

WIDER Working Paper No. 2013/023

Measuring government performance in public opinion surveys in Africa

Towards experiments?

Michael Bratton*

March 2013

Abstract

In examining the study of government performance, this paper asks whether field experiments can improve the explanatory precision of results generated by public opinion surveys. Survey research on basic health and education services sub-Saharan Africa shows that the perceived ‘user friendliness’ (or ease of use) of services drives popular evaluations of government performance. For the reliable attribution of causality, however, surveys and field experiments, combined in a variety of mixed research designs, are more rigorous and reliable than either method alone. The paper proposes a menu of such designs.

Keywords: government performance, surveys, experiments, health, education, Africa
JEL classification: H4, I00, I18, I25

Copyright © UNU-WIDER 2013

*Michigan State University, USA, email: mbratton@msu.edu

This study has been prepared within the UNU-WIDER project ‘ReCom—Research and Communication on Foreign Aid’, directed by Tony Addison and Finn Tarp.

UNU-WIDER gratefully acknowledges specific programme contributions from the governments of Denmark (Ministry of Foreign Affairs, Danida) and Sweden (Swedish International Development Cooperation Agency—Sida) for ReCom. UNU-WIDER also gratefully acknowledges core financial support to its work programme from the governments of Denmark, Finland, Sweden, and the United Kingdom.



The World Institute for Development Economics Research (WIDER) was established by the United Nations University (UNU) as its first research and training centre and started work in Helsinki, Finland in 1985. The Institute undertakes applied research and policy analysis on structural changes affecting the developing and transitional economies, provides a forum for the advocacy of policies leading to robust, equitable and environmentally sustainable growth, and promotes capacity strengthening and training in the field of economic and social policy making. Work is carried out by staff researchers and visiting scholars in Helsinki and through networks of collaborating scholars and institutions around the world.

www.wider.unu.edu

publications@wider.unu.edu

UNU World Institute for Development Economics Research (UNU-WIDER)
Katajanokanlaituri 6 B, 00160 Helsinki, Finland

Typescript prepared by Janis Vehmaan-Kreula at UNU-WIDER.

The views expressed in this publication are those of the author(s). Publication does not imply endorsement by the Institute or the United Nations University, nor by the programme/project sponsors, of any of the views expressed.

1 Introduction

Citizens have a stake in the quality of government. As consumers of public goods and services, the general public is uniquely situated to judge how governments are handling various policy tasks—from the delivery of basic services to the creation of jobs—that directly affect popular welfare. Moreover, if people live in a democracy—the preferred form of political regime around the world—they enjoy rights to express opinions about government performance.

This paper contends that public opinion surveys based on nationally representative probability samples are a suitable method for describing and analysing public reactions to government performance. Well-designed surveys can capture patterns in popular sentiment and behaviour across different geographic regions or social groups. Comparative surveys can track trends in public opinion over time, juxtapose satisfaction with services across policy sectors, and contrast one government's performance with another. If the right questions are asked, survey research can even reveal the foundations—cognitive, affective, or evaluative—on which people form opinions about government performance.

But there are drawbacks to any kind of observational research. A typical problem is endogeneity, which refers to the circularity or reciprocity embedded in social relationships. Does A cause B? Or, vice-versa, does B cause A? For example, do people think governments perform well because they enjoy easy access to desired public services? Or do they judge service accessibility in terms of their general impression of how well governments are performing? It is insufficient to simply observe that a relationship exists; we also need to unravel which comes first, preferably by specifying some kind of causal mechanism. In a related problem, public policy practitioners face the issue of attribution. They want to know whether a given policy, programme or project leads to intended social outcomes. The challenge is to distinguish the consequences of a policy intervention from a welter of other contributory factors. Where a planned policy intervention is only one variable in a kaleidoscopic social setting, policy impact is difficult to discern.

The twin features of multi-causality and endogeneity are endemic to social reality and thus intrinsic to social research (Franzese 2007). Public opinion has multiple roots, being based simultaneously on a person's social background, exposure to government, and deeply held identities, interests and ideas. In short, any explanation of how people judge government performance is bound to be multi-causal. Moreover, to avoid cognitive dissonance, people endeavor to think and act consistently, thus raising the prospect of endogeneity. Untangling the Gordian knot of social causation is a perennial challenge for researchers.

According to scientific norms, the laboratory experiment is the best method—a 'gold standard' if you will—for addressing issues of causality (Druckman et al. 2006). Experimental research designs allow researchers to maximize control over objects of interest. But many of the most compelling questions in social science do not lend themselves to inquiry under the artificial conditions of a laboratory. Thus researchers endeavour to extend the logic of experimental research to the collection and analysis of field data. Randomized controlled trials (RCTs), designed *ex ante* and rigorously conducted, seek to simulate laboratory controls. Even where RCTs are not feasible, quasi-experimental research designs, for example conducted *ex post* on extant data, can provide a measure of insight into causality.

Experiments and quasi-experiments have long informed research in development economics (e.g., Banerjee and Duflo 2009; Barrett and Carter 2010; Duflo 2006) and recently in comparative politics (e.g., Fearon et al. 2009; Humphreys et al. 2006; Humphreys and Weinstein 2009; Moehler 2010). These studies often employ social surveys as measurement tools to generate baseline indicators or response data from experimental groups. Currently, researchers are beginning to apply the logic of experiments to the structure of surveys through panel designs (in which the same respondents are interviewed over time in order to trace and attribute change), via vignettes embedded in questionnaires (that expose survey respondents to manipulated stimuli in order to assess differential effects) or behavioural games run parallel to surveys (to test theories about the effects of economic, social or policy incentives).

This paper examines the study of government performance through public opinion surveys in Africa and asks whether the logic of scientific experimentation can improve explanatory precision. The central argument is that surveys and field experiments, combined in a variety of mixed research designs, are more rigorous and reliable than either method alone.

The first section examines the concept of government performance and introduces points of convergence between survey and experimental methods. Data from the Afrobarometer are then employed to make operational measurements of government performance in two sectors—primary education and basic health care—and to explain performance outcomes as seen by citizens. Addressing multi-causality, I argue that the perceived accessibility (or ‘user-friendliness’) of social services drives popular satisfaction with government performance. Recognizing that a degree of endogeneity remains in this result, however, the paper discusses a menu of econometric and experimental methods to address it. A conclusion makes the case that, far from being rival methods, survey and experimental research designs are complementary.

2 Framework

2.1 Government performance

The object of interest in this paper is *government performance*, that is, the capacity of governing authorities to provide public goods and services. Public goods and services are benefits intended for shared consumption from which nobody can be legally or effectively excluded. These benefits may be general political guarantees (such as national defense, law and order, or civil rights) or particular socio-economic commodities (such as transport infrastructure, electricity supply, or health and education services). Either way, government performance concerns the track record of state agencies at addressing collective social needs.

In studying government performance, researchers refer to best practices in the delivery of public goods and services, which are usually measured in terms of the outputs, outcomes or impacts of public programmes, including citizen assessments thereof. At the heart of the enterprise is performance management: the planning, programming, implementation, measurement, monitoring, evaluation, and reporting of programme results. In a high-performing government, each agency defines problems and goals, identifies programmes to address goals, names leaders, and outlines strategies for implementation, along with key

performance indicators. Ideally, agency heads are held accountable for ensuring the performance of the institutions and policies for which they are responsible (GPS 2012).

As researchers, we need to be clear about the specific points of reference for government performance. Are we measuring attributes of government institutions, government policies, or government personnel? A well-designed research instrument would explicitly distinguish among these alternative objects. Using classic principles of public administration, we may differentiate *institutions* by level of government—central, sub-national, or local—or by function—such as specialized technical or service activities (Fesler 1949). As far as public *policies* are concerned, we may distinguish extractive, regulatory or distributive policies or make precise reference to specific policy sectors such as health or education. When it comes to government *personnel*, we want to know how citizens evaluate the performance of various officials, especially front-line agents who interact directly with the general public.

2.2 Surveys versus experiments?

The question arises as to the most appropriate method for assessing government performance, particularly as experienced by citizens. Whatever method is chosen—whether a broad national survey of mass opinion or a field experiment focused on a local policy intervention—researchers immediately face the issue of case selection. Simply stated, which units should be chosen for observation and analysis? Because it is difficult (and anyway unnecessary) to measure the characteristics of all relevant units, most evaluation research is designed around the logic of sampling. In surveys and experiments alike, researchers face the challenge of drawing samples that represent relevant populations of eligible citizens. For both surveys and experiments, randomization is the preferred selection mechanism, meaning that every appropriate person has an equal probability of being selected into the sample.

Sampling strategies differ mainly in terms of the scope of the population under consideration. For national surveys of government performance, researchers wish to refer to the citizenry as a whole, often meaning adult individuals of voting age. The latest official census, updated where necessary with projections for population growth and movement, affords a suitable sampling frame. Although a national cross-section may be constructed in several stages to take into account various demographic, administrative or cultural strata in the population, the final product is one sample representing the country as a whole. For experiments on government performance, researchers require more fine-grained information about citizen exposure to particular government programmes. As such, they assign respondents to treatment and control groups, each on a random basis. The procedure produces at least two distinct but comparable samples, or more if the researcher wishes to measure different levels of treatment. Because survey and experimental methods share the underlying logic of randomization, however, there is no insuperable barrier to mixing these methods in hybrid research designs.

3 Measuring government performance

This paper uses data from the Afrobarometer, an independent, large-scale, comparative public opinion research project (Afrobarometer 2013). To be valid and reliable, indicators of government performance must refer to specific objects of measurement and be stated in terms that ordinary people can comprehend.

3.1 Institutional performance

The structure of modern government resembles a vertical pyramid composed of central, regional and local tiers. And within each of these tiers, distinctive institutions perform various executive, legislative, judicial, service delivery or other functions. Recognizing this diversity, but being unable to exhaustively cover every kind of government structure and function, the designers of the Afrobarometer make strategic choices.

The first addresses central government. The survey instrument measures the public's trust in a list of specific apex institutions such as the presidency, the parliament, the electoral commission, the tax department, the police, the army and the national courts of law. The second tier is local government. In recent Afrobarometer surveys, respondents are asked to evaluate how well or badly their municipal or district council is performing quotidian functions such as the maintenance of roads, the regulation of land use, and sanitation (Bratton 2012).

3.2 Personnel performance

Popular views about the performance of government derive not only from individual experiences with institutions; they also depend on daily face-to-face interactions. In practice, the way that agents of the state routinely relate to citizens may decide whether the government gets high or low marks for performance. Are the public's experiences with officials marked by respect and responsiveness? Or are they marred by rudeness and neglect? Critical too is whether rent seeking and corruption creep into official behaviour.

To tap these aspects of government performance, the Afrobarometer deploys batteries of questions on the perceived job performance and honesty of particular government officials, including police officers, tax collectors, customs agents and elected local government councillors. Beyond general perceptions of official corruption (which can be inflated by gossip and rumor) other items ask about individual experience with demands for bribes.

3.3 Policy performance

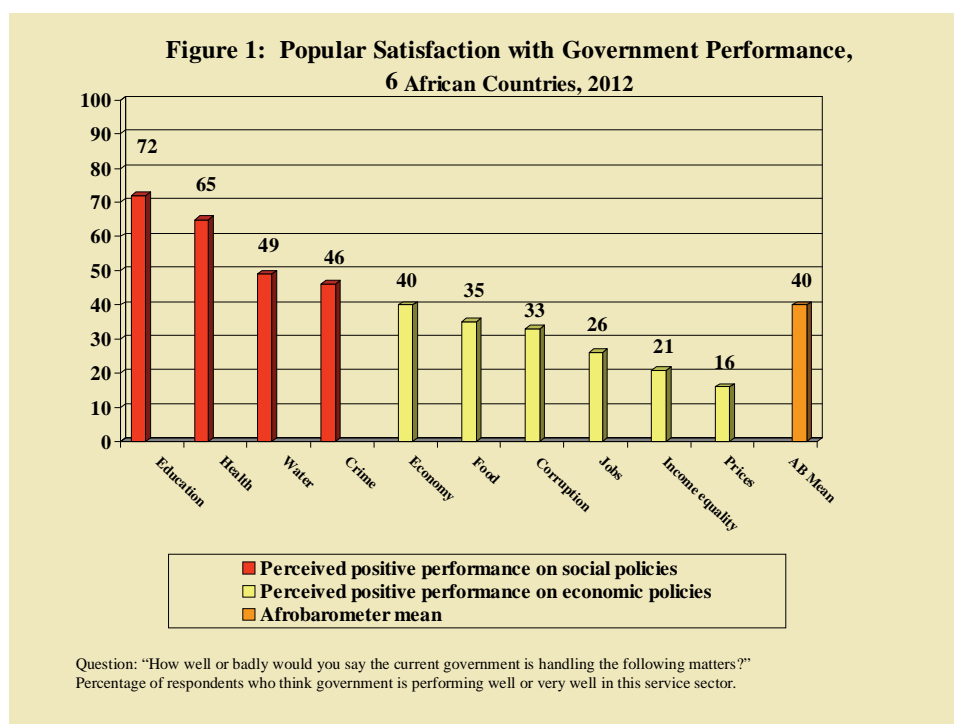
To capture popular perceptions of policy performance, the Afrobarometer applies a root question—'how well or badly is the government handling the following matters?'—to policies in more than a dozen sectors. These range from macro-economic policies—such as job creation and inflation control—to micro-social policies—like delivering household water and controlling crime.

As will be shown, the Afrobarometer has devoted more effort to measuring policy performance than any other aspect of government performance. The survey has delved in detail into two fundamental government services: primary education, and basic health care. The choice is appropriate because many people encounter these services as they try to improve family welfare. Health and education rank as the top social priorities among the most important problems that the government should address (Bratton 2009). And both sectors have received prior attention from researchers using experimental methods (e.g., Banerjee et al. 2010a; Glewwe et al. 2000).

4 Government performance in Africa

This paper employs public opinion data for six African countries in 2012—Benin, Cape Verde, Kenya, Mauritius, South Africa, and Uganda—to describe profiles of government performance. The starting point is popular satisfaction with the delivery of public goods and services across a variety of policy sectors. Figure 1 suggests that Africans distinguish among these domains and arrive at separate judgments about each. In the social sectors, people consider that governments are performing quite well. About two-thirds approve of government performance in the delivery of education and health services (72 and 65 per cent respectively). Perceived government performance on every social service, including domestic water supply and crime control, exceeds the Afrobarometer mean (40 per cent) for government performance.

Not so with economic policies; only 40 per cent thought that the governments in their countries were doing well at managing the national economy. Similarly, more people scored governments as doing ‘badly’ rather than ‘well’ on a range of activities from ensuring food security to controlling corruption. Moreover, performance at all economic tasks was evaluated as falling at or below the Afrobarometer mean for government performance. At the extreme, only about one quarter of respondents gave a positive rating to performance at job creation, inflation control and closing the gap between rich and poor.



Source: Afrobarometer, Round 5, 2012, www.afrobarometer.org

A factor analysis on all responses about satisfaction with government performance generates two components that validate the sharp distinction in perceived performance between social and economic sectors. Reliability analysis also confirms, however, that an overarching concept bridges this distinction, allowing an overall index of ‘government performance’. The country mean scores for this index along a four-point scale from ‘very badly’ to ‘very well’ have face validity. On average, citizens rate government performance relatively well in Mauritius, a more developed country than Uganda. Importantly, though, since scores for all

countries fall below the mean on the scale, the Africans we interviewed are apparently more dissatisfied than satisfied with overall government performance.

The analysis below employs three indicators of government performance: ‘satisfaction with health services’, ‘satisfaction with education services’, and ‘satisfaction with basic social services’, which is an average construct of both (health and education) indicators. The construct is permissible because the people who are satisfied with health services tend to also be satisfied with education services, and vice versa. But what are the main determinants of satisfaction with these aspects of government policy performance?

4.1 Accessibility of services

Infrastructure

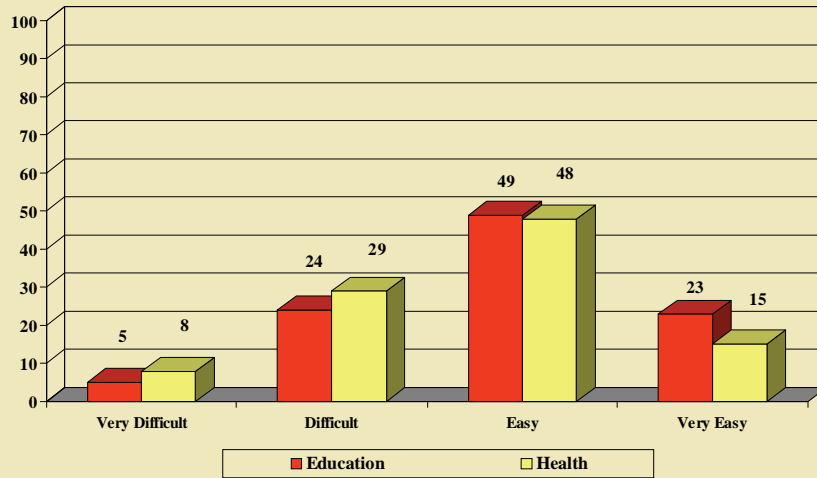
One possible source of public satisfaction is the physical proximity of service infrastructure in the locations where people live. After all, an individual’s prospect of gaining access to services starts from the convenient availability of a nearby service outlet (World Bank 2004). The Afrobarometer measures the proximity of service infrastructure by the observed presence or absence of primary schools and health clinics in every primary unit in the survey sample. In the six countries studied, the density of service infrastructure is everywhere greater for schools than clinics. On average, more than four-fifths of adults live in areas with access to a local primary school compared to less than three-fifths who possess ready access to a local health clinic.

User-friendliness

Quite apart from physical proximity, the accessibility of services depends upon the organizational feature of ‘user-friendliness’. From a user’s perspective, services may be simple, transparent and inclusive or they may be formal, complex and exclusionary. For poor or illiterate people, the approachability of a bureaucratic state may be a prime consideration. The relevant survey questions are direct: ‘In your experience, how easy or difficult is it to obtain the following services: A place in a primary school for a child? How about medical treatment at a nearby clinic? Or do you never try to get these services from the government?’.

An initial examination of the data shows that people find it easier to get a child into school than to get medical attention (Figure 2). Whereas, across six African countries in 2012, 72 per cent reported that it is easy to gain access to a basic educational service, some 63 per cent said the same about medical care.

**Figure 2: Ease of Access to Education and Health Services:
Popular Estimates, 6 African Countries, 2012**



"In your experience, how easy or difficult is it to obtain the following services:
A place in a primary school for a child? Medical treatment at a nearby clinic?
Or do you never try to get these services from government?"
Figures exclude those who never tried to use these services.

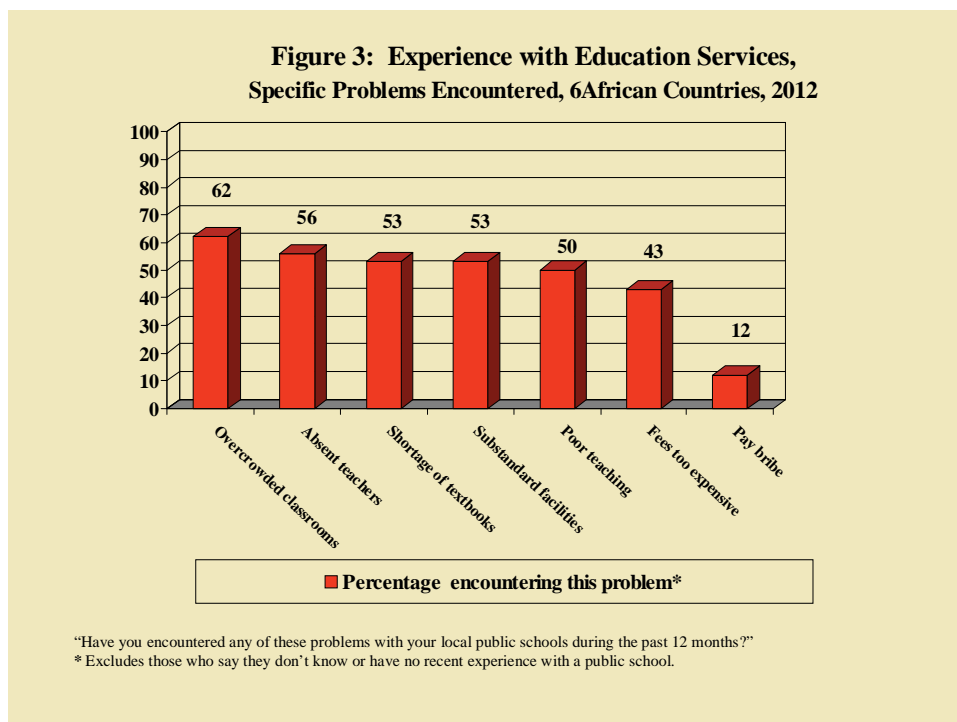
Source: Afrobarometer, Round 5, 2012, www.afrobarometer.org

4.2 Experiences with education services

Which obstacles—of service availability, quality, and cost—arise most frequently? For education, the survey asked, ‘Have you encountered any of these problems with your local public schools during the past 12 months?’. A list of six problems was then read out, ranging from ‘overcrowded classrooms’ to ‘unable to pay’. Respondents were also asked separately if they had ever, in the previous year, had to ‘pay a bribe, give a gift, or do a favour’ to a government official in order to ‘get a place in a primary school for a child’.

Figure 3 compares the reported frequency of problems arising with education services. In this case, only those persons are counted who have had contact with primary schools during the previous 12 months.

Because popular demand for education exceeds the supply of school facilities, overcrowded classrooms are the most common problem, reported by 62 per cent of users. It is cited most often in countries like Kenya and Uganda with universal free primary education (UPE). The related problem of teacher absenteeism arises almost as frequently (56 per cent). About half of all African users register objections to a trio of core issues: shortages of textbooks and other teaching materials, substandard school buildings and facilities, and the low quality of instruction.



Source: Afrobarometer, Round 5, 2012, www.afrobarometer.org

Is public schooling too expensive? Are users unable to pay? In the litany of client problems, the costs of primary schooling assume low priority. Fewer than half of all respondents say that the expense of required fees inhibits them from sending children to school. In this instance, the provision of UPE unexpectedly fails to reduce popular concern about the costs of schooling: indeed, in 2012, Kenyans and Ugandans are *more* likely than other Africans surveyed to complain about the high levels of school fees. Even in these countries, parents still face a bevy of unofficial charges and expenses.

Finally, few users (12 per cent) say they confront demands for illegal payments from teachers or school administrators, for example in return for school placement. Such corruption reportedly hardly ever happens in Mauritius or Cape Verde (so say under 2 per cent), but it is said to be common in Kenya and Uganda (over 20 per cent). Perhaps because teachers and administrators feel overstretched by the influx of waves of new pupils under UPE, they feel justified in seeking illicit rents.

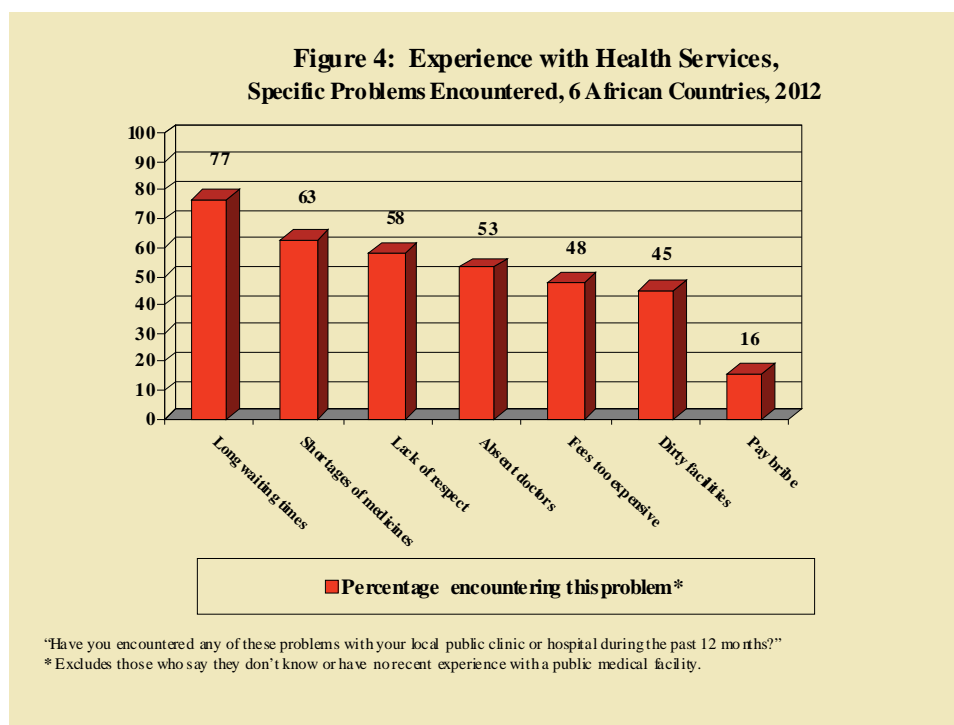
For the record, it is worth noting that specific problems of access to education form a composite whole. In other words, people who perceive one obstacle in gaining access to education are likely to see other obstacles too. An average index of this factor—labelled as ‘service experiences (education)’—is expected to predict popular satisfaction with education services. In order to tease out the relative influence of specific problems encountered, however, each problem will first be entered separately in the explanatory models that follow.

4.3 Experiences with health care services

A parallel set of questions was asked about health care: ‘Have you encountered any of these problems with your local public clinic or hospital during the past 12 months?’. A list of six problems was offered, ranging from ‘long waiting times’ to ‘lack of attention or respect from

staff'. Respondents were also asked separately if they had ever had to 'pay a bribe, give a gift, or do a favour' to a government official in order to get treatment.

Figure 4 breaks down the recent experiences of persons who attempted to use clinics and hospitals. On average, slightly more users report a specific problem with health services (51 per cent) than with education services (47 per cent).



Source: Afrobarometer, Round 5, 2012, www.afrobarometer.org

This discrepancy is most evident in relation to overcrowded facilities, where three quarters (77 per cent) of clinic users complain about 'long waiting times'. By a clear margin, delays in delivery at the point of service are the biggest problem. On any given day, urban hospitals are typically unable to accommodate all patients; long lines of applicants regularly assemble outside rural clinics; and, too often, some people are turned away at the end of the day without consultation or treatment. Relative to effective client demand, which is higher for healthcare than schooling, health services are in even shorter supply than education services.

Nearly two thirds of healthcare clients also report shortages of medicines and medical supplies. Regardless of whether a sound health infrastructure has been built, local clinics may lack basic commodities needed for routine preventive care. Huge cross-country disparities are evident. Whereas over 80 per cent or more of Kenyans and Ugandans express concern about the under-provisioning of health facilities, the same proportion in Mauritius does not.

About half of all users of health services also find fault with other common problems: staff who behave disrespectfully toward patients (58 per cent), doctors who fail to arrive at work (53 per cent), the high cost of consultations or medicines (48 per cent), or dirty or unsanitary facilities (43 per cent). The data show that these sorts of complaints are *least* common in the relatively richer countries of Mauritius and South Africa and more prevalent in relatively poorer countries like Benin and Uganda.

Across all countries, however, health service problems are significantly more common in rural than urban areas. Waiting times are longer at rural clinics in part because of the sparser coverage of health infrastructure in remote areas; medicines are less readily available due partly to long supply lines from the capital city; and doctors are more often absent, in part because professionals are reluctant to serve at distant outposts. In the only exception to this general tendency, clients claim that medical staff—nurses, technicians, and clerks—are more likely to treat them rudely at urban hospitals and clinics.

Finally, few Africans (16 per cent) say they encounter demands from health care workers for illegal payments ('bribes'). Demands for bribes tend to be more common at rural clinics, perhaps because Ministry officials find it difficult to supervise field staff in outlying areas.

As in the education sector, specific experiences of health access cohere into a single reliable factor. This convergence enables the construction of an average index called 'service experiences (health)'. Moreover, there is a marked coincidence between 'service experiences (education)' and 'service experiences (health)'. On these grounds, a meta-index of 'service experiences (combined)' is employed in analysis below.

4.4 Corruption

Popular encounters with official corruption corrode service satisfaction (Barr et al. 2009; Olken 2007). As a key component of bad governance, the multifaceted concept of corruption is worth measuring from various angles. One would expect that, if ordinary people perceive official corruption to be prevalent, negative effects on service satisfaction would follow. The Afrobarometer asks, 'how many government officials do you think are involved in corruption?'. This question taps the popular reputation of service providers independently of whether an individual respondent has ever actually paid a bribe. In absolute terms, public workers in Africa have yet to win reputations for honesty; on average, 39 per cent of the Africans we interviewed perceive that 'most' or 'almost all' are corrupt.

A gulf exists between perceptions of corruption and the act of handing over an inducement. As reported earlier, an average of only 14 per cent said they did the latter in the past year in return for education or health services. As such, people either over-estimate the extent to which corruption pervades their society, or they under-report their own involvement in the socially disapproved act of paying a bribe. Or, most likely, both these biases are present in the data. As such, the real level of illegal exchanges of public goods for private gain probably lies somewhere between these extreme estimates (i.e. between 14 and 39 per cent).

For purposes of analysis, we will want to explore whether perceptions or experiences of corruption have the bigger impact on service satisfaction. And we will wish to confirm that such effects always run in the expected, negative direction.

5 Explaining service satisfaction

This section of the paper addresses these issues, among others. The goal is a comprehensive, multivariate explanation of popular satisfaction with public services. As discussed, the object of study is represented by several versions of the dependent variable, namely satisfaction with education services, satisfaction with health services, and an average index of both services.

The proposed explanatory factors are the independent variables surveyed above under the rubrics of social structure, service accessibility, service experiences, and corruption.

What, then, are the sources of public satisfaction (and dissatisfaction) with service delivery in a sample of six African countries? The OLS regression models in Table 1, which control for country fixed effects (not shown), tell a similar story regardless of the way that satisfaction is measured. The strongest and most statistically significant relationships are highlighted in bold and their explanatory power is ranked in parentheses.

The most important consideration—always ranked first—is service accessibility. What matters most is whether clients consider services to be ‘easy to use’. This feature of ‘user-friendliness’ captures whether citizens regard public services as being uncomplicated to operate. It addresses the extent to which clients wish to avoid entanglements in administrative red tape and formal interactions with officials whom they feel to be socially distant. If ordinary people can get a child into school or receive primary health care with a minimum of such hassles, they are likely to be satisfied with service delivery.

Table 1: Sources of Service Satisfaction,
6 African Countries, 2012

	EDUCATION SERVICES	HEALTH SERVICES	BOTH SERVICES
Social Structure			
Poverty	-.077*** (3)	-.073*** (4)	-.079*** (4)
Service Accessibility			
Availability of service infrastructure	.064*** (4)	.050***	.063***
Perceived ease of service access	.181*** (1)	.232*** (1)	.266*** (1)
Service Experiences			
Fees too expensive	-.058**	-.078*** (3)	
Shortages of supplies	-.004	-.045**	
Poor quality teaching /treatment	-.133*** (2)	-.007***	
Poor/dirty facilities	-.029	-.086*** (2)	
Service experiences (education)			-.113*** (2)
Service experiences (health)			-.082*** (3)
Corruption			
Perception of corruption	-.029*	-.022	-.023
Experience of corruption	-.006	-.051***	-.044***
Constant	2.415***	2.299***	2.060***

Cell entries are standardized regression coefficients = beta. (Explanatory rank in parentheses). *Constant* = B.
Country fixed effects included but not shown. Weak predictors (beta <.015) not shown. ***p <.001, **p <.01, *p <.05

Source: Afrobarometer, Round 5, 2012, www.afrobarometer.org

It is important to note that the subjective element of service accessibility is more critical than the objective one. As Table 1 shows, physical infrastructure—whether there is a school or clinic in the locality—remains less important to service satisfaction. The relative superiority of subjective over objective criteria raises a challenge for government ministries responsible for health and education in Africa. Success at service delivery is not simply a matter of building more clinics and schools. Instead, it requires an organizational commitment to an ethic of customer service by which the client comes to feel that his or her needs are being considered and addressed.

As for social structure, the poverty status of users—measured in terms of shortages of basic human needs—is the key social consideration. The connection of poverty to service (dis)satisfaction is at least three times as strong as the average for other societal influences like gender, age or residential location (not shown). And poverty’s negative impact is consistent for both health and education services. The poorer a person, the less likely is he or she to be satisfied with government performance at social delivery.

This robust result indicates that existing services embody an anti-poor bias, at least in the opinion of the poor themselves. To the extent that poverty is more prevalent in the rural areas of every country surveyed, a bias in accessibility is offset and obscured by the apparent willingness of rural residents to accept lower quality services. This combination of facts suggests that, African governments that seek to improve popular perceptions of performance, would be well advised to design pro-poor health and education policies and to direct these services initially to urban populations.

Service experiences—that is, the quality of users’ encounters with service providers—are also part of a complete explanation. With regard to education services, poor quality of teaching matters most to respondents, many of whom are parents of school age children. By contrast, two other factors loom large in shaping popular satisfaction with healthcare services: the expense of service fees and substandard facilities (‘dirty clinics’).

In other words, clients seem willing reluctantly to tolerate the inconvenience of excess of demand for services above the available supply (as manifest in overcrowded classrooms and long waiting times at clinics). They seemingly value education and health care so highly that they resign themselves to congestion as an unavoidable cost of accessing scarce services. But they nonetheless insist that teachers provide high quality education and that the government equip medical staff with the necessary infrastructure and supplies to do their jobs properly.

Finally, what is the impact of corruption? Table 1 indicates that, if people enter into corrupt transactions with health clinic officials, their satisfaction with services drops by significant margins. In other words, actual encounters with bribery have negative effects on service satisfaction that are independent of, and larger than, mere perceptions of official graft. It is worth noting that this result departs from an original analysis of 18 African countries in 2005, which showed that paying bribes to get services marginally increased service satisfaction in the health sector (Bratton 2009). The adjudication of this discrepancy awaits the availability of 2012 data from a larger number of countries.

6 Explaining government performance

To summarize, citizens in several African countries appraise the delivery of social services principally in terms of ‘user-friendliness’, that is, whether these services are easy to use. This feature has less to do with the physical proximity of schools and clinics and more to do with customer relations, that is, whether citizens find front-line workers to be approachable and responsive. Citizens want to know that government officials understand their needs and are committed to public service.

But how far does ‘user-friendliness’ travel? While it may be an essential feature of social services, does it also apply to government performance writ large? After all, the overall performance of the government encompasses not only the delivery of health and education

services (which citizens rate in a mostly positive light), but also to the management of the economy (which citizens rate negatively) (Figure 1). One might expect, therefore, that a government with relatively strong performance in the social sectors could still find difficulty in obtaining public satisfaction across the board, including with regard to its economic policies.

But this is not what we find when we apply the same regression model to a broad index of government performance (derived from all items in Figure 1). Rather, a customer-friendly set of social services appears to create a halo of goodwill that enhances public perceptions of the government's overall performance. Table 2 shows that the perceived ease of access to health and education services remains the principal determinant of citizen assessments of government performance *as a whole*. To be sure, such positive assessments are offset if citizens experience unhelpful encounters with staffs at government clinics and schools. But this result only serves to underline the importance of the treatment handed out by front-line public workers. In this case, perceptions (not experience) of corruption loom much larger in the public's assessment of overall government performance, perhaps because respondents are suspicious of elite behaviour in policy arenas, like economic management, that are technically complex and opaque to public scrutiny.

Table 2: Sources of Perceived Government Performance, 6 African Countries, 2012

	GOVERNMENT PERFORMANCE
Social Structure	
Poverty	-.095***
Service Accessibility	
Availability of service infrastructure	.037*
Perceived ease of service access	.172*** (1)
Service Experiences	
Service experiences (education)	-.119*** (3)
Service experiences (health)	-.107*** (4)
Corruption	
Perception of corruption	-.139*** (2)
Experience of corruption	-.035**
Constant	2.013***

Cell entries are standardized regression coefficients = beta. (Explanatory rank in parentheses). *Constant = B*.

Country fixed effects included but not shown. Weak predictors (beta <.015) not shown. ***p <.001, **p <.01, *p <.05

Source: Afrobarometer, Round 5, 2012, www.afrobarometer.org

7 Towards experiments?

So far, the empirical results in this paper have been based on observational data. Issues of multi-causality were addressed by modelling government performance with multivariate regression under *ceteris paribus* conditions. In so doing, we learned that public opinion about government performance depends more on the quality of service provision than on the physical quantity of service outlets.

But what about causal direction? Given that the ‘user-friendliness’ of services is the best explanation of popular satisfaction, we cannot dismiss a potential loop of endogeneity. Might the causal arrow run in a reverse direction? Citizens who are broadly satisfied with the performance of the incumbent government might be led to conclude that such services are easy to use. Indeed, it could be argued that ‘ease of use’ is simply a proxy for (that is, an indirect way of measuring) policy or institutional performance. This concern is heightened by an inherent weakness of survey data: both factors are perceptions offered by the same informant, and thus likely to be mutually constructed.

So how can we distinguish the chicken from the egg? How can we be certain that we have attributed performance outcomes correctly to the quality of face-to-face interactions between service providers and their clients? This section of the paper distinguishes econometric or statistical approaches to addressing the causal problem from experimental approaches.

A simple test for causality—sometimes called ‘poor man’s exogeneity’—is the passage of time. Actions that occur in the past can be assumed to be independent of events that occur later. As a starting point, it is reasonable to assume that people form evaluations of government performance after an encounter with a service provider. Thus experience of user-friendliness is causally prior to service satisfaction. But social preferences can confound this seeming causal connection; citizens bring expectations to their interactions with governments. Supporters of opposition political parties, for example, may be predisposed to judge performance harshly regardless of ease of service access. Conversely, loyalists of an incumbent government may be overly generous in overlooking deficits in the responsiveness of state agents. And other contaminating orientations may remain even if partisan identification is controlled for.

To tease out reciprocal feedback loops, researchers may use simultaneous equation models (Bollen 1989; Greene 2002; Barretto and Howland 2005). This statistical technique employs regression analysis in two stages as a means of detecting causation within observational data. Unlike a single-equation model in which a dependent variable (y) is a function of independent variables (x), the y variable is used simultaneously as a predictor in its own right. To control for endogeneity (x causes y causes x), simultaneous equations include instrumental variables (IVs) that are chosen because they are uncorrelated with (exogenous to) the error term in the explanation. If the IV turns out to correlate with x when it is used as a dependent variable, however, then we can conclude that we are in the presence of reverse (or bidirectional) causation.

Instrumental variables, however, are difficult to identify, are often conceptually vague, and are easily misapplied. Moreover, non-specialists find simultaneous equations hard to execute and interpret. Thus the technique is rarely used outside econometrics, for example in analysis aimed at policy practitioners. To employ it here, we would have to find an IV that predicts both government performance and the user-friendliness of services, but which is independent of both. One is hard pressed to conceive of such a factor, let alone to find an operational indicator for all the ten thousand cases of survey respondents in the present study. The only conceivable source is another item in the same survey, but we already know that survey respondents adjust their attitudes by avoiding cognitive dissonance, defaulting to political partisanship, or relying on informational shortcuts, thus inviting endogeneity.

This is not to say that two-stage least squares models with instrumental variables can never be used with survey data. A separate Afrobarometer study asked which comes first: state

building or democracy? Having established that citizens associate the rule of law with an effective state, the authors demonstrate that this public good is produced not just by effective states but also by democratic regimes (Bratton and Chang 2006). As such, while democracy is most easily built on the foundations of an established state, state building is also a product of democratization.

Given the complexity of applying *ex-post* econometric fixes to survey observations, however, researchers tend to prefer *ex-ante* research designs based on experimental logic as a possible way forward. Random assignment of treatments in controlled field trials would seem to improve the odds of reliably attributing causality. Is it therefore possible to design field experiments on government performance at social service delivery? A limited literature has begun to find positive effects of participatory monitoring of service providers by citizens in the developing world, including in sub-Saharan Africa (SSA) (Paul 2000; Bjorkmann and Svensson 2009; Olken 2010). An experiment would be most pertinent to the concerns of this paper if the focus were on education or health services at the grassroots level and if the treatment of interest were the user-friendliness of service delivery. Such an experiment would contrast an accessible delivery mechanism with one that citizens found difficult to use; the experiment would manipulate the way that staff treats clients.

The relevant services are a primary school place for a child and medical treatment for oneself or one's family. An experiment in the educational sector could therefore examine the effects on popular satisfaction of a generous school admissions policy. The treatment would throw open the school doors to any child who wished to attend by doing away with the need for documentary proof of the child's identity, age and place of residence or requirements that parents purchase school uniforms or school supplies. The control group would be admissions procedures as currently constituted. An experiment in the health sector could follow similar lines. Clinic staff would be required to admit all-comers without any requirement for verification, registration or payment. All clients, not only the indigent, would receive free consultation and treatment without any fee or co-payment. Again, the control group would be exposed to extant procedures.

At issue would be the impact of these treatments on popular satisfaction with service delivery. The main hypothesis would be that, relative to existing procedures, open and accessible treatments would induce citizens to express greater satisfaction. Providing the sample of participants were large enough, the experiment could be refined to include different packages of treatments as a way of discerning, for example, whether identity documents, fee payments, corruption or other hurdles to service accessibility were blocking satisfaction.

The Afrobarometer data suggest that client experiences are distinctive in each service sector. In the African countries we studied, citizens drew most attention to poor quality teaching in schools as opposed sub-standard (dirty) facilities in the health clinics. Experiments could be designed to randomly vary these treatments in schools and clinics. Perhaps policy initiatives could improve teacher quality in randomly selected primary schools, whether through retraining or improved salaries, or upgrade the medical supply chain or building maintenance schedules for randomly selected health clinics. The treatment-control logic would then be applied to a comparative analysis of service satisfaction among clients of service outlets 'with and without' policy improvements.

An experimental study on the use of insecticide-treated bed nets to combat malaria in Kenya confirmed the importance of the 'user friendliness' of health services (Cohen and Dupas

2008). It questioned whether the imposition of a nominal charge for a service ensured that clients value the good provided, thereby avoiding wastage. The study showed unexpectedly that pregnant women who received free bed nets were no less likely to use them than others who paid a subsidized price. Because cost sharing dampened demand, the study concluded that free distribution (the more accessible option) would save lives, especially among young children.

Under field conditions, mundane and moral considerations often disrupt the abstract purity of experimental research designs. Government officials, even if open to evidence-based policy research (a big ‘if’), commonly resist treating target communities differentially because they wish to avoid charges of political favouritism. Moreover, the ethics of human development demand that no group be deprived of opportunity to improve wellbeing simply for research purposes. For this reason, opportunities for access to services would have to be added to treatment groups rather than subtracted from control groups. But these sorts of honourable considerations constrain the ability of researchers to manipulate the full range of variance in the objects of their interest.

Most importantly, the results of field experiments cannot be assessed without data. The question arises therefore as to where such facts and figures come from. Experimental researchers always have to rely on information from research subjects including, very commonly, data elicited through public opinion and other social surveys. The final section of this paper discusses a menu of options for taking advantage of complementarities in experimental and survey research methods.

8 A menu of mixed methods

We first consider three integral permutations in which surveys and experiments are combined (*ex-ante*) into a single research design: (a) surveys *around* experiments; (b) surveys *within* experiments, and; (c) experiments *within* surveys. We then turn to two permutations in which separate surveys or experiments are deployed (*ex-post*) in sequential fashion; (d) experiments *beyond* surveys and (e) surveys *beyond* experiments. The pros and cons of each admixture are discussed in the context of interventions to improve the delivery of basic public services in sub-Saharan Africa.

8.1 Surveys around experiments

A quasi-experimental ‘before-and-after’ research design is a straightforward way to combine surveys and experiments. This design brackets a project intervention between two surveys administered at different times. As a first step, a ‘before’ survey provides baseline data about the beliefs, attitudes and behaviours of a randomly selected sample of prospective clients. The ‘before’ survey precedes the project intervention, which then takes place. At some subsequent time, when project effects and impacts are expected to have appeared, an ‘after’ survey is administered to a random sample of project participants in order to take a second measurement of the variables of interest. Researchers then compare ‘before’ and ‘after’ results and attribute any differences to the project.

The key assumption of before-and-after experimental designs is that changes in outcomes from time t to time $t+1$ are likely to have a causal connection. But challenges remain. First, attribution: is the project the only variable factor in the period between surveys? Or can

outcomes be ascribed to other, uncontrolled contextual changes? Second, how can researchers contain research costs and obtain timely results when the design requires mounting field teams on at least two separate occasions?

Third, even if baseline and participant samples are randomly drawn, do they amount to well-matched cross-sections? After all, certain participants—often persons who are more proximate, educated or well connected—may select themselves into a project. If these individuals bring distinctive values, attitudes and behaviours into the project, they may cause selection bias. One possible solution is to employ a panel design—meaning that the same people are interviewed before and after the project. But a panel design has its own shortcomings, the most common of which is the inevitable attrition of respondents over time, sometimes to the point that the second or subsequent samples become too small to support reliable generalizations.

As an omnibus instrument, the Afrobarometer lends itself to use as a source of baseline data when surveys are required *around* experiments. It offers cost savings and convenience to policy analysts who are unable to mount customized surveys of field conditions prior to planned interventions. Due to limitations of sample size, however, these data can only be disaggregated reliably to the region, provincial or state level, and cannot be used to track projects at the district level or below (subject to exceptions described in the next section). But, because this survey employs a national probability sample, it can be used for monitoring and evaluating planned government interventions on a national scale.

For example, a donor agency commissioned an evaluation of the impact of civic education programmes in Zambia based on pilot Afrobarometer surveys. The first survey followed the emerging Afrobarometer standard of a randomized national probability sample; the second covered a smaller snowball sample of programme participants identified by project leaders. Initial results suggested that civic education had positive impacts on political knowledge and occasionally changed popular values like political tolerance. These results survived statistical controls for citizens' education levels and exposure to mass media. The problem of self-selection was addressed by imposing statistical controls for political activism, measured in terms of these individuals' political efficacy and organizational skills. As a result, the study concluded that persons exposed to civic education remained slightly more likely than the average Zambian to register to vote (Bratton et al. 1999; see also Finkel 2002, Finkel et al. 2012).

8.2 Surveys within experiments

An alternative approach embeds surveys within experiments as instruments to measure the differential effects of treatments and controls. This 'with-and-without' design does not require a baseline observation from an earlier period. Instead, researchers randomly assign participants in a service project to Treatment Groups ($T_1 \dots T_k$, depending on levels of service) and a Control Group (C). When project outcomes have become visible, researchers administer one common survey questionnaire to all participants. They expect respondents in Groups $T_1 \dots T_k$ (the groups with various project treatments) to show greater and more positive project effects than respondents in Group C (the standard control group, that is, without treatment).

A ‘with-and-without’ design has the advantage of roughly approximating laboratory conditions in the field. The quality of the simulation depends on the researcher’s success at isolating presumed causal factors (treatments) while holding other conditions constant (controls). Much also depends on whether subjects are assigned to treatment and control groups on a random or matched basis, and are therefore comparable.

But field settings pose inconvenient obstacles. First, it is difficult to perfectly match treatment and control groups. One common way to limit environmental variance is to locate the study in neighbouring communities; in this case, one community receives the public service and the other does not. But the advantages of geographic proximity may be offset if interactions among neighbours contaminate the purity of the treatment-control distinction. Second, the intense scrutiny of researchers in a confined locality may have an unintended Hawthorne Effect, whereby the very act of observing research subjects changes their attitudes or behaviours. Third and finally, field experiments, especially if conducted on a pilot basis within strict boundaries, suffer a similar problem to laboratory experiments. One can never be certain that results are externally valid, that is, that they apply outside the experimental setting, for example, when service programmes are scaled up to wider populations (Barabas and Jerit 2010).

Omnibus public opinion surveys like the Afrobarometer survey are *not* designed for use *within* a field experiment. To serve this purpose, modifications are required to the survey’s sampling protocol and questionnaire instrument. A purposive over-sample must be added that is large enough to encompass treatment and control groups, both randomly selected, within a specific project area. In a second requirement, questionnaires must include items that explore how respondents react to particular project goals, procedures and outcomes. For this reason, service providers who wish to conduct a ‘with-and-without’ experiment are advised to design their own survey sample and instrument.

Note, however, that the Afrobarometer has assisted an international development assistance agency with this task in Uganda, Mali, and Senegal. The subject matter was local government performance. The treatment groups were local councils that received donor-funded institution-building inputs; control groups were councils that did not. In Uganda, the project focused on strengthening local government councils in 16 of the country’s 80 districts, with the remaining 62 districts serving as a control. The donor sponsored an over-sample of 100 cases in each of 16 treatment districts and added several questions about local government performance. This collaboration offered project managers the opportunity to test for impact at both the collective level (i.e., 16 districts combined) and, at the level of individual districts within a wider margin of error (Logan 2011). Unfortunately, to our knowledge, the donor agency never made full use of the data in order to derive policy relevant results about the efficacy of their local government strengthening programmes. This shortcoming suggests that, where sponsors lack in-house research capacity, analysis should be subcontracted to the survey agencies that collect the data.

8.3 Experiments within surveys

Given the state of the art in public opinion studies, *experiments within surveys* lie at a research frontier. This approach embeds randomized controlled experiments inside survey instruments.

Researchers use vignettes—or capsule stories—as a means to expose survey respondents to varying stimuli (‘treatments’) *within an interview*. A sub-set of interviewees, who serve as a control group, receives no variations of treatment. The stimuli are hypothetical scenarios in which the researcher systematically manipulates a variable that is expected to influence outcomes of interest. For example, South African survey respondents, when exposed to contrasting stories about the eviction of a squatter, were found to be more likely to see the act as legitimate if they believed that it occurred by means of due process of law (Gibson 2008). Other researchers varied the names of candidates for public office in order to reveal that perceived candidate ethnicity influences voters’ expressed voting intentions (Conroy-Krutz 2012; Dunning and Harrison 2010).

On the positive side, experiments within surveys allow researchers to move beyond passive observation by actively varying an object of interest. As such, the method offers a means to increase leverage over causal inference. On the negative side, hypothetical vignettes are just that: hypothetical. The researcher is left wondering whether any observed impacts are accompanied by behavioural changes in the real world. Moreover, vignettes take up a lot of scarce space and time in survey questionnaires. And random assignment of respondents is logistically challenging under the far-flung field conditions of a social survey, thus inviting error.

8.4 Experiments beyond surveys

Two final approaches allow researchers to triangulate on a given research problem, even if multiple methods were not envisaged as part of an original research design. For example, researchers who wish to test for theorized causal mechanisms may choose to add experiments based on economic games as supplements to surveys (Camerer 2003).

Take ethnicity, a factor relevant to government performance in Africa. Existing studies have established that ethnic diversity, especially if reinforced by inter-group inequality, inhibits public goods provision (Miguel 2004; Baldwin and Huber 2010). A survey in urban Kampala, Uganda both confirmed this negative relationship and provided a frame for randomly selecting participants for a field experiment (Habyarimana et al. 2007). The researchers asked participants to play economic games and provided cash incentives to join others in community projects to provide clean drains, garbage collection, and security from crime. Participants rewarded or punished potential collaborators (shown in photographs) according to whether or not they saw them as co-ethnics. Analysis of these games showed that ethnic co-operation was strategic; it did not grow out of shared cultural preferences but rather from rational expectations of reciprocity or sanctions.

If experiments conducted after surveys can reveal *why* empirical relationships exist, then they promise theoretical progress. Game theory is certainly a useful device for exploring behavioural outcomes that respond to economic incentives. But, even if conducted in the field, behavioural games do not fully replicate real-world contexts and thus do not entirely vitiate the problem of external validity associated with laboratory experiments. Moreover, as applied in the literature so far, games seem better suited to modelling client behaviour than the performance of government service providers. Finally, while scholars may wish to excavate the theoretical foundations of behaviour, it is far from clear that practitioners need to do so; the latter may be content to know ‘*what works*’ in policy terms without necessarily delving deeply into *why* it does so.

8.5 Surveys beyond experiments

In a final combination of methods, researchers may use the results of field experiments to improve the construction survey instruments. The purpose here is to test the external validity of field experiments on large populations. For example, several experimental studies from various parts of the world suggest that public information campaigns can increase popular satisfaction and induce participatory behaviour (Boeri and Tabellini 2010; Collier and Vicente 2011; Gerber and Green 2000). The external validity of such results can be tested with questionnaire items about the availability, nature and spread of public information in national or cross-national surveys.

For example, experiments that monitor the effects of public report cards on the performance of elected political representatives would seem especially well suited to the study of social service delivery (Humphreys and Weinstein 2012; Banerjee et al. 2010b). There is no logical barrier to using report cards to inform the general public about the quality of services at schools and clinics, especially in the urban areas of Africa where respondents have a measure of choice in the consumption of education or health services (e.g., public or private, traditional or modern). To my knowledge, no such experiment has yet been conducted. But, if it were, its results would be helpful in designing questionnaire items that tap respondents' levels of knowledge about service provision, including from report cards. In this way, experiments would strengthen survey instruments that are used to test theory and praxis about government performance.

9 Conclusion

This paper has argued that, as tools for explaining government performance, public opinion surveys and field experiments are potentially compatible research methods. Both methods employ random selection of cases, either to match treatment and control groups or to make inferences to wider populations. By combining surveys and experiments in mixed research designs, social scientists are able to generate more dependable results than by using either method in isolation. To take the example used here, mixed methods may help sort out whether the 'user friendliness' of social services is a precursor or a product of government performance.

At minimum, surveys are well suited to discover the attitudes and behaviour of a government's clientele: the citizens. This kind of raw data can provide an informative foundation for experimental research. Moreover, the large numbers of cases in survey data enable statistical consideration of multi-causality, that is, the diversity and relative weight of rival explanations of government performance. Public opinions tend to be mutually endogenous, however, which limits their utility in discerning the direction of underlying causes. This is where field experiments come in. Experimental controls allow analysts to manipulate a selected aspect of government performance such as the ease with which clients are able to access social services. The inference is that any observable difference between treatment and control groups is causally attributable to the experimental treatment.

Among various options, the most promising approach is to apply mixed survey and experimental methods to the field performance of government agencies, especially in developing countries where knowledge is often uncertain about 'what works'. The researcher may choose to embed surveys and experiments within a single, integral design or opt instead to employ sequential combinations of these methods on an as-needed basis. Much depends on

the evolving state of theoretical knowledge and on practical considerations such as the size of research budgets and the availability of existing data. But, all told, the future looks promising for the continued use of surveys and experiments as complementary devices for the measurement of the impacts of government performance in the field.

References

- Afrobarometer (2013). 'Survey and Methods'. www.afrobarometer.org.
- Baldwin, K. and J. Huber (2010). 'Economic versus Cultural Differences: Forms of Ethnic Diversity and Public Goods Provision'. *American Political Science Review*, 104(4): 644-62.
- Banarjee, A. and E. Duflo (2009). 'The Experimental Approach to Development Economics'. *Annual Review of Economics*, 1(1): 151-78.
- Banarjee, A., E. Duflo, and S. Khemani (2010a). 'Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation of Education in India'. *American Economic Journal: Economic Policy*, 2(1): 1-30.
- Banarjee, A., S. Kumar, R. Pande, and F. Su (2010b). 'Do Informed Voters Make Better Choices? Experimental Evidence from Urban India'. Unpublished paper, Massachusetts Institute of Technology, October.
- Barabas, J. and J. Jerit (2010). 'Are Survey Experiments Externally Valid?'. *American Political Science Review*, 104(2): 226-43.
- Barr, A., M. Lindelow, and P. Serneels (2009). 'Corruption in Public Service Delivery: An Experimental Analysis'. *Journal of Economic Behaviour and Organization*, 72(1): 225-39.
- Barrett, C. and M. Carter (2010). 'The Power and Pitfalls of Experiments in Development Economics: Some Non-Random Reflections'. *Applied Economic Perspectives and Policy*, 32(4): 515-48.
- Barretto, H. and F. Howland (2005). *Introductory Econometrics*. New York: Cambridge University Press.
- Bjorkman, M. and J. Svensson (2009). 'Power to the People: Evidence from a Randomized Experiment of a Citizen Report Card Project in Uganda'. *Quarterly Journal of Economics*, 124: 735-70.
- Boeri, T. and G. Tabellini (2010). 'Does Information Increase Political Support for Pension Reform?'. Unpublished paper, Bocconi University, April.
- Bollen, K. (1989). *Structural Equations with Latent Variables*. New York: Wiley.
- Bratton, M. (2009). 'Are You Being Served? Popular Satisfaction with Health and Education Services in Africa'. In G. Bland and C. Arnson (eds), *Democratic Deficits: Addressing Deficits of Sustainability and Consolidation Around the World*. Washington, DC: Woodrow Wilson Center.
- Bratton, M. (2012). 'Citizen Perceptions of Local Government Responsiveness in Sub-Saharan Africa'. *World Development*, 40(3): 516-27.
- Bratton, M., P. Alderfer, G. Bowser, and J. Temba (1999). 'The Effects of Civic Education on Political Culture: Evidence from Zambia'. *World Development*, 27(5): 807-24.

- Bratton, M. and E. Chang (2006). 'State Building and Democratization: Forwards, Backwards, or Together?'. *Comparative Political Studies*, 39(9): 1059-83.
- Camerer, C. (2003). *Behavioural Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton University Press.
- Cohen, J. and P. Dupas (2008). 'Free Distribution or Cost-Sharing? Evidence from a Malaria Prevention Experiment'. *National Bureau of Economic Research Working Paper No. 14406*.
- Collier, P. and P. Vicente (2011). 'Votes and Violence: Evidence from a Field Experiment in Nigeria'. *Working Paper*, Centre for the Study of African Economies, University of Oxford.
- Conroy-Krutz, J. (2012). 'Information and Ethnic Politics in Africa'. *British Journal of Political Science*, First View online, September.
- Druckman, J., D. Green, J. Kuklinski, and A. Lupia (2006). 'The Growth and Development of Experimental Research in Political Science'. *American Political Science Review*, 100(4): 627-36.
- Duflo, E. (2006). 'Field Experiments in Development Economics': Paper prepared for the World Congress of the Econometric Society, January.
- Dunning, T. and L. Harrison (2010). 'Crosscutting Cleavages and Ethnic Voting: An Experimental Study of Cousinage in Mali'. *American Political Science Review*, 104(1): 21-39.
- Fearon, J., M. Humphreys, and J. Weinstein (2009). 'Can Development Aid Contribute to Social Cohesion after Civil War? Evidence from a Field Experiment in Post-Conflict Liberia'. *American Economic Review: Papers and Proceedings*, 99(2): 287-91.
- Fesler, J. (1949). *Area and Administration*. Birmingham: University of Alabama Press.
- Finkel, S. (2002). 'Civic Education and the Mobilization of Political Participation in Developing Democracies'. *Journal of Politics*, 64(4): 994-1020.
- Finkel, S., J. Horowitz, and R. Rojo-Mendoza (2012). 'Civic Education and Democratic Backsliding in the Wake of Kenya's Post-2007 Election Violence'. *Journal of Politics*, 74(1): 52-65.
- Franzese, R. (2007). 'Multicausality, Context Conditionality and Endogeneity'. In C. Boix and S. Stokes (eds), *The Oxford Handbook of Comparative Politics*. Oxford: Oxford University Press.
- Gerber, A. and D. Green (2000). 'The Effects of Canvassing Phone Calls and Direct Mail on Voter Turnout: A Field Experiment'. *American Political Science Review*, 94(3): 653-63.
- Gibson, J. (2008). 'Group Identities and Theories of Justice: An Experimental Investigation Into the Justice and Injustice of Land Squatting in South Africa'. *Journal of Politics*, 70: 700-16.
- Glewwe, P., M. Kremer, S. Moulin, and E. Zitzewitz (2000). 'Retrospective versus Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya'. *National Bureau of Economic Research Working Paper No. 8018*.
- Government Performance Summit (GPS) (2012). <http://governmentperformance.org/index.htm>.
- Greene, W. (2002). *Econometric Analysis*. New York: Macmillan, fifth edition.

- Habyarimana, J., M. Humphreys, D. Posner, and J. Weinstein (2007). 'Why Does Ethnic Diversity Undermine Public Goods Provision?'. *American Political Science Review*, 101(4): 709-45.
- Humphreys, M., W. Masters, and M. Sandbu (2006). 'The Role of Leaders in Democratic Deliberations'. *World Politics*, 58(2): 583-622.
- Humphreys, M. and J. Weinstein (2009). 'Field Experiments and the Political Economy of Development'. *Annual Review of Political Science*, 12: 367-78.
- Humphreys, M. and J. Weinstein (2012). 'Policing Politicians: Citizen Empowerment and Political Accountability in Uganda'. Columbia University, draft paper, April.
- Logan C. (2011). 'The Uses of the Afrobarometer in Policy Planning, Program Design and Evaluation'. *Afrobarometer Briefing Paper* No. 98. www.afrobarometer.org.
- Miguel E. (2004). 'Tribe or Nation? Nation Building and Public Goods in Kenya versus Tanzania'. *World Politics*, 56(3): 327-62.
- Moehler, D. (2010). 'Democracy, Governance and Randomized Development Assistance'. *The Annals of the American Academy of Political and Social Science*, 628: 30-46.
- Olken, B. (2007). 'Monitoring Corruption: Evidence from a Field Experiment in Indonesia'. *Journal of Political Economy*, 115(2): 200-49.
- Olken, B. (2010). 'Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia'. *American Political Science Review*, 104(2): 243-67.
- Paul, S. (2000). 'Making Voice Work: the Report Card on Bangalore's Public Services'. Paper No. RPG 68207M for a project on Effectiveness of Client Surveys in Improving Service Delivery, World Bank.
- World Bank (2004). *World Development Report, 2004: Making Services Work for Poor People*. Washington, DC: World Bank.