perceptions of the performance of representatives at the local level over the 2005 to 2017 period. The core of the analysis was the application of a set

of difference-in-difference models to assess the effects of the scorecard initiative both over time and across scorecard and non-scorecard districts.

The analyses allowed us to causally attribute the effects deriving from the scorecard initiative, but the findings indicate that the impact has been scattered. Our analyses show that the scorecard initiative has had some small impact on policy making of district councils. We find that there is a difference in the under-exhaustion of budgets between scorecard and non-scorecard districts – this seems to point to greater budgetary restraint of local government councils in scorecard districts. Also, we find some indication that districts with more competitive elections in district constituencies perform better on one service-delivery indicator, i.e. the primary school leaving exam pass rate.

Further, our analysis of Afrobarometer survey data suggests that the implementation of the scorecard has an impact on citizens' perceptions of the performance of local councillors. Our analyses report the robust finding that citizens of scorecard districts feel on average that local councillors are less corrupt than citizens of non-scorecard districts. This result may be seen as indication of the trust-enhancing effect of the scorecard. On the basis of this, we may hypothesize that the overall climate of interaction between citizens and their representatives will improve over time as a result of the introduction of the scorecard. We expect that it will take time before the changes in perception actually trickle down to improve the interactions between citizens and local councillors, and that the effect on the quality of service delivery may be seen only over a longer period if civic monitoring is maintained.

Afrobarometer data also point to some potential effects on service delivery: the quality of road maintenance in scorecard districts is judged better than that in non-scorecard districts, although this finding seems limited to districts with a higher level of competition for constituency seats. Although the analyses indicate that there is a potential impact of political competition on perceived performance and responsiveness of representatives, doubts remain on the robustness of these findings due to the nature of Afrobarometer data.

Overall, our analyses have some important implications for the literature on 'transparency for accountability' interventions (Kosack and Fung, 2014: 68). In the first place, we provide further information on the difficulties involved in shortening the so-called 'long route' of accountability (Waddington et al., 2019: 6). Even though the scorecard initiative is a good example of a 'strategic intervention' (Fox, 2015: 352), particularly given its resources, longevity, annual recurrence and related information campaigns, the impact of providing performance information about politicians and local political institutions on service delivery seems to be rather limited. This finding reinforces Waddington et al.'s (2019: 78) conclusion that 'ultimately, [the shortened long route of accountability] remains too long to identify short-term effects on service delivery'. Next, our analyses suggest that the introduction of the scorecard may have had some positive results on the behaviour of local politicians. In particular, our findings on the budgetary restraint observed by councils in scorecard districts and on citizen perceptions of corruption among politicians in those districts may be the source of some optimism about the scorecard's potential. Finally, the finding that political competition seems to have some positive effect on service delivery related to educational performance and road maintenance may be judged positively.

While it is easy to highlight some of the positive aspects of the research reported in this study, it is also important to be cautious about the value of the scorecard as an accountability mechanism. The relationship between accountability mechanisms and political and policy outcomes appears to be very complex, and dependent on a range of variables. For this reason, more research is needed about this complex relationship, as well as about the transmission mechanisms that are at work to produce enhanced responsiveness

of representative bodies to their constituencies. In this regard, our finding that some effects operate only when there is true competition among representatives points at the importance of classical mechanisms of democracy. The interaction between competitiveness and accountability should, in our view, get a primary focus of future investigations into political representation.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We thank Rose Namara for her support in collecting and updating the data and Fauzul Muna for the research assistance provided. We have benefited from the comments given by participants in the ECPR workshop "Citizen Perspectives on Accountability in Developing Democracies" in May 2021, notably Adriana Algarin, Leticia Barbabela, Sergiu Gherghina, Marcia Grimes, Miquel Pellicer and Eva Wegner. We gratefully acknowledge the helpful comments provided by two anonymous reviewers of *World Development*. The original research for this paper was supported by the Policy and Operations Evaluation Department (IOB) of the Dutch Ministry of Foreign Affairs. The original results of the research were reported in Hout et al. (2016).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.worlddev.2022.106011>.

## References

- ACODE (2021). Publications/local governments. <https://acode-u.org/local-government>.
- Adida, C., Gottlieb, J., Kramon, E., & McClendon, G. (2017). Reducing or reinforcing in-group preferences? an experiment on information and ethnic voting. *Quarterly Journal of Political Science*, 12(4), 437–477.
- Adida, C., Gottlieb, J., Kramon, E., & McClendon, G. (2019). Under what conditions does performance information influence voting behavior? Lessons from Benin. In T. Dunning, G. Grossman, M. Humphreys, S. D. Hyde, C. McIntosh, & G. Nellis (Eds.), *Information, accountability, and cumulative learning: lessons from Metaketa I* (pp. 81–117). Cambridge: Cambridge University Press.
- Adida, C., Gottlieb, J., Kramon, E., & McClendon, G. (2020). When does information influence voters? The joint importance of salience and coordination. *Comparative Political Studies*, 53(6), 851–891.
- Afrobarometer (2020). Uganda data. <http://afrobarometer.org/data/338>.
- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry preschool, and early training projects. *Journal of the American Statistical Association*, 103(484), 1481–1495.
- Arias, E., Balán, P., Larreguy, H., Marshall, J., & Querubín, P. (2019). Information provision, voter coordination, and electoral accountability: Evidence from Mexican social networks. *American Political Science Review*, 113(2), 475–498.
- Arias, E., Larreguy, H., Marshall, J., & Querubín, P. (2019). When does information increase electoral accountability? Lessons from a field experiment in Mexico. In T. Dunning, G. Grossman, M. Humphreys, S. D. Hyde, C. McIntosh, & G. Nellis (Eds.), *Information, accountability, and cumulative learning: lessons from Metaketa I* (pp. 118–155). Cambridge: Cambridge University Press.
- Ashworth, S. (2012). Electoral accountability: Recent theoretical and empirical work. *Annual Review of Political Science*, 15(1), 183–201.
- Askim, J. (2007). How do politicians use performance information? An analysis of the Norwegian local government experience. *International Review of Administrative Sciences*, 73(3), 453–472.
- Bainomugisha, A., Mbabazi, J., Muhwezi, W. W., Bogere, G., Atukunda, P., Ssemakula, E. G., ... Ayesigwa, R. (2020). *The Local Government Councils Scorecard FY 2018/19: The next big steps, consolidating gains of decentralisation and repositioning the local government sector in Uganda. ACODE Policy Research Paper 96*. Kampala: ACODE.
- Banerjee, A.V., Kumar, S., Pande, R. & Su, F. (2011). Do informed voters make better choices? Experimental evidence from urban India. [http://scholar.harvard.edu/files/rpande/files/doinformedvoters\\_nov11\\_0.pdf](http://scholar.harvard.edu/files/rpande/files/doinformedvoters_nov11_0.pdf).

- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, 57(1), 289–300.
- Besley, T., & Burgess, R. (2002). The political economy of government responsiveness: Theory and evidence from India. *Quarterly Journal of Economics*, 117(4), 1415–1451.
- Bidwell, K., Casey, K., & Glennerster, R. (2020). Debates: Voting and expenditure responses to political communication. *Journal of Political Economy*, 128(8), 2880–2924.
- Boas, T. C., Hidalgo, F. D., & Melo, M. A. (2019a). Horizontal but not vertical: Accountability institutions and electoral sanctioning in Northeast Brazil. In T. Dunning, G. Grossman, M. Humphreys, S. D. Hyde, C. McIntosh, & G. Nellis (Eds.), *Information, accountability, and cumulative learning: Lessons from Metaketa I* (pp. 257–286). Cambridge: Cambridge University Press.
- Boas, T. C., Hidalgo, F. D., & Melo, M. A. (2019b). Norms versus action: Why voters fail to sanction malfeasance in Brazil. *American Journal of Political Science*, 63(2), 385–400.
- Börang, F., & Grimes, M. (2021). Social accountability and quality of government: Effectiveness and recursive effects. In A. Bågenholm, M. Bauhr, M. Grimes, & B. Rothstein (Eds.), *The Oxford handbook of the quality of government* (pp. 268–290). Oxford: Oxford University Press.
- Bovens, M. (2005). Public accountability. In E. Ferlie, L. E. Lynn, & C. Pollitt (Eds.), *The Oxford handbook of public management* (pp. 182–208). Oxford: Oxford University Press.
- Brixi, H. (2010). New development: Innovative governance – Using the citizen scorecard in China. *Public Money & Management*, 30(4), 257–260.
- Buntaine, M.T., Jablonski, R., Nielson, D.L., & Pickering, P.M. (2018). SMS texts on corruption help Ugandan voters hold elected councillors accountable at the polls. *Proceedings of the National Academy of Sciences of the United States of America*, 115(26), 6668–6673.
- Buntaine, M. T., Bush, S. S., Jablonski, R. S., Nielson, D. L., & Pickering, P. M. (2019). Budgets, SMS texts, and votes in Uganda. In T. Dunning, G. Grossman, M. Humphreys, S. D. Hyde, C. McIntosh, & G. Nellis (Eds.), *Information, accountability, and cumulative learning: lessons from Metaketa I* (pp. 188–220). Cambridge: Cambridge University Press.
- Chong, A., De La A.L., O., Karlan, D., & Wantchekon, L. (2015). Does corruption information inspire the fight or quash the hope? A field experiment in Mexico on voter turnout, choice, and party identification. *The Journal of Politics*, 77(1), 55–71.
- Deiniger, K., & Mpuga, P. (2005). Does greater accountability improve the quality of public service delivery? Evidence from Uganda. *World Development*, 33(1), 171–191.
- Devas, N., & Grant, U. (2003). Local government decision-making – Citizen participation and local accountability: Some evidence from Kenya and Uganda. *Public Administration and Development*, 23(4), 307–316.
- Dunning, T., Grossman, G., Humphreys, M., Hyde, S. D., McIntosh, C., & Nellis, G. (2019). Meta-analysis. In T. Dunning, G. Grossman, M. Humphreys, S. D. Hyde, C. McIntosh, & G. Nellis (Eds.), *Information, accountability, and cumulative learning: lessons from Metaketa I* (pp. 315–374). Cambridge: Cambridge University Press.
- Epstein, P., Wray, L., & Harding, C. (2006). Citizens as partners in performance management. *Public Management*, 88(10), 18–22.
- Ferraz, C., & Finan, F. (2008). Exposing corrupt politicians: The effects of Brazil's publicly released audits on electoral outcomes. *Quarterly Journal of Economics*, 123(2), 703–745.
- Fox, J. A. (2015). Social accountability: What does the evidence really tell? *World Development*, 72, 346–361.
- Gainer, M. (2015). *Listening to the public: A citizen scorecard in the Philippines, 2010–2014*. Innovations for Successful Societies Program, Princeton: Princeton University. [https://successfultsocieties.princeton.edu/sites/successfultsocieties/files/Philippines%20Listening%20to%20the%20Public\\_ToU\\_1.pdf](https://successfultsocieties.princeton.edu/sites/successfultsocieties/files/Philippines%20Listening%20to%20the%20Public_ToU_1.pdf).
- Galiani, S., Gertler, P., & Schargrodsy, E. (2005). Water for life: The impact of the privatization of water services on child mortality. *Journal of Political Economy*, 113(1), 83–120.
- Gottlieb, J. (2016). Greater expectations: A field experiment to improve accountability in Mali. *American Journal of Political Science*, 60(1), 143–157.
- Grossman, G., & Hanlon, W. W. (2014). Do better monitoring institutions increase leadership quality in community organizations? Evidence from Uganda. *American Journal of Political Science*, 58(3), 669–686.
- Grossman, G., & Michelitch, K. (2018). Information dissemination, competitive pressure, and politician performance between elections: A field experiment in Uganda. *American Political Science Review*, 112(2), 280–301.
- Harding, R. (2015). Attribution and accountability: Voting for roads in Ghana. *World Politics*, 67(4), 656–689.
- Hout, W., Brouwers, R., Fisher, J., Namara, R., Schakel, L., & Wagner, N. (2016). *Policy review good governance: Uganda country study. Report Commissioned by the Policy and Operations Evaluation Department (IOB), Ministry of Foreign Affairs* Available at <http://hdl.handle.net/1765/102964>. The Hague:: International Institute of Social Studies.
- Humphreys, M. & Weinstein, J.M. (2012). Policing politicians: Citizen empowerment and political accountability in Uganda – Preliminary analysis. Working Paper S-5021-UGA-1, London: International Growth Centre.
- James, O. (2011). Performance measures and democracy: Information effects on citizens in field and laboratory experiments. *Journal of Public Administration Research and Theory*, 21(3), 399–418.
- Joshi, A. (2013). Do they work? Assessing the impact of transparency and accountability initiatives in service delivery. *Development Policy Review*, 31(S1), S29–S48.
- Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard – Measures that drive performance. *Harvard Business Review*, 70(1), 71–79.
- Kosack, S., & Fung, A. (2014). Does transparency improve governance? *Annual Review of Political Science*, 17(1), 65–87.
- Muyomba-Tamale, L., & Cunningham, K. (2017). Holding governments accountable for service delivery: The Local Government Councils Scorecard Initiative in Uganda. *Commonwealth Journal of Local Governance*, 20, 187–205.
- Pande, R. (2011). Can informed voters enforce better governance? *Experiments in low-income democracies, Annual Review of Economics*, 3(1), 215–237.
- Platas, M., & Raffler, P. (2019). Candidate videos and vote choice in Ugandan parliamentary elections. In T. Dunning, G. Grossman, M. Humphreys, S. D. Hyde, C. McIntosh, & G. Nellis (Eds.), *Information, accountability, and cumulative learning: lessons from Metaketa I* (pp. 156–187). Cambridge: Cambridge University Press.
- Raffler, P. (2019). *Does political oversight of the bureaucracy increase accountability? Field experimental evidence from an electoral autocracy*: Harvard University, mimeo. [http://piaraffler.com/wp-content/uploads/2019/06/Raffler\\_Political-Oversight\\_June2019.pdf](http://piaraffler.com/wp-content/uploads/2019/06/Raffler_Political-Oversight_June2019.pdf).
- Ravindra, A. (2004). *An assessment of the impact of Bangalore citizen report cards on the performance of public agencies*. ECD Working Paper 12. Washington, D.C.: World Bank Operations Evaluation Department.
- Thampi, G. (2011). Informed public opinion and official behavior change. In S. Odugbemi & T. Lee (Eds.), *Accountability through public opinion: From inertia to public action* (pp. 359–376). Washington, D.C.: World Bank.
- Thinnyane, H., & Siebörger, I. (2017). MobiSAM: Reflections from a four year case study using technology to increase public participation in local government in South Africa. *The Journal of Community Informatics*, 13(1), 25–49.
- Tumushabe, G., Mushemeza, E.D., Muyomba-Tamale, L., Lukwago, D. & Ssemakula, E. (2010). *Monitoring and assessing the performance of local government councils in Uganda: Background, methodology and score card*. ACODE Policy Research Series 31, Kampala: ACODE. <https://www.acode-u.org/uploadedFiles/PRS31.pdf>.
- Tumushabe, G., Muyomba, L.T., & Ssemakula, E. (2011). *Uganda Local Government Councils Score-card report 2009/2010: Political accountability, representation and the state of service delivery*. ACODE Policy Research Series 42, Kampala: ACODE. <https://www.acode-u.org/uploadedFiles/PRS42.pdf>.
- Tumushabe, G., Muyomba-Tamale, L., Ssemakula, E. G., & Muhumuza, T. (2013). *Uganda Local Government Councils Scorecard 2012/13: The big service delivery divide*. ACODE Policy Research Series 60, Kampala: ACODE.
- Van Reybrouck, D. (2016). *Against elections: The case for democracy*. London: The Bodley Head.
- Waddington, H., Sonnenfeld, A., Finetti, J., Gaarder, M., John, D., & Stevenson, J. (2019). Citizen engagement in public services in low- and middle-income countries: A mixed-methods systematic review of participation, inclusion, transparency and accountability (PITA) initiatives. *Campbell Systematic Reviews*, 15(1–2), e1025.
- Wantchekon, L., & Vermeersch, C. (2011). Information, social networks, and the demand for public goods: Experimental evidence from Benin. In S. Odugbemi & T. Lee (Eds.), *Accountability through public opinion: From inertia to public action* (pp. 123–135). Washington, D.C.: World Bank.
- Wild, L., & Harris, D. (2011). *The political economy of community scorecards in Malawi*. London: Overseas Development Institute.
- World Bank (2003). *World Development Report 2004: Making services work for poor people*. Washington, D.C.: World Bank.
- Wray, L., & Hauer, J. (1997). Performance measurement to achieve quality of life: Adding value through citizens. *Public Management*, 79(8), 4.