UP School of Economics
Discussion Papers

Discussion Paper No. 0808 August 2008

Can information about local government performance induce civic participation? Evidence from the Philippines

by
Joseph J. Capuno* and Maria Melody S. Garcia**

*University of the Philippines
**University of Rome – Tor Vergata

Can information about local government performance induce civic participation? Evidence from the Philippines

Joseph J. Capuno[a*,] and Maria Melody S. Garcia[b]
[a]*University of the Philippines*
[b]*University of Rome – Tor Vergata*

Abstract

The question of whether people are motivated to engage in civic activities once informed of their local government's performance is relevant to decentralized governance. Applying propensity score matching technique on a unique household-level dataset from the Philippines, it is found that the knowledge of an index of local government performance has positive and statistically significant effects on the likelihood of membership in local organizations and participation in local projects. The estimated average treatment effects are robust to choice of comparison groups, matching algorithm and to possible effects of unobserved variables. Thus, transparency in local government activities can deepen citizenship.

JEL Classifications: D83, H70, C21

[*] Corresponding author. E-mail address: jjcapuno@up.edu.ph. The authors would like to acknowledge the institutional support from The Ford Foundation, the Philippine Center for Policy Studies, UP School of Economics and other local partners during the pilot of the GOFORDEV Index. Thanks are also due to the seminar participants in Turin, Italy where an earlier draft of this paper was presented..

1. Introduction

This study investigates whether information about local government performance can motivate people to participate or engage in civic activities using a unique set of household-level data from the Philippines. The issue bears on the design of decentralization policies and governance mechanisms to make sub-national or local governments better providers of local public services. Such policies or mechanisms that could strengthen accountability, transparency or democratic participation could improve the performance of local governments, which are often said to be more susceptible to elite capture than the national government (Platteau, 2000; Bardhan and Mookherjee, 2002; Platteau and Gaspart, 2003).

The importance of participation to development has been well discussed. One strand of the literature emphasizes the instrumental value of participation. In cases where program beneficiaries are involved in the design or implementation of public projects or activities, for example, the general outcomes are found to be generally positive (Isham et al., 1995; Blair, 2000; Jennings, 2000). Arguably, when more people participate different views and concerns are considered which could result in greater support for the collective decisions made. Also, participation in social or community-level organizations enables the local residents to acquire information about their local governments (Pierre, 1998; Alatas et al., 2002; Best and Dustan, 2006). Thus, in the process the officials' accountability is enhanced, and possibly trust or social capital is also developed (Putnam 1995). Another strand in the literature underscores the intrinsic value of participation, i.e., participation as a constituent component of development. Participation or social inclusion serves to fulfill the need of the marginalized sectors to become a part of the larger community (Schugurensky, 2003; Ohmer, 2007). In the political realm, for example, people may value their right to vote on its own, even though they may not exercise it during elections.

Various studies have been made on the determinants of political or civic participation. Some of these studies underline the importance of demographics, leadership, incentives of individuals, opportunities, and cost and benefit analysis (Olson, 1965; Moe, 1980; Gaspart et al., 1998; Alesina and La Ferrara, 2000; Weinberg and Jütting, 2000; Lowndes et al., 2001a; Agrawal and Gupta, 2005). Based on focus group discussions among minorities or members of excluded groups, Lowndes et al. (2001b) conclude that citizens are motivated to participate if local authorities are responsive and keep them informed of outcomes. Using US data, Knoke (1988) reported that the type of incentives determine the degree of involvement or participation. In general, these findings lend support to the model of the "rational voter" that emphasize the relative costs and benefits of participation in shaping individual decisions (Olson, 1965; Moe 1980).

