

Does Accountability Enhance Service Delivery? Assessment of a Local Scorecard Initiative in Uganda¹

Paper for presentation at the workshop ‘Citizen Perspectives on Accountability in Developing Democracies’, ECPR Joint Sessions of Workshops, May 2021

Wil Hout^{a,b}, Natascha Wagner^a and Binyam A. Demena^a

Abstract

This article assesses whether the Local Government Council’s Score Card Initiative, implemented in Uganda since 2009, achieved its intended impact of increasing accountability in public spending and service delivery. We analyse a district-level panel dataset (2005-2016) with administrative data, as well as Afrobarometer data on citizen perceptions (2008-2017). Empirically, we exploit the phasing in of the scorecard for a meticulous difference-in-difference framework with district-specific trends. The results show limited measurable impacts of the scorecards on policy making and public services. Scorecard districts appear to spend less of their budgets in comparison with non-scorecard districts. Impacts on service delivery cannot be detected. Yet, the scorecard impacts on citizen perceptions of local councillors’ corruption. Our results provide a quantitative contribution to the literature on accountability by demonstrating that civil society reporting mechanisms about political representatives only trickle down slowly to improved services. Like earlier research, we find that impacts also depend on political competitiveness.

Keywords: Uganda; accountability; democracy; representation; governance; local government; service delivery; scorecard; difference-in-difference

^a International Institute of Social Studies, Erasmus University Rotterdam, The Netherlands.

^b Corresponding author: Wil Hout, International Institute of Social Studies, Erasmus University Rotterdam, Kortenaerkade 12, 2518AX The Hague, The Netherlands. Email: hout@iss.nl.

¹ The original research for this paper was done with support from the Policy and Operations Evaluation Department (IOB) of the Dutch Ministry of Foreign Affairs (see Hout et al. 2016). We thank Rose Namara for her support in collecting and updating the data and Fauzul Muna for the research assistance provided.

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 element in representative democracy. Such a political regime consists of a chain of principal-agent relations, in which political influence is shifted upwards, from citizens to members of representative institutions 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 improved 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 importantly 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 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 impacts on *political behaviour* of voters/citizens or representatives. In this paper, we present the findings of a study on how 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 accountability mechanisms. First, most studies on the input side focus on the effects of enhanced information about political performance on the perceptions of individual voters or representatives, but leave the outcome of improved information, for instance on service provision, undiscussed. Second, despite the wealth of experimental studies, there is still major controversy over whether information provision has 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).

In this article, we present findings of a longitudinal analysis of the impact of the Local Government Council's Score Card Initiative (LGCSCI, hereafter referred to as the 'scorecard initiative'), implemented since 2009 with external financial support by Ugandan NGO Advocates Coalition for Development and Environment (ACODE). The goal of the scorecard initiative is to deepen democratic governance by means of yearly, evidence-based performance assessments of local government councillors. 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 five rounds 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. Second, 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 their political representatives has rather small impacts on service delivery, mainly by affecting budgetary discipline. We also show that citizen perceptions are more favourable towards councillors in scorecard districts, as councillors are perceived as being less corrupt. In line with the earlier results of Grossman and Michelitch (2018) we find that impacts depend on political competitiveness at the district level.

This 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. 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

A literature search on the impact of different forms of information provision and performance assessment (including scorecards) on public accountability identified 35 studies, which differ significantly in terms of research design, methods and units of analysis (see Appendix 1). Most of the reviewed studies attempt to determine whether the introduction of accountability mechanisms and forms of (enhanced) information provision impact 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 is orientated to the *effect* that accountability mechanisms have on the performance of political functionaries or institutions in terms of service delivery. With regard to research methods, about half of the reviewed studies use field experiments to determine the impact of accountability mechanisms; 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 accountability mechanisms and of the provision of information about the performance of representatives. A smaller number of studies are either undecided about the size or nature of the impact or find that accountability mechanisms and performance information have no impact at all.

The majority of field experiments report a positive accountability 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. 2019; 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, 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 new accountability mechanisms 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 purely national public goods electoral platforms, while clientelistic platforms 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 accountability mechanisms impact positively on citizen and voter positions (Ravindra 2004; Gainer 2015; Harding 2015; Thinyane and Siebörger 2017), some studies fail to find an influence of information provision (Brixi 2010; Ashworth 2012).

Finally, a group of studies that focus on service delivery, policy outputs and incumbent behaviour similarly lead to 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. Various studies conclude that the introduction of new accountability mechanisms, including variants of scorecards, have positive effects for 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 had a positive effect on service delivery, and one-third showed no impact. In their experimental study of the Ugandan local government scorecard, Grossman and Michelitch (2018) find that random dissemination of the results to constituents improved the performance of politicians but did not have an impact on service delivery. Likewise, Wild and Harris' (2011) report that the introduction of a 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 their focus on decision making is limited. While Kosack and Fung (2014) synthesise studies on the performance of a range of service providers across a variety of countries, their conclusions do not necessarily relate to decision making in political institutions. Grossman and Michelitch's (2018) study of Uganda mainly focuses on 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 generalisable conclusions about decision making in local councils.

The literature review leads to two conclusions. First, the overview demonstrates that the majority of studies has focused on the effect of information provision and accountability instruments on citizen 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 citizens and their representatives have received relatively less attention. Secondly, the limited harvest of studies of policy outputs and service delivery shows that previous research has left a gap in our knowledge of the impact of accountability tools on the output side. Our research on the Ugandan local government scorecard is intended to start filling this gap by linking the implementation of the scorecard to budgetary outcomes and service delivery, next to focusing on citizens' perceptions of local councillor performance.

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 for improved democratic governance in Uganda, established by Denmark, Ireland, The Netherlands, Norway, Sweden and the UK. The scorecard initiative is implemented by ACODE, a Kampala-based NGO focused on public policy research and advocacy, in collaboration with the Uganda Local Governments Association (ULGA).

The scorecard initiative is a 10-year programme to strengthen citizens' demand for effective public service delivery and accountability. More specifically, its 'central premise is 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). The scorecard initiative contains three key elements. First, it provides annual assessments of district councils' and councillors' performance based on a scorecard.² Secondly, the initiative leads to the publication of feedback reports on the assessments, both at the level of included districts and across Uganda. Finally, the initiative 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). Increased capacity of the local councillors is expected to result in better oversight of local service providers and thus in better service delivery.

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).

Attempts to improve monitoring instruments for service delivery are directed to the 'demand side', which implies that citizens and their organisations should be '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). In the scorecard's logic, existing mechanisms for horizontal accountability are supplemented by vertical accountability instruments, which should lead to more influence of citizens, both through the exercise of their voting rights, and through the activities of citizens' organisations.

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. 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 of political leaders

² The actual scorecard and its content can be found in Tumushabe et al. (2010). It is divided in four sections: (i) legislative roles, (ii) accountability, (iii) planning and budgeting, and (iv) service delivery along seven sub-categories. Simple scoring questions are used that, in the ideal case, add up to a score of 100.

of their roles, as well as more awareness among citizens about the responsibilities of those political leaders (Tumushabe et al. 2013: 68). This should, in turn, strengthen the accountability of local councillors and positively impact the quality of service delivery and the use of development budgets at the local level.

According to its initiators, the positive impact of the introduction of the scorecard may 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 failed multi-party politics (Tumushabe et al. 2013: 68-70). Factors concerning the embedding of local governments include the distortions inherent to Uganda's decentralisation policy, related to the use of decentralisation for clientelistic purposes, as well as 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 show 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).

The theory of change underlying the scorecard initiative holds that the introduction of the scorecard instrument has a positive impact on public service delivery through five interconnected channels.³ 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 will lead to increased demand for better services. Fourth, this would result in enhanced capacity of local councillors, which includes increased awareness of their formal roles and responsibilities, and ultimately in improved performance. Fifth, the scorecard initiative would result in the greater ability of key stakeholders, such as civil society, to lobby for better services at the sub-national level.

Implementation of the annual scorecard process starts with the filling of the scorecard by ACODE researchers with the use of interviews with local councillors and collected written evidence. First, the scorecard assesses how elected political leaders (district councillors, chairpersons, speakers) and the district council as institution perform in view of their tasks and responsibilities specified in the Constitution, the Local Governments Act and other legal provisions. Next, data are verified in field visits to service-delivery units and interviews with service consumers. Finally, ACODE researchers organise focus-group discussions with community groups on councillors' performance in all sub-counties of districts involved in the scorecard (Tumushabe et al. 2010: 27-33).

³ 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).

The scorecard was 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 the number was increased until the current number of 35 districts was reached, comprising almost 30% of all districts.⁴ 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 decentralisation has been accounted for as this has led to the division 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.

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		✓	✓	✓	✓

⁴ 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. The number of districts has increased again since the completion of our study.

Nakapiripiriti		✓	✓	✓	✓
Nebbi	✓	✓	✓	✓	✓
Ntungamo	✓	✓	✓	✓	✓
Nwoya				✓	✓
Rukungiri		✓	✓	✓	✓
Sheema					✓
Soroti		✓	✓	✓	✓
Tororo			✓	✓	✓
Wakiso			✓	✓	✓
Total	9	19	26	30	35

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 accountability and service delivery. We study three sets of service delivery outcomes: (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.⁵ As outlined in the theory of change, measured by the scorecards and based on earlier analyses within scorecard districts, accountability and budgeting are assessed and impacts on education and health are expected (Tumushabe et al., 2010, 2013).

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 representative, the Afrobarometer dataset suffers from the limitation that not all variables are covered in each round and thus sample sizes per outcome 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. The outcome variables covered in most 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

⁵ 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.

Table 2 presents descriptive statistics for the budgetary and service delivery outcome variables, as well as the control variables included in the analyses. Statistically significant differences exist between treated units and control units (including pre-treatment scorecard and non-scorecard districts). While at first sight one may be inclined to conclude that the scorecard intervention has had an impact, the reported differences, in particular with respect to control variables, rather suggest that a rigorous empirical design needs to be employed that controls for confounding factors since the scorecards have not been randomly implemented and, along the dimension of control variables, we only find balance in two out of five variables.

For the sample as a whole (across districts and over time) we observe that the average district has a budget of 16.450 billion Ugandan Shillings (UGX), which corresponds to roughly €4 million (based on the exchange rate in March 2020). Average spending per citizen is UGX 53,078 (or €13). 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 over 4%. On average, districts run a delay of 3.4 months in sending their budgetary reports to the Ministry of Local Government, and local revenues contribute as little as UGX 0.45 billion (or around €110,000) to the budget. We include two indicators that are related directly to service delivery: the primary school leaving exam pass rate (62.19%) and the number of health centres in the district (11.22 on average).

Next, we turn to the control variables. With the exception of population growth and district reorganisation, 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 326,669 and that population grows by 3.07% annually. Access to clean drinking water, which is a proxy for the local poverty level in the absence of complete poverty data, is available for around 66% of district inhabitants. 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 Ugandan districts since 2009 (except for the capital district of Kampala⁶); since some districts were created after 2006, we have an unbalanced panel. The indicator on district reorganisation (with a value of 1 indicating the event of the split) shows that only around 1.4% of the observations in our dataset result from the event of a district split. The final control variable is the average win margin in the 2011 local elections⁷ (information on earlier elections proved unavailable); the data indicate that the average distance

⁶ 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.

⁷ 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 average majority, the proportion of unopposed constituencies, the proportion of councillors belonging to the ruling National Resistance Movement (NRM) and the party affiliation (NRM or non-NRM) of the district chair. These alternative indicators did not have explanatory power in the analyses reported in section 5 of this article.

between the directly elected councillors and the respective runners-up was 16.8% on average. Finally, we present the share of treated districts: this reflects the number of observations in the dataset that are subject to the scorecard intervention. Overall, 17.6% of all district-year observations have been subject to the scorecard intervention.

Table 3 presents descriptive statistics for Afrobarometer data on citizens' perceptions on local councillors' performance. All but one outcome variables (trust in the national parliament) show significant differences between scorecard and non-scorecard districts. On average, citizens perceive some of their local councillors to be corrupt (score of 1.38 on a 0-3 scale), while they indicate that they have a fairly good impression of the councillors' overall performance (score of 2.70 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 1.19 on a 1-3 scale) and their low level of trust in the local government council (score of 1.34 on a 0-3). The trust in local government matches with the low level of trust in the Ugandan national parliament (score of 1.33 on a 0-3 scale). Finally, citizens perceive the quality of road maintenance by local government as moderately poor (score of 2.27 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 Table A1 in the Appendix.

Table 2: Summary statistics, 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)	16.40	9.70	23.20	13.00	15.00	8.14	0.000
Per capita district budget (UGX)	53,078.35	24,456.72	59,847.39	22,879.27	51,621.23	24,552.39	0.000
Share of the budget spent (%)	95.982		93.687		96.478		0.000
Delay in reporting to the central government (months)	3.441	1.699	3.000	1.987	3.536	1.616	0.000
Local contribution to revenues (billions of UGX)	0.450	0.719	0.631	0.759	0.412	0.704	0.000
Primary school leaving exam pass rate (%)	62.190		66.172		61.346		0.000
Health centres and hospitals (number)	11.224	9.160	17.661	11.203	9.846	8.022	0.000
Control variables							
Population (number)	326,669	213,206	405,504	296,141	309,794	186,781	0.000
Population growth (%)	3.070		3.045		3.076		0.774
Access to clean drinking water (%)	66.088		73.868		64.422		0.000
District was split (proportion)	0.014		0.011		0.015		0.660
Average win margin in 2011 local elections (proportion)	0.168	0.160	0.224	0.127	0.156	0.163	0.000
Treatment							
Scorecard intervention	0.176						

Note: The sample size for the outcome variables is identical to the maximum number of available observations per outcome (see Table 4). The summary statistics of the control variables and the yearly observations are based on the largest number of observations available (1,072). The total number of observations for units under treatment is 189, except if there are missing data. The total number of observations in the control group is 883, except if there are missing data.

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

Outcome variables	Overall sample		Treated			Control			Difference in means (p-value)
	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	
Councillors are perceived as being corrupt (ordinal, 0-3, 'none' to 'all of them')	1.377	0.778	2,809	1.335	0.765	9,277	1.390	0.782	0.001
Perceived performance of local government councillors (ordinal, 1-4, 'strongly disapprove' to 'strongly approve')	2.699	0.854	2,828	2.665	0.877	9,695	2.709	0.847	0.018
Councillors listen to what people have to say (ordinal, 0-3, 'never' to 'always')	1.190	0.928	2,025	1.051	0.911	7,169	1.229	0.929	0.000
Trust in the local government council (ordinal, 0-3, 'not at all' to 'a lot')	1.338	0.575	2,861	1.278	0.58	9,761	1.356	0.573	0.000
Trust in parliament (ordinal, 0-3, 'not at all' to 'a lot')	1.330	0.574	2,880	1.312	0.564	9,682	1.335	0.577	0.059
Quality of maintenance of local roads by local government (ordinal, 1-4, 'very bad' to 'very well')	2.270	0.911	2,057	2.222	0.911	7,275	2.284	0.911	0.006

Note: The sample size for the outcome variables is identical to the maximum number of available observations per outcome (see Table 5).

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 analysis. 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:

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

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, access to clean drinking water and a dummy variable for the event of district reorganisation to capture structural change). In addition, we conduct an analysis controlling for the win margin 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. The treatment, i.e. the scorecard intervention, at the district level is denoted by D_{dt} . The district-specific fixed effect is captured by μ_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 (ε_{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 the 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 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 observations 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; coefficient estimates

can be directly interpreted as marginal effects. In addition, we estimate an ordered probit model as a robustness check.

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. The results of the difference-in-difference specification with the full set of control variables in Panel B indicate not only that all coefficient estimates decrease in magnitude, 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 positive effects are wrongly attributed to the scorecard initiative, which results 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 4.7% less 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 budgets year by year between 2009 and 2016 (reflected in positive coefficients associated with the time dummies, not reported). Studying the impact of the scorecard initiative on total public budget might be 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. This may indicate 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, operationalised 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 is already reflected in the simple comparison of means (Panel A), which shows that non-scorecard districts spend 2.8% more of their budget than scorecard districts. The multivariate analysis further reinforces this finding, indicating that the simple comparison is likely to even be an underestimate.

Delays 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.

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 primary school leaving exam and have significantly more health facilities, this finding disappears 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 16% 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 scorecards 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. 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. Our analysis does not identify any relationship between political competition and budgetary policy outcomes. There is some indication that districts with a higher average win margin delay reporting to the central level. At the maximum win margin of 0.641, reporting is delayed by 1.3 months, but this finding is statistically significant only at the 15% level. 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 (Table 2) this implies a reduction of the pass rate by almost 2 percentage points (p -value < 5%).

Finally, we analyse the relationship between political competition and the scorecard initiative. Only two results stand out across outcome variables. There is some indication that competitive scorecard districts show under-exhaustion of their budgets. While the effect seems practically relevant, it is statistically significant only at the 14% level. Moreover, we find that scorecard districts with higher average win margins tend to delay their reporting more (p -value = 0.062). Overall, our findings suggest that political competition (and its interaction with the scorecard initiative) is not a key determinant of differences in district outcomes. Once we account for multiple hypothesis testing all impacts disappear.

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

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 govern- ment	[5] Local contribution to revenues (log)	[6] Primary school leaving exam pass rate	[7] Number of health centres and hospitals
Panel A: Simple comparison of means							
Scorecard intervention	0.432*** (0.000)	0.167*** (0.001)	-2.791*** (0.006)	-0.536*** (0.000)	2.036*** (0.000)	4.826*** (0.002)	7.815*** (0.000)
Panel B: Main difference-in-difference results with the full set of controls							
Scorecard intervention	-0.047 (0.284)	0.013 (0.674)	-3.853* (0.093)	-0.001 (0.998)	1.197 (0.225)	-0.283 (0.770)	1.743 (0.157)
Panel C: Main specification with electoral results of 2011 elections							
Scorecard intervention	-0.046 (0.295)	0.013 (0.680)	-3.763* (0.098)	0.012 (0.977)	1.192 (0.228)	-0.343 (0.719)	1.748 (0.159)
Avg. win margin in 2011 elections	0.246 (0.487)	-0.044 (0.867)	14.671 (0.261)	2.093 (0.144)	-0.879 (0.838)	-10.976** (0.040)	0.764 (0.896)
Panel D: Electoral results interacted with treatment							
Scorecard intervention	-0.023 (0.637)	0.003 (0.925)	-1.619 (0.417)	-0.297 (0.438)	0.905 (0.460)	0.099 (0.940)	1.905 (0.166)
Avg. win margin in 2011 elections	0.260 (0.465)	-0.050 (0.851)	16.032 (0.210)	1.916 (0.180)	-1.043 (0.809)	-10.711** (0.042)	0.854 (0.885)
Interaction (scorecard x average win margin)	-0.136 (0.528)	0.056 (0.719)	-13.143 (0.132)	1.897* (0.062)	1.761 (0.702)	-2.733 (0.532)	-0.968 (0.806)
Observations	1,064	1,067	1,064	1,072	1,072	1,069	1,072

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 access to clean drinking water as a proxy of district level poverty, population size (log), population growth rate (log) and district reorganisation as observable structural change. The panel dataset contains information for the years 2005 to 2016 and for 111 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.

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 almost -0.20 is sizeable compared to the average score on this outcome variable of 1.38 that was reported in Table 3. This particular finding is in line with the claim of the scorecard initiative that the instrument represents an additional layer of monitoring of 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' perception 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 the perceptions about the work of the 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 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 districts with a higher average win margin, the performance of the local government councillors is rated far lower (Panel C, Column 2), and trust in the local council (Panel C, Column 4) and the national parliament (Panel C, Column 5) is significantly lower and road maintenance is judged much more poorly (Panel C, Column 6). The findings indicate that power concentration at the local level results in more critical evaluation of local and national 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 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. 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, we conclude that the scorecard intervention has impacted on the perceptions of corruption among local councillors, as well as 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 Table A4 in the Appendix. Our earlier findings about these variables are repeated here, as we find little or no impact of national political dynamics on outcomes at the district level.

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
Panel A: Simple comparison of means						
Scorecard intervention	-0.055 (0.240)	-0.043 (0.384)	-0.178*** (0.001)	-0.077** (0.011)	-0.023 (0.351)	-0.063 (0.301)
Panel B: Main difference-in-difference results with the full set of controls						
Scorecard intervention	-0.196** (0.016)	0.008 (0.936)	-0.004 (0.979)	0.032 (0.654)	0.078 (0.237)	-0.021 (0.888)
Panel C: Main specification with electoral results of 2011 elections						
Scorecard intervention	-0.200** (0.014)	-0.002 (0.983)	-0.026 (0.856)	0.026 (0.722)	0.072 (0.280)	-0.099 (0.520)
Average win margin in 2011 local elections	-0.389 (0.170)	-1.005*** (0.000)	-0.927 (0.241)	-0.597** (0.038)	-0.510* (0.067)	-3.185** (0.015)
Panel D: Electoral results interacted with treatment						
Scorecard intervention	-0.231** (0.036)	-0.045 (0.724)	-0.197 (0.670)	0.040 (0.588)	0.112 (0.128)	0.593 (0.167)
Average win margin in 2011 local elections	-0.423 (0.138)	-1.050*** (0.000)	-1.141 (0.238)	-0.582** (0.040)	-0.469* (0.087)	-2.314 (0.121)
Interaction (scorecard x average win margin)	0.177 (0.559)	0.249 (0.461)	0.618 (0.674)	-0.085 (0.603)	-0.227 (0.140)	-2.504* (0.064)
Observations	12,086	12,523	9,194	12,622	12,562	9,332

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 with each group. In addition, the regressions control for district characteristics (population size, population growth, access to clean water, district reorganisation) and individual characteristics related to respondents' age, gender, home language, education level, religion, consumption needs, frequency of media usage and attendance of public meetings (see Appendix, 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.

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 abovementioned outcomes in scorecard and non-scorecard districts are compared prior to treatment. Panel A of Table 6 does not show any difference in coefficients for five outcome variables for the years 2005 to 2008, which reassures us that the districts were following similar patterns in budget allocation and service delivery and that the parallel trends assumption holds. Only for the local contribution to revenues (Column 5) the parallel trend cannot be established: scorecard districts already collected higher local revenues in the period 2005 to 2008 compared to non-scorecard districts. Since we do not identify any impact of the intervention in the main specification we are not concerned about the lack of a parallel trend in local revenues.

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.

Table 6: Parallel trends and placebo interventions for the administrative data

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] Number of health centres and hospitals
Panel A: Assessing the parallel trends between treatment and control districts prior to the intervention							
Parallel trend	0.235 (0.500)	0.291 (0.157)	1.773 (0.922)	0.313 (0.838)	57.891*** (0.000)	2.930 (0.697)	15.031 (0.422)
Panel B: Placebo treatment taking place in 2008							
Placebo treatment	0.020 (0.856)	0.069 (0.398)	-4.599 (0.194)	0.117 (0.796)	1.982 (0.591)	0.742 (0.636)	0.432 (0.890)
Panel C: Placebo treatment taking place in 2007 and 2008							
Placebo treatment	-0.018 (0.859)	-0.021 (0.757)	-1.669 (0.722)	-0.120 (0.811)	-2.268 (0.462)	-0.423 (0.779)	-4.324 (0.417)
Panel D: Placebo treatment taking place in the period 2006-2008							
Placebo treatment	-0.031 (0.773)	-0.052 (0.523)	3.873 (0.424)	0.020 (0.964)	-1.817 (0.595)	-0.655 (0.395)	1.859 (0.631)
Observations	262	262	262	264	264	264	264

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 access to clean drinking water as a proxy of district-level poverty, population size (log), population growth rate (log) and district reorganisation 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.

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 only be established for four of the six outcome variables. This implies that our findings with regard to local councillors' performance (Panel A, Column 2) and responsiveness (Panel A, Column 3) need 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.

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 52 out of over 100 districts, and as a consequence our parallel trends estimations are limited. This limitation is also reflected in the coefficient estimates: the reported coefficients for councillor performance and responsiveness (Panel A, Columns 2 and 3) are extremely large in absolute terms, which suggests that the data do not fully support the analysis that was performed. In contrast, the analysis of the full sample includes 103 districts, and for most districts the Afrobarometer survey contains more than one observation. Despite the limitation of the parallel trends analysis for the Afrobarometer data, we find it reassuring that the outcome 'councillor corruption' (Panel A, Column 1) reflects equal trends across future scorecard and non-scorecard districts and that the reported coefficient estimate is reasonable. Thus, available evidence supports our main finding about perceived lower levels of corruption among local councillors in scorecard 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 52 districts, carries over into this analysis. Similar to Panel A, the analysis of the placebo intervention also reports an impact for councillor performance and responsiveness (Panel B, Columns 2 and 3). Compared to the parallel trends analysis, coefficient estimates are smaller in absolute terms and show a change in sign. We also find an impact of the placebo treatment on road maintenance (Panel B, Column 6). 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 limitations deriving from Afrobarometer's sampling.

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 Table A2 in the Appendix. Our earlier findings are fully corroborated by this analysis. The ordered probit analysis only brings out 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 4.19 percentage points more likely that no councillor is perceived as corrupt and 5.77 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 5.36 and 4.60 percentage points, respectively. With all coefficients having a p -value below 2%, the impact of the scorecard intervention is reinforced.

Table 7: Parallel trends and placebo intervention for 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
Panel A: Assessing the parallel trends between treatment and control districts prior to the intervention						
Parallel trend	-1.134 (0.619)	-65.092*** (0.000)	-23.347*** (0.000)	-8.851 (0.274)	3.984 (0.735)	-0.188 (0.971)
Panel B: Placebo treatment taking place in 2008						
Placebo treatment	0.830 (0.966)	1.147*** (0.000)	0.703** (0.013)	-0.188 (0.317)	-35.197 (0.135)	0.389** (0.042)
Observations	3,771	4,028	4,057	4,072	3,989	4,111

Notes: Only pre-intervention periods, i.e. the years 2005 and 2008, are considered. All regressions contain district fixed effects, time fixed effects and district-specific time trends 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.

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 implementation, 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 accountability and transparency instruments beyond the narrow political domain where the representatives operate.

The first part of our analyses focused on administrative data on 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 2008-2017 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 attribute the effects deriving from the scorecard initiative, but the findings indicate that the impact has been scattered. Our analyses indicate that the scorecard initiative has had

limited 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 scorecard districts with more competitive elections in district constituencies perform better on one service-delivery indicator, 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 hypothesise 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.

Afrobarometer data also point to some effect on service delivery: the quality of road maintenance in scorecard districts is judged significantly better than in non-scorecard districts, although this finding seems limited again 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 further advance the literature on political accountability and transparency. Importantly, the sustained implementation of instruments to provide citizens with more information about their political representatives may have a positive impact on relevant political and policy outcomes. In particular, policy decisions of representative bodies may be affected, and certain forms of service delivery may be improved. Obviously, more research is needed about the complex relationship between accountability mechanisms and political and policy outcomes, and 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.

Bibliography

- Adida, C., Gottlieb, J., Kramon, E. and 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. and 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 and G. Nellis (eds) *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*, Cambridge: Cambridge University Press, pp. 81-117.
- Adida, C., Gottlieb, J., Kramon, E. and 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', retrieved 14 June 2020 from <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. and 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. and 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 and G. Nellis (eds) *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*, Cambridge: Cambridge University Press, pp. 118-155.
- 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.
- Banerjee, A.V., Kumar, S., Pande, R. and Su, F. (2011) 'Do Informed Voters make Better Choices? Experimental Evidence from Urban India', retrieved 14 June 2020 from http://scholar.harvard.edu/files/rpande/files/doinformedvoters_nov11_0.pdf.
- Benjamini, Y. and 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. and 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. and Glennerster, R. (2019) 'Debates: Voting and Expenditure Responses to Political Communication', *Journal of Political Economy*, <https://doi.org/10.1086/706862>.
- Boas, T.C., Hidalgo, F.D. and 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 and G. Nellis (eds) *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*, Cambridge: Cambridge University Press, pp. 257-286.
- Boas, T.C., Hidalgo, F.D. and Melo, M.A. (2019b) 'Norms Versus Action: Why Voters Fail to Sanction Malfeasance in Brazil', *American Journal of Political Science* 63(2): 385-400.
- Bovens, M. (2005), 'Public accountability', in E. Ferlie, L.E. Lynn and C. Pollitt (eds), *The Oxford Handbook of Public Management*, Oxford: Oxford University Press, pp. 182-208.
- 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. and 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. and Pickering, P.M. (2019) 'Budgets, SMS Texts, and Votes in Uganda', in T. Dunning, G. Grossman, M. Humphreys, S.D. Hyde, C. McIntosh and G. Nellis (eds) *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*, Cambridge: Cambridge University Press, pp. 188-220.
- Chong, A., De La O, A.L., Karlan, D. and 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. and Mpuga, P. (2005) 'Does Greater Accountability Improve the Quality of Public Service Delivery? Evidence from Uganda', *World Development* 33(1): 171-191.
- Devas, N. and 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. and Nellis, G. (2019) 'Meta-Analysis', in T. Dunning, G. Grossman, M. Humphreys, S.D. Hyde, C. McIntosh and G. Nellis (eds) *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*, Cambridge: Cambridge University Press, pp. 315-374.
- Epstein, P., Wray, L. and Harding, C. (2006) 'Citizens as Partners in Performance Management', *Public Management* 88(10): 18-22.
- Ferraz, C. and 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.

- Gainer, M. (2015) 'Listening to the Public: A Citizen Scorecard in the Philippines, 2010-2014', Innovations for Successful Societies Program, Princeton: Princeton University, retrieved 14 June 2020 from https://successfulsocieties.princeton.edu/sites/successfulsocieties/files/Philippines%20Listening%20to%20the%20Public_ToU_1.pdf.
- Gottlieb, J. (2016) 'Greater Expectations: A Field Experiment to Improve Accountability in Mali', *American Journal of Political Science* 60(1): 143-157.
- Grossman, G. and 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. and 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. and Wagner, N. (2016), 'Policy Review Good Governance: Uganda Country Study', Report Commissioned by the Policy and Operations Evaluation Department (IOB), Ministry of Foreign Affairs, The Hague: International Institute of Social Studies. Available at <http://hdl.handle.net/1765/102964>.
- Humphreys, M. and 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.
- Kaplan, R.S. and Norton, D.P. (1992) 'The Balanced Scorecard – Measures that Drive Performance', *Harvard Business Review* 70(1): 71-79.
- Kosack, S. and Fung, A. (2014) 'Does Transparency Improve Governance?', *Annual Review of Political Science* 17(1): 65-87.
- Lierl, M. and Holmlund, M. (2019) 'Performance Information and Voting Behavior in Burkina Faso's Municipal Elections: Separating the Effects of Information Content and Information Delivery', in T. Dunning, G. Grossman, M. Humphreys, S.D. Hyde, C. McIntosh and G. Nellis (eds) *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*, Cambridge: Cambridge University Press, pp. 221-256.
- Muyomba-Tamale, L. and 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. and Raffler, P. (2019) 'Candidate Videos and Vote Choice in Ugandan Parliamentary Elections', in T. Dunning, G. Grossman, M. Humphreys, S.D. Hyde, C. McIntosh and G. Nellis (eds) *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*, Cambridge: Cambridge University Press, pp. 156-187.
- Raffler, P. (2019) 'Does Political Oversight of the Bureaucracy Increase Accountability? Field Experimental Evidence from an Electoral Autocracy', Harvard University, mimeo, retrieved 14 June 2020 from 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 and T. Lee (eds) *Accountability through Public Opinion: From Inertia to Public Action*, Washington, D.C.: World Bank, pp. 359-376.
- Thinnyane, H. and 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. and 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.
- Tumushabe, G., Muyomba-Tamale, L., Ssemakula, E.G. and 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.
- VNG International (2014) *Evaluation of the ACODE Scorecard for Local Government*, Kampala: VNG International.
- Wantchekon, L. and Vermeersch, C. (2011) 'Information, Social Networks, and the Demand for Public Goods: Experimental Evidence from Benin', in S. Odugbemi and T. Lee (eds) *Accountability through Public Opinion: From Inertia to Public Action*, Washington, D.C.: World Bank, pp. 123-135.
- Wild, L. and Harris, D. (2011) 'The Political Economy of Community Scorecards in Malawi', London: Overseas Development Institute.
- Wray, L. and Hauer, J. (1997) 'Performance Measurement to Achieve Quality of Life: Adding Value through Citizens', *Public Management* 79(8): 4.

APPENDICES (TO BE SHARED ONLINE)

Appendix 1: Overview of research on public accountability and scorecards

Study	Method	Regional focus	Unit of analysis	Dependent variable	Results	
Adida et al. (2017)	Field experiment	Benin (137 villages)	Voters	Electoral behaviour	+	Information about the legislative performance of incumbent politicians running in the election impacts on electoral behaviour; ethnicity is intervening variable.
Adida et al. (2019)*	Field experiment	Benin (30 communes)	Voters	Electoral behaviour	0	Information about the legislative performance of incumbent politicians has no impact on electoral behaviour.
Adida et al. (2020)	Field experiment	Benin (30 communes)	Voters	Electoral behaviour	+	Information about the legislative performance of incumbent politicians impacts on electoral behaviour; impact is mediated by salience of the information and attitudes of others in their community.
Arias, Balan et al. (2019)	Field experiment	Mexico (678 electoral precincts)	Voters	Electoral behaviour	+	Information about the performance of incumbents impacts on electoral behaviour; electoral sanctioning of poor performers is enhanced by coordination of voters within networks.
Arias, Larreguy et al. (2019)*	Field experiment	Mexico (678 electoral precincts)	Voters	Electoral behaviour	0	Information about the performance of incumbents doesn't impact in the expected direction; information about good performance and malfeasance both benefits the incumbents.
Ashworth (2012)	Literature review	Not applicable	Voters and incumbent politicians	Electoral behaviour, incumbent behaviour	0	Information about incumbent performance is less important for voters than expectations of payoffs in the future; incumbents are driven more by desire to impress voters with competence and ideology than by serving voters' interests.
Askim (2007)	Survey	Norway (750 councillors)	Members of local government councils	Councillor behaviour	+	Performance information impacts on behaviour of local councillors at different stages of decision-making process (agenda setting, decision stage and implementation); impact is greatest on councillors involved in elderly care, administrative affairs and education.
Banerjee et al. (2011)	Field experiment	India, Delhi (775 polling stations in slums)	Voters	Electoral behaviour, voter turnout	+	Information about incumbent performance and qualifications impacts on electoral behaviour, voter turnout and vote buying practices.
Besley and Burgess (2002)	Panel data	India (16 states)	State governments	Public goods provision, service delivery	+	Newspaper circulation and electoral accountability (electoral competition and turnout) lead state governments to become more responsive to falls in food production and crop flood damage by paying attention to food distribution and calamity relief expenditure.
Bidwell et al. (2019)	Field experiment	Sierra Leone (224 polling centres)	Voters, elected MPs	Electoral behaviour, MP accountability	+	The spread of videotaped debates between candidates impacts on electoral behaviour; participation in debates impacts on accountability of elected MPs.
Boas et al. (2019a)*	Field experiment	Brazil, Pernambuco State (75 municipalities)	Voters	Electoral behaviour	0	Information about the performance of incumbent politicians has no impact on electoral behaviour.
Boas et al. (2019b)	Field experiment and vignettes	Brazil, Pernambuco State (47 municipalities)	Voters	Electoral behaviour	+/0	Presentation of vignettes about malfeasance of public officials impacts on electoral behaviour; information about malfeasance of the incumbent mayor does not impact on electoral behaviour.

Study	Method	Regional focus	Unit of analysis	Dependent variable	Results
Brixi (2010)	Survey	China (5 cities)	Citizens	Citizen perceptions	0 Information about the quality of service delivery in citizen scorecards creates insufficient incentives for citizens to make informed choices and hold local government agencies accountable.
Buntaine et al. (2018)	Field experiment	Uganda (27 districts)	Voters	Electoral behaviour	+/- Information about the performance of incumbent politicians impacts on electoral behaviour; information about the performance of district chair does not impact on electoral behaviour.
Buntaine et al. (2019)*	Field experiment	Uganda (762 villages)	Voters	Electoral behaviour	+/- Information about the performance of incumbent politicians impacts on electoral behaviour; information about the performance of district chair does not impact on electoral behaviour.
Chong et al. (2015)	Field experiment	Mexico (12 municipalities)	Voters	Electoral behaviour, voter turnout	+ Information about the performance of incumbent politicians impacts on electoral behaviour and voter turnout
Devas and Grant (2003)	Literature review	Kenya and Uganda (local governments)	Local decision-making, citizen participation	Local government accountability to citizens	0 Committed local leadership, external pressure and increased availability of information are important but not sufficient conditions for local government accountability to citizens.
Dunning et al. (2019)	Meta-analysis of 5 field experiments	3 African, 2 Latin American countries	Voters	Electoral behaviour	0 Information about the performance of incumbent politicians do not shape voters' evaluations of candidates, nor electoral behaviour.
Ferraz and Finan (2008)	Experiment	Brazil (373 municipalities)	Mayoral elections	Support for incumbent mayor	+ Information about the performance of incumbent mayors impacts on the likelihood of re-election; effects are more pronounced in municipalities with local radio stations.
Gainer (2015)	Policy analysis	The Philippines	Citizen scorecard	Citizens' perceptions of government performance	+ Introduction of a citizen scorecard led to a decrease of failing and acceptable grades given to government offices and to an increase of good and excellent grades.
Gottlieb (2016)	Field experiment	Mali (95 rural communes)	Voters	Electoral behaviour	+ Participation in a civics course impacts on electoral behaviour and on the likelihood to sanction poor performing local government representatives.
Grossman and Hanlon (2014)	Survey	Uganda (287 village-level groups)	Community leaders	Payoffs for community leaders	+ Monitoring impacts on accountability of community leaders by increasing the penalties for those who shirk in office; there is a trade-off with quality as some candidates do not choose leadership if they have more rewarding options.
Grossman and Michelitch (2018)	Field experiment	Uganda (20 districts)	Local council members	Council member performance, policy outputs	+/- Dissemination of scorecard impacts on council members' performance, but only in competitive constituencies; the scorecard does not impact on service delivery overall but influences the number of development projects and development spending in competitive districts.
Harding (2015)	Election and administrative data	Ghana (438 electoral areas)	Elections	Incumbent vote share	+ Information about the quality of service delivery impacts on electoral behaviour.
Humphreys and Weinstein (2012)	Field experiment	Uganda (225 constituencies and 332 MPs)	Parliamentary scorecard	MP behaviour, election results	0 Introduction of a scorecard on the performance of Members of Parliament does not impact on the behaviour of those MPs, nor does it impact on their prospects for re-election.

Study	Method	Regional focus	Unit of analysis	Dependent variable	Results
James (2011)	Field and laboratory experiments	UK, Exeter (service users and students)	Citizens	Citizens' perceptions, electoral behaviour	0/+ Information about the performance public service providers does not impact on electoral behaviour; information impacts on citizens' perceived performance and satisfaction
Kosack and Fung (2014)	Literature review (16 experiments)	Asia, Africa and US	Transparency initiatives for service providers	Citizens' perceptions of quality of service delivery	+/0 Two-thirds of reviewed studies report positive effects of transparency initiatives on service delivery; one-third report unsuccessful outcomes.
Pande (2011)	Literature review (13 experiments)	Not applicable	Voters	Electoral behaviour	+ The reviewed studies report that information about the performance of politicians impacts on electoral behaviour; voters in low-income settings are receptive to information about politician performance.
Platas and Raffler (2019)	Field experiment	Uganda (11 parliamentary constituencies)	Voters	Electoral behaviour, voter turnout	+ Screening of videos of candidate MPs answering policy questions impacts on electoral behaviour and voter turnout in the case of alignment between voters' and candidates' policy preferences; there is no impact when policy preferences are not aligned.
Raffler (2019)	Field experiment and interviews	Uganda (260 local governments)	Local council members	Service delivery	+ Dissemination of financial information and training modules impacts on monitoring of the bureaucracy by local politicians and on service delivery; the impact is limited to local governments where the chairperson does not belong to the ruling party.
Ravindra (2004)	Interviews	Bangalore, India (agencies, government, NGOs, media)	Citizen report card	Citizens' perceptions	+ The Citizen Report Card impacts on public awareness of quality of services and increased demands for service improvement; public service agencies become more client oriented.
Thampi (2011)	Case studies	Delhi, India (people's audit), Kenya, 3 cities (citizen report cards)	Information provision	Service delivery	+ Information provision about user experiences with public services to the general public impacts on public service delivery and facilitates internal reforms within utilities.
Thinyane and Siebörger (2017)	Survey and interviews	Makana municipality, South Africa	Accountability initiative	Communication between citizens and municipality	+ Social accountability monitoring through mobile technology impacts on communication between citizens and government; pressure from citizens does not force municipalities to listen.
Wantchekon and Vermeersch (2011)	Field experiment	Benin (9 non-competitive electoral districts)	Voters	Electoral behaviour	- Exposure to purely national public goods electoral platforms instead of clientelistic electoral platforms impacts on electoral behaviour and leads to sanctioning of candidates who use them; organisation membership dampens the negative response of voters to public goods platforms.
Wild and Harris (2011)	Case studies	Malawi (agriculture and education in 2 districts)	Community scorecard	Service delivery	0 Community scorecard does not impact on local service delivery; service delivery is constrained by patronage relationships and the centrality of the Presidency in decision making.

Note: * means that the study is part of the meta-analysis in Dunning (2019).

Appendix 2: Tables

Table A1: Summary statistics for control variables, Afrobarometer citizen perceptions of local councillors' performance

Control variables	Overall sample		Treated		Control		Difference in means (<i>p</i> -value)
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	
District level control variables							
Population (number)	484,951	284,019	571,095	424,045	459,701	220,926	0.000
Population growth	3.023	1.24219	3.009	1.309	3.027	1.222	0.501
Access to clean drinking water (%)	66.860		72.808		65.117		0.000
District was split (proportion)	0.013		0.031		0.008		0.000
Individual level control variables							
Age of the respondent	34.968	13.156	35.231	13.081	34.891	13.178	0.224
Respondent is female	0.497		0.499		0.496		0.828
<i>Home language (Excluded category: Other small languages)</i>							
Acholi	0.044		0.065		0.038		0.000
Ateso	0.063		0.076		0.060		0.002
Langi	0.064		0.063		0.064		0.876
Luganda	0.150		0.196		0.136		0.000
Lugbara	0.028		0.025		0.028		0.372
Lumasaba	0.042		0.077		0.032		0.000
Lusoga	0.084		0.110		0.077		0.000
Rukiga	0.044		0.037		0.047		0.038
Runyankole	0.101		0.106		0.100		0.368
Runyolo	0.032		0.051		0.027		0.000
<i>Education levels (Excluded category: No education)</i>							
Primary education	0.436		0.418		0.442		0.027
Secondary education	0.339		0.344		0.338		0.568
Higher education	0.099		0.118		0.094		0.000
<i>Religion (Excluded category: None, traditional or other religion)</i>							
Christian	0.871		0.879		0.868		0.137
Muslim	0.112		0.100		0.115		0.027
<i>Living conditions (Excluded category: Living conditions are identical)</i>							
(Very) bad living conditions	0.501		0.492		0.504		0.299
(Very) good living conditions	0.282		0.306		0.275		0.001
<i>Consumption needs: Number of times gone without enough food in past 12 months (Excluded category: At most once or twice)</i>							
Several times gone without food	0.192		0.197		0.190		0.425
Many times gone without food	0.093		0.085		0.096		0.070
Always gone without food	0.035		0.023		0.038		0.000

<i>Watching TV (Excluded category: Never watches TV)</i>				
Watches TV at most once a week	0.109	0.110	0.108	0.837
Watches TV several times a week	0.065	0.073	0.063	0.061
Watches TV every day	0.076	0.134	0.060	0.000
<i>Listening to the radio (Excluded category: Never listens to the radio)</i>				
Listens to the radio at most once a week	0.037	0.047	0.034	0.002
Listens to the radio several times a week	0.075	0.106	0.066	0.000
Listens to the radio every day	0.782	0.744	0.794	0.000
<i>Reading of newspapers (Excluded category: Never reads the newspaper)</i>				
Reads newspaper at most once a week	0.197	0.197	0.197	0.926
Reads newspaper several times a week	0.111	0.124	0.108	0.017
Reads newspaper every day	0.044	0.050	0.043	0.092
<i>Attendance of public meetings</i>				
Attended public meetings now and then	0.260	0.255	0.261	0.535
Attended public meetings several times	0.360	0.356	0.362	0.585
Attended public meetings often	0.150	0.153	0.149	0.580
Treatment				
Scorecard intervention	0.227			

Note: For the calculation of the summary statistics of the control variables and the yearly observations we took the largest number of observations available, i.e. 12,622. The number of treated (control) observations is 2,861 (9,761).

Table A2: Results for the Afrobarometer data employing an ordered logit model

	Councillors are perceived as being corrupt (ordinal)	Performance of local government councillors (ordinal)	Councillors listen to what people have to say	Trust in the local government council	Trust in the parliament	Quality of maintenance of local roads by local government
Scorecard intervention (coefficient estimate)	-0.278** (0.014)	0.005 (0.974)	-0.008 (0.965)	0.066 (0.644)	0.136 (0.299)	-0.031 (0.870)
Marginal effects per category	0.042** (0.015)	-0.001 (0.974)	0.002 (0.965)	-0.011 (0.643)	-0.023 (0.298)	0.009 (0.870)
	0.058** (0.014)	-0.001 (0.974)	0.000 (0.965)	-0.012 (0.645)	-0.026 (0.299)	0.002 (0.870)
	-0.054** (0.017)	0.001 (0.974)	-0.001 (0.965)	0.004 (0.645)	0.009 (0.299)	-0.008 (0.870)
	-0.046** (0.012)	0.001 (0.974)	-0.001 (0.965)	0.019 (0.644)	0.040 (0.298)	-0.004 (0.870)
Observations	12,086	12,523	9,194	12,622	12,562	9,332

Note: Ordered logit results employing the same control variables as specified in the note of Table 5. Standard errors are clustered at the district level; corresponding *p*-values are presented in parentheses. First the coefficient associated with the treatment effect is shown and then the corresponding marginal effect per outcome category. More information on the outcome categories is presented in section 4.1.

Table A3: Results for the budgetary outcomes controlling for NRM related electoral outcomes

	Total public spending (log)	Per capita public spending (log)	Share of the budget spent	Delay in reporting to the central level	Local contribution to revenues (log)	Primary school leaving exam pass rate	Number of health centres and hospitals
Panel A: Main specification with electoral results for district chair of 2011 elections							
Scorecard intervention	-0.048 (0.281)	0.013 (0.679)	-3.851* (0.092)	-0.005 (0.990)	1.178 (0.230)	-0.304 (0.753)	1.762 (0.161)
District chair is member of the NRM as result of the 2011 local elections	-0.042 (0.556)	-0.020 (0.731)	0.159 (0.967)	-0.342 (0.470)	-1.596 (0.304)	-1.362 (0.379)	1.560 (0.373)
Panel B: Electoral results for district chair interacted with treatment							
Scorecard intervention	-0.011 (0.824)	0.024 (0.506)	-1.549 (0.435)	-0.236 (0.480)	1.068 (0.346)	-0.336 (0.785)	2.467 (0.142)
District chair is member of the NRM as result of the 2011 local elections	-0.019 (0.797)	-0.013 (0.833)	1.641 (0.655)	-0.490 (0.318)	-1.666 (0.294)	-1.383 (0.378)	2.012 (0.283)
Interaction (Scorecard x District chair NRM)	-0.087 (0.276)	-0.027 (0.630)	-5.492* (0.057)	0.552 (0.112)	0.262 (0.858)	0.078 (0.954)	-1.687 (0.315)
Panel C: Main specification with electoral results for NRM share in 2011 elections							
Scorecard intervention	-0.041 (0.318)	0.016 (0.587)	-3.612 (0.103)	0.002 (0.996)	1.152 (0.244)	-0.454 (0.635)	1.818 (0.152)
NRM share	0.340 (0.108)	0.179 (0.229)	13.892 (0.181)	0.161 (0.870)	-2.676 (0.379)	-9.998*** (0.000)	4.381 (0.327)
Panel D: Electoral results for NRM share interacted with treatment							
Scorecard intervention	-0.006 (0.915)	0.016 (0.655)	-1.376 (0.482)	-0.248 (0.496)	0.864 (0.475)	-0.002 (0.999)	1.702 (0.199)
NRM share	0.332 (0.114)	0.179 (0.227)	13.374 (0.184)	0.228 (0.823)	-2.599 (0.387)	-10.116*** (0.000)	4.412 (0.324)
Interaction (Scorecard x NRM share)	-0.087 (0.322)	-0.001 (0.983)	-5.456 (0.118)	0.610 (0.132)	0.703 (0.689)	-1.118 (0.504)	0.282 (0.861)
Observations	1,064	1,067	1,064	1,072	1,072	1,069	1,072

Note: All regressions contain district fixed effects, time fixed effects and district-specific time trends. In addition, the regressions control for access to clean drinking water as a proxy of district level poverty, the population size (log), the population growth rate (log) and the event of a district split as one observable structural change. The panel dataset contains information for the years 2005 to 2016 and for 111 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.

Table A4: Results for the Afrobarometer outcomes controlling for NRM related electoral outcomes

	Councillors are perceived as being corrupt (ordinal)	Performance of local government councillors (ordinal)	Councillors listen to what people have to say	Trust in the local government council	Trust in the parliament	Quality of maintenance of local roads by local government
Panel A: Main specification with electoral results for district chair of 2011 elections						
Scorecard intervention	-0.197** (0.015)	0.004 (0.971)	-0.035 (0.775)	0.030 (0.668)	0.078 (0.238)	-0.031 (0.831)
District chair is member of the NRM as result of the 2011 local elections	-0.065 (0.319)	-0.214** (0.044)	-0.644*** (0.000)	-0.075 (0.313)	-0.012 (0.798)	-0.210 (0.271)
Panel B: Electoral results for district chair interacted with treatment						
Scorecard intervention	-0.249** (0.012)	-0.044 (0.730)	-0.084 (0.575)	0.064 (0.436)	0.118 (0.120)	-0.000 (0.997)
District chair is member of the NRM as result of the 2011 local elections	-0.101 (0.133)	-0.248** (0.012)	-0.670*** (0.000)	-0.050 (0.507)	0.017 (0.705)	-0.194 (0.317)
Interaction (Scorecard x District chair NRM)	0.124 (0.149)	0.115 (0.295)	0.063 (0.757)	-0.081 (0.198)	-0.097 (0.123)	-0.040 (0.853)
Panel C: Main specification with electoral results for NRM share in 2011 elections						
Scorecard intervention	-0.198** (0.015)	-0.002 (0.987)	-0.067 (0.598)	0.030 (0.675)	0.077 (0.245)	-0.085 (0.581)
NRM share	-0.085 (0.527)	-0.545*** (0.006)	-1.199*** (0.003)	-0.094 (0.489)	-0.044 (0.698)	-1.216*** (0.004)
Panel D: Electoral results for NRM share interacted with treatment						
Scorecard intervention	-0.251** (0.026)	-0.013 (0.923)	-0.487* (0.052)	0.063 (0.480)	0.124 (0.142)	0.280 (0.180)
NRM share	-0.082 (0.537)	-0.544*** (0.006)	-1.372*** (0.000)	-0.097 (0.487)	-0.048 (0.677)	-1.065*** (0.004)
Interaction (Scorecard x NRM share)	0.122 (0.319)	0.025 (0.858)	0.602 (0.107)	-0.074 (0.376)	-0.108 (0.173)	-0.522 (0.119)
Observations	12,086	12,523	9,194	12,622	12,562	9,332

Note: All regressions contain district fixed effects, time fixed effects and district-specific time trends. In addition, the regressions control for district characteristics (population size, population growth, access to clean water, dummy variable for the event of a district split) and individual characteristics related to respondents' age, gender, home language, education level, religion, consumption needs, frequency of media usage and attendance of public meetings. The dataset contains information for the years 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.