Given the instrumental and intrinsic values of participation, several initiatives have been undertaken to widen the people's civic involvement. Among the well-known initiatives include the various participatory methodologies adopted in Bolivia that enabled non-government organizations (NGOs) to exercise greater influence over local governments (Blackburn and De Toma, 1998). A similar initiative in India empowered a community organization called the *Mazdoor Kisan Shakti Sangathan* (Association for the Empowerment of Workers and Peasants) to perform audits on local governments which led to the identification of cases of mis-spending of public funds and made some local officials accountable for the missing amounts (Goetz and Gaventa, 2001). Another well known initiative in India was the Citizen's Report Cards in Mumbai, Bangalore and Calcutta which basically announces the citizen's overall satisfaction with the services provided by government agencies in the concerned cities (Goetz and Gaventa, 2001). A similar survey was undertaken in the Philippines in 2000 (World Bank, 2001) and in Rwanda in 2004 (OSSREA, 2006).

The public announcement of performance ratings can have their desired effects on local government performance if the people will act on the information, or at least the responsible agencies or officials believe the people will do so. The issue of whether information alone can induce civic or social participation is investigated here, using household level data from a local governance project done in 2001-2003. Under this project, an index that captures the overall performance of 12 city and municipal governments in two Philippine provinces was piloted. The impact of the public dissemination of the index is assessed using propensity score matching techniques. The overall results indicate that information about local government performance encourage membership in local organizations and engagement in local public projects. Thus, the findings support the call for greater transparency and wider participation in local governance to help develop better citizenry.

The rest of the paper is organized as follows. In section 2, democratic participation under decentralization in the Philippines is briefly reviewed. Then, the local governance project that yielded the data used here is described in section 3. The evaluation framework is presented in section 4, and followed by description of its implementation in section 5. The results of the evaluation are analyzed in section 6. Short concluding remarks end the paper.

2. Democratic participation under decentralization

Several avenues opened up for democratic participation in local governance in the Philippines since the enactment of the Local Government Code of 1991. For one, the people can directly participate in legislations through referendum to propose, amend or enact local ordinances. Aside from voting for their candidates during local elections, they can also join in various local consultative bodies (LCBs) mandated to be established in all local government units (provinces, cities, municipalities and *barangays* or villages). Representatives of people's organizations (POs), non-government organizations (NGOs) and other private sector groups are regular members of LCBs which essentially review, recommend or provide inputs to local policies or programs in health, education, peace and order, and local development.

Since 1991, most local government units (LGUs) have constituted the mandatory local school boards, local health boards, local development councils, peace and order councils, local prequalification, bids and awards committees, and people's law enforcement boards. By 2000, for example, these local consultative bodies were already established in nearly all 71 cities (out of a total 96) that submitted reports to the Bureau of Local Government Supervision (Department of the Interior and Local Government, 2002). At least a fourth of the total members in the local development councils were from the private or non-government sector. Also, the duly elected president of the local parents-teacher association joins the local school board, and that two representatives of local POs or NGOs participate in the activities of the local bidding committees. Partly because of this feature of the Code, a wide range of POs, NGOs or voluntary organizations has also been established such as those by workers, urban poor, farmers and fisherfolks, teachers, businessmen (e.g., Rotary Club and Lion's Club), mothers, credit cooperatives, indigenous people and transport groups. Several coalitions of these groups also fielded their own candidates in national elections to vie for the reserved seats in Congress for the marginalized sectors.

Besides these venues for participation, more and better information about local governments was also made available to the local constituents. Besides their regular coverage in mass media, local government performances are also monitored and made known to the public through several quantitative and qualitative performance measures, customer or citizen satisfaction surveys, and awards and recognitions for exemplary LGUs. As of 2002, there were already 32 of these performance benchmarking and awards schemes administered by various national government offices, donor agencies, academic institutions and private sector groups. Among the famous ones are the province-level Human Development Index and the *Galing Pook* Awards for best practices in service delivery and local governance (Capuno, 2007). The results of these assessments are usually announced to the public.

With the opening of the venues for participation and the access to better information since 1986, social or civic participation have also deepened in the Philippines. The estimated number of NGOs registered with the Securities and Exchange Commission grew from 27,100 in 1986 to 50,800 by 1992 (Campos and Hellman, 2005). Because of the costly registration process, it is believed though that there are more NGOs, POs and other voluntary organizations than are officially recorded. There is also an increasing proportion of families with at least one member who was involved in at least one legitimate PO or local association for community development. Based on a series of

Annual Poverty Indicator Survey (APIS), the proportion rose from 16 percent in 1998 to 19.1 in 1999 and to 27 percent in 2002. Arguably, some of these households took advantage of the various civic opportunities present in their communities, while others were induced to take actions after knowing of the favorable or unfavorable performance of their local governments, and maybe the rest had both motivations. Untangling the different motives and measuring their relative effects on participation could help design better mechanisms or institutions for good local governance.

3. Data from a local governance project

The issue of whether informed citizens are more likely than others to engage in social or civic activities has not been fully investigated in the Philippines before, partly due to dearth of suitable data. In this paper, a unique dataset is used which contains household-level information on civic engagements, socioeconomic status and demographic characteristics. More importantly, the dataset include information on the household's knowledge of local government performance not usually collected in official surveys. The dataset comprises three rounds of household surveys undertaken to evaluate the impact of a local governance project implemented in two Philippine provinces. The especial feature of the project was its quasi-experimental set up in which data for treatment and control areas were collected.

The dataset was generated under the Good Governance and Local Development (GGLD) Project of the Philippine Center for Policy Studies (PCPS). Under the GGLD Project, the Governance for Local Development Index (GI) developed and piloted for two years (2001-2003) in12 component cities and municipalities in the provinces of Bulacan and Davao del Norte. The pilot test was undertaken, among others, to investigate if information about local government performance has any impact on the local citizens' level of participation and trust of their officials, and on the responsiveness of the LGUs to their announced performance ratings.

The pilot test of the index was set up as a quasi-experiment. In this setup, both treatment and control sites were selected to account for the wide differences in the levels of socioeconomic conditions across and within provinces in the country. In choosing the pilot provinces, all provinces were first grouped into two based on their relative levels development. Then one sample from each group was randomly picked. These are the Bulacan for the group with better than national average score in the Human Development Index and incidence of poverty in 2000, and Davao del Norte for the group with lower than average performance on both counts (Human Development Network, 2002; National Statistical Coordination Board, 2004). Then in each of the two provinces, the component sub-provincial LGUs were clustered based on their incomes into highly developed and less developed areas. From each cluster, three LGUs were randomly selected as pilot areas, with two of them as treatment sites and the other one as control site. The six pilot sites in Bulacan are San Jose del Monte City and the municipalities of Angat, Baliwag, Bustos, Guiguinto and Plaridel. In Davao del Norte, the six pilot areas are the cities of Island Garden City of Samal, Panabo and Tagum, and the towns of Asuncion, Braulio E.Dujali and Sto. Tomas.

The main project activities in the 12 pilot sites were the generation of the GI scores and their public dissemination. The two sets of activities were undertaken only in the eight treatment sites. In contrast, only the generation of the GI scores was done in the four control sites. To carry out all activities in these sites, local partners were formally contracted and given logistical and technical support. All local partners were trained and provided funds and materials in the conduct of surveys, data processing, and, where appropriate, public presentations and information dissemination. In each province, the local partners in the four treatment sites were two LGUs (local planning and development office) and two civil society organizations (NGOs, business groups and academic institutions), and those in the two control sites were civil society organizations as well (Table 1).

Table 1. The Pilot areas and the local partners

Relative Levels of Development	Bulacan			Davao del Norte		
	Treatment Areas		Control Areas	Treatment Areas		Control Areas
	LGU Partner	Civil Society Partner	Civil Society Partner	LGU Partner	Civil Society Partner	Civil Society Partner
High	San Jose del Monte City (City Planning and Development Office)	Baliwag (Soroptimist International of Baliwag)	Plaridel (Bulacan State University-Bustos Campus*, Rotary Club of Bustos**)	Panabo City (City Planning and Development Office)	Sto. Tomas (Davao Provinces Rural Development Institute, Inc.)	Tagum City (St. Mary's College-Tagum City*, University of Southeastern Philippines**)
Low	Guiguinto (Municipal Planning and Development Office)	Angat (Rotary Club of Angat)	Bustos (Bulacan State University-Bustos Campus*, Rotary Club of Bustos**)	Braulio E. Dujali (Municipal Planning and Development Office)	Island Garden City of Samal (LAWIG Foundation)	Asuncion (PhilNet-Rural Development Institute*, University of Southeastern Philippines**)

Notes: Names in parentheses are those of the local area partners.
* Local partner in 2001-2002 only.
** Local partner in 2002-2003 only.

The respective performances of the concerned LGUs in 12 pilot areas were assessed using the GI along three performance domains. The first domain is *public service needs*, which is measured with five indicators of access to and adequacy of basic services and the perceived effectiveness of the LGU in improving family welfare. The second domain is *expenditure prioritization*, which is indicated by the share of health, education and other basic services in total fiscal outlays. The last domain is *participatory development*, which captures with four indicators the extent of the functioning of the LCBs and the level of *barangay*-level public consultations. Ranging from zero (lowest) to 100 (highest), the scores in ten component indicators of GI were calculated based on household surveys and official documents from the LGUs such as audited financial reports and minutes of the LCB meetings.

The GI scores are publicly announced in the treatment areas through printed materials and public presentations. The materials distributed included posters, stickers and *komiks* (a popular reading fare using comic strips), which were translated into Tagalog language for the Bulacan areas and Bisaya language for the Davao del Norte areas. The number of *komiks* and posters distributed were proportional to the local population (Table 2). Moreover, local partners presented the GI scores in public forums at least three times in 2001 and then four times in 2002, when an additional forum was conducted exclusively for local officials.

The local partners were advised to distribute the information materials in all barangays and invite as many participants in the public fora as possible. The numbers of participants in the public fora and of the disseminated GI materials is shown in Table 2. In many of the treatment sites, the local partners disseminated the posters, stickers and *komiks* in public places like jeepney or tricycle terminals, municipal or city halls and market places. Also, they invited many members of local organizations to the presentations. In those places, therefore, the people who had knowledge of GI (and who were informed of their LGU's performance) may have been specifically targeted or self-selected. Since the actual distribution of GI materials and the invitation to GI presentations was not fully randomized, the evaluation of the GI's impact on civic participation would clearly become biased if based solely on the targeted groups.

Table 2. Public presentations and information materials

Pilot Areas	Number of Participants in Public Presentations*				Number of Information Materials Distributed				
	2001		2002		Komiks		Posters		Stickers
	Total	Non-gov't (%)	Total	Non-gov't (%)	2001	2002	2001	2002	2002
Bulacan	496	61	565	58	2000	3001	20000	8000	4000
Angat	99	95	126	82	198	397	1983	793	1000
Baliwag	116	75	163	47	526	1053	5263	2105	1000
Guiguinto	164	53	174	46	269	538	2688	1075	1000
San Jose del Monte City	117	30	102	63	1007	1013	10066	4027	1000
Davao del Norte	428	44	596	45	1999	2999	18999	6034	4000
B. E. Dujali	141	50	102	40	35	172	352	345	1000
Panabo City	87	15	224	28	907	1305	9069	2644	1000
Samal City	99	38	119	32	530	763	5304	1527	1000
Sto. Tomas	101	66	151	83	527	759	5274	1518	1000

Notes:
For 2001, the total number of *komiks* and posters distributed is equivalent to 30 per cent and 3 per cent of the local population, respectively.
For 2002, the total number of *komiks* and posters distributed is equivalent to 10 per cent and 5 per cent of the local population, respectively.
For 2002, the total number of stickers is equivalent to the following percentages of the local population: 13 per cent in Angat, 5 per cent in Baliwag, 9 per cent in Guiguinto, 2 per cent in San Jose del Monte, 29 per cent in Braulio E. Dujali, 4 per cent in Panabo City, 7 per cent in Island Garden City of Samal, and 7 per cent in Sto. Tomas.
The required numbers of public presentations in each area were three and four in 2001 and 2002, respectively.

To address the selection problem in the analysis of the GI's impact, three rounds of random household surveys were conducted in the 12 pilot areas. The surveys were scheduled to ensure appropriate collection of control and treatment data. The first survey was conducted in April-May 2001 - i.e. before the local partners undertook any activity – to obtain baseline information. This was followed by a mid-project assessment survey in

February- March 2002, after the local partners have completed their assigned activities in April –August 2001. After the second survey, the local partners again undertook the generation of GI scores and, for some, also the dissemination of the GI scores in March – September 2002. A final, post- project assessment survey was conducted in February- March 2003. The information collected in three surveys include household-level socioeconomic and demographic characteristics, knowledge of the GI, civic participation, trust of the local officials, and satisfaction with local government performance. To ensure the comparability of survey data from the 12 sites, the same sampling design and survey instrument were used. Specifically, a random sample of 100 household respondents who were at least 18 years old were interviewed with the same questionnaire in each of the 12 sites per survey.[1] Thus, the dataset assembled can be used with an appropriate evaluation methodology to obtain an unbiased estimate of the GI's impact on civic participation.

4. Evaluation framework

This section presents the commonly-used evaluation technique for non-experimental observational studies that is adopted here. This technique, called the propensity score matching (PSM), accounts for the non-random aspect of treatment at the individual level (i.e., exposure to GI through public fora or printed media). Since the local citizens were exposed to GI purposively, event through the treatments sites were randomly selected, a simple comparison of the means of participation rates in the treatment and control sites could lead to an overestimate of the true effects of GI. The bias may be due to special or predisposing characteristics or circumstances of those who participated in the GI presentations or received the GI materials not otherwise found in other residents in the pilot areas.

To address the selection problem, the propensity score matching (PSM) method is applied here. Originally developed in the statistics literature (Rosenbaum and Rubin 1983, 1984), the PSM technique has been used to obtain an unbiased estimate of the effects of various social interventions, such as job training programs (Heckman, Ichimura and Todd, 1997), pro-poor social health insurance (Trujillo, Portillo and Vernon, 2005), and education programs (Callahan, Wilkinson and Muller, 2008). Essentially, the PSM solves the selection problem by comparing the mean outcomes of the treated individuals and that of their matched control groups (of untreated individuals) with the same pre-treatment characteristics. In this case, whatever difference there is in the mean outcomes of the two groups is attributed to the treatment or intervention.

Following the PSM-based evaluation framework presented in Trujillo, Portillo and Vernon (2008) and in Dehejia and Wahba (2002), let M_{i1} be the value of the outcome variable (say, membership status) for the ith individual in the treatment group (say, the residents in the areas where GI scores were announced) and M_{i0} be the value for the residents in the control sites. When treatment is randomized, the effect of the treatment is the difference between the mean outcomes in the two groups. This is called the average treatment effect (ATE), as given by

$$\tau = E(M_{i1}) - E(M_{i0}) \qquad (1)$$

However, when treatment is not randomized, the estimate of the average treatment effect in (1) would be biased. The bias could be due to observed or unobserved characteristics that predispose certain individuals to participate in the treatment.

Instead of ATE, the average treatment effect on the treated (ATT) is calculated in non-experimental data to obtain an unbiased estimate of the effect of treatment but only on those who actually participated in the treatment. Estimating the ATT involves first finding the counterfactual outcome for the treated (i.e., the outcome had they not been treated), and then getting the difference between the mean outcome values with and without treatment for those who actually participated in the treatment. Formally, the ATT is given by

$$\tau_{|G=1} = E(M_{i1} \mid G_i = 1) - E(M_{i0} \mid G_i = 1) \qquad (2)$$

where G=1 means that the individual is treated (say, exposed to GI materials) and G=0 means that she is not. The problem here is that the individual can only be either treated or not treated at a given time. This means that it is possible to estimate $E(M_{i1}|G_i=1)$ but not $E(M_{i0}|G_i=1)$.

If $E(M_{i0}|G_i=1)$ in (2) is replaced with $E(M_{i0}|G_i=0)$,which can be estimated, a potential estimator of $\tau_{|G=1}$. However, this estimator of $\tau_{|G=1}$ τ is going to be biased as well if M_{i0} for the treated group and the comparison group systematically differ. The critical problem then is to find for each treated individual the appropriate comparator from a potential comparison group. The appropriate comparator for the ith individual (treated) should have the same pre-treatment characteristics or covariates, $\boldsymbol{X}_i$. But since the set of potential covariates could be large, matching the treated individual with a potential comparator could be difficult..

This particular problem ("curse of the dimensionality of $\boldsymbol{X}$") can be avoided with the use of PSM. Formally, the propensity score for the ith individual is defined as $p(\boldsymbol{X}_i)$=prob$(G_i=1|\boldsymbol{X}_i)=E(G_i|\boldsymbol{X}_i)$, i.e., the conditional probability of being treated defined over a set of observable variables or pre-treatment covariates. As shown in Rosenbaum and Rubin (1983), the propensity score has the following property

$$M_{i1}, M_{i0} \perp G_i \mid X_i \Rightarrow M_{i1}, M_{i0} \perp G_i \mid p(X_i) \qquad \forall i \qquad (3)$$

In words, if the outcomes are independent of treatment status after conditioning on observable covariates, then the same can be said after conditioning on the propensity score derived from the same observable covariates. Since the propensity score is a scalar, its use simplifies the matching of treated individuals with a suitable comparison unit.

An implication of (3) is that $E(M_{i0}|p(X_{\mathrm{i}}), G_i=0)$ can be used as an unbiased estimate of $E(M_{i0}|p(X_i), G_i=1)$, which cannot be calculated. With this implication, the ATT can now be computed as

$$\begin{aligned}\tau_{|G=1} &= E(M_{i1} \mid p(X_i), G_i = 1) - E(M_{i0} \mid p(X_i), G_i = 1) \\ &= E(M_{i1} \mid p(X_i), G_i = 1) - E(M_{i0} \mid p(X_i), G_i = 0) \qquad (2')\end{aligned}$$

The great advantage of using (2') as an estimator of ATT is that it is easy to compute the mean outcomes of the treated individuals and that of the non-treated individuals with the same propensity scores based on the same set of covariates as that of the treated individuals.

The use of PSM to derive an unbiased estimate of ATT rests on two crucial assumptions. The first assumption, called *unconfoundedness*, requires that the propensity scores should be based on all variables that could influence treatment assignment and potential outcomes simultaneously. That is, the treated units and comparison units should have balanced observable characteristics *and* that there should not be unobserved characteristics that could lead to selection bias. If the unconfoundedness assumption is not satisfied, then (3) would not be valid and (2') as well. The second requirement is that matching should be done along *common support*, the intersection of the conditional probabilities of the treated individuals and matched comparison units. Over this range, a comparator with the same *X* values exists for each treated individual. To satisfy these two requirements, various methods have been developed in the literature. Some of these methods are applied here, as discussed in the next section.

5. Implementation of the matching methodology

The use of PSM to obtain an unbiased estimate of ATT entails two critical steps. The first step is choosing the matched control subsample. This step involves the identification of the appropriate control group, the estimation of the propensity scores, and using the propensity scores to match the treated units with the subsample units from the control group. The implementation of these three sub-steps to obtain the appropriate comparison units is presented here. The second step in the PSM-based evaluation is to obtain the average treatment effects on the treated, which is discussed in section 6.

(a) *Choosing the comparison group*

To partly satisfy the unconfoundedness assumption, it is ideal to have the treatment group and the control group culled from similar, if not exactly the same, surveys. The use of similar or the same surveys would ensure that the two groups would have the same observable covariates. Similarities in sampling scheme or survey questionnaires would take into account the underlying market incentives, institutional constraints or the wider socioeconomic environment facing the individual respondents. It is found that the selection of comparison groups from a dissimilar survey could lead to a significant bias in the estimate of treatment effects (Heckman and Smith, 1995).

Some of these potential survey-related problems are avoided here with the use of the impact assessment data for the GI. As mentioned in section 3, this dataset was generated from the three rounds of household surveys conducted in the 12 pilot sites. In each pilot site, a total of 100 respondents were randomly selected in survey round. From this dataset, a total of 178 respondents from the eight treatment sites reported social or civic participation. Specifically, these individuals claimed membership in local organizations or personal participation in planning, implementation, monitoring or evaluation of local programs or projects. These treated individuals are matched to a subsample of individuals with the same characteristics from three potential comparison groups.

Each of the three comparison groups comprises survey respondents from areas where the GI was not disseminated, because either these are control sites (i.e., where the GI was explicitly kept from public knowledge) or the GI was not yet introduced anywhere (i.e., all 12 pilot during the baseline period). The first comparison group consists of 1,200 individuals included in the baseline survey. The second comparison group consists of 800 individuals from the four control sites that were interviewed during the mid-term and post-pilot evaluation surveys. The 1,200 observation units in the first group and the 800 observation units in the second group are combined together to form the third comparison group. Clearly, the individuals from these three groups were motivated to participate by factors other than the information about local government performance as contained in the GI materials.

(b) *Estimating the propensity scores*

To derive the propensity scores, three probit models are used here. One probit model is applied on an enriched sample consisting of the 1,200 observation units from the first comparison group and the 178 observation units from the treatment sites. The second probit model is estimate using the pooled sample consisting of the 800 observation units from the second comparison group and the 178 observation units from the treated sites. The last probit model is estimated based on the pooled sample that combine the same 178 treated individuals with those in the third comparison group. The descriptive statistics and definitions of the regression variables used are presented in Table 3.

In the specification of the probit model, the explanatory variables included are those that influence the likelihood of treatment. The treatment variable in this case is knowledge of GI, which indicates whether the individual has read a GI *komiks*, saw a GI poster or attended a public presentation of the GI. The regressors included pertain to the respondent's socioeconomic characteristic (log of income, college education, family size, monthly electric bill, employment status, house ownership status) and demographic features (age, household headship, spouse, male). Also included among the regressors are indicators of possible exposure to similar LGU performance measures (other index)[2], of those who reside in the *poblacion* and other densely populated barangays that may have been specifically targeted by local partners during the information campaign. A dummy variable for areas where the incumbent mayors were re-elected in the May 2001 elections are likewise introduced.

Table 3. Variable definitions and descriptive statistics

Variable	Definition	Obs.	Mean	Std. Dev	Min.	Max
Member	1= if member of any local organization in the city or municipality of residence; 0= otherwise	3600	0.249	0.432	0	1
Participation	1=if personally involved in the planning, implementation, monitoring or evaluation of local government projects or programs; 0= otherwise	3600	0.223	0.416	0	1
Knowledge of GI	1=if respondent read a komiks, saw a poster or attended a public presentation about the GOFORDEV Index; 0=otherwise	3600	0.051	0.219	0	1
Other index	1=if aware of the Human Development Index, Minimum Basic Needs, Galing Pook Awards or Clean and Green Awards; 0=otherwise	3600	0.359	0.480	0	1
College	1=if the respondent went to or finished college; 0=otherwise	3600	0.256	0.437	0	1
Age	Age in years of the respondent	3598	41.8	14.8	18	90
Male	1=if the respondent is male; 0=otherwise	3600	0.307	0.461	0	1
Household head	1=if the respondent is the household head; 0=otherwise	3600	0.387	0.487	0	1
Spouse	1=if the respondent is the spouse of the household head; 0=otherwise	3600	0.461	0.499	0	1
Family size	Number of family members	3592	5.19	2.25	0	1
Electric bill	Average monthly electric bill for the last six months (in pesos)	3509	464.85	658..69	0	20000
Regular job	=1 if the respondent has a regular job or a source of income for the past six months; 0=otherwise	3600	0.562	0.496	0	1
Government employee	1=if the respondent is a government employee or worker; 0=otherwise	3600	0.063	0.243	0	1
Income, ln	Natural logarithm of monthly family income	3557	8.61	0.942	4.14	12.43
High density barangay	1=if resident in highly populated barangays (village); 0=otherwise	3600	0.713	0.452	0	1
Owner	1=if the respondent or his/her family is the owner of the house and lot they reside in; 0=otherwise	3600	0.668	0.471	0	1
Married	1=if respondent is married; 0=otherwise	3600	0.849	0.359	0	1
Re-elected Mayor	1=if the current city/municipal mayor was re-elected in the May 2001 local elections; 0=otherwise	3600	0.806	0.396	0	1

Corresponding to each of the enriched samples used, three sets of probit regression results are presented in Table 4. In the first set (Baseline year), the statistically significant variables are other index, high-density barangays, household head, electric bill*spouse, owner, regular job and re-elected mayor. The first five of these variables have negative coefficients, and the last two variables have positive coefficients. In contrast, only two variables are found statistically significant in the second set of regression results (Two-year pilot). One these variables has negative coefficient (high density barangay), while the other has positive coefficient (re-elected mayor). The third set of results (All) is roughly similar to those in the first set. In this case, the statistically significant explanatory factors are high-density barangay, household head, electric bill, electric bill*spouse, owner and re-elected mayor. Likewise, the coefficients of high-density barangay, household head, electric bill*spouse and owner are found negative, while that of electric bill and re-elected mayor are positive. The estimated pseudo R^2 ranges from 0.0848 (baseline year) to 0.2493 (Two-year pilot). In all cases, the χ^2 test statistics indicate that the regressors are simultaneously different from zero. In each of the regressions, the actual number of observations used is less than the maximum because several of the observations had missing data for some of the variables.

Table 4. Probit model of the probability of knowledge of Gofordev Index
Dependent variable: Knowledge of Gofordev Index (GI)

Independent variables	Baseline year		Two-year pilot*		All	
	Coeff.	Robust Std. error	Coeff.	Robust Std. error	Coeff.	Robust Std. error
Other index	-0.228	0.120[c]	0.156	0.125	-0.141	0.109
High-density barangay	-0.480	0.120[a]	-0.767	0.129[a]	-0.519	0.110[a]
Income, ln	-0.065	0.071	-0.083	0.060	-0.067	0.058
College	-0.187	0.133			-0.119	0.119
Age	-0.023	0.023	-0.0003	0.022	-0.015	0.019
Age*age	0.0003	0.0002	-0.000	0.0002	-0.0002	0.0002
Male	0.075	0.179	0.241	0.158	0.136	0.148
Household head	-0.453	0.232[c]	-0.241	0.221	-0.386	0.202[c]
Spouse	-0.133	0.256	0.352	0.249	0.011	0.221
Family size	0.005	0.024	0.028	0.030	0.013	0.023
Electric bill	0.0002	0.0001	0.0001	0.0001	0.0002	0.000[c]
Electric bill*spouse	-0.0005	0.0003[c]	-0.0004	0.0002	-0.0005	0.0002[c]
Owner	-0.359	0.125[a]	-0.086	0.133	-0.261	0.114[b]
Married	-0.059	0.166	-0.242	0.175	-0.127	0.142
Government employee	0.281	0.198	0.205	0.203	0.249	0.179
Regular job	0.228	0.124[c]	0.073	0.133	0.177	0.112
Re-elected Mayor	0.449	0.117[a]	1.316	0.118[a]	0.673	0.102[a]
Constant	0.320	0.779	-0.476	0.706	-0.350	0.665
Pseudo R^2	0.0848		0.2493		0.1039	
Wald χ^2	52.04		193.47		79.45	
Prob>χ^2	0.000		0.000		0.000	
No. of observations	1329		954		2105	

Notes:
The results are obtained using sampling weights.
[a] significant at p<0.01, [b] significant at p<0.05, [c] significant at p<0.10.
*The variable college is dropped to satisfy the balancing property in the propensity score estimation.

While the probit models will be used primarily to estimate the propensity scores (conditional probabilities), they show some interesting results. In general, the evidence suggest that those with relatively high socioeconomic status (house owners, high electric bills) or household heads (who might be busy or away from home during day-time when GI materials are distributed) are less likely to be aware of the GI. Also, those residing in far-flung barangays are more likely to be aware of GI than those who live in the *poblaciones*, a result which partly attests to effort made by the local partners to disseminate GI information as widely as possible. There is also some evidence that interest in the performance of the LGUs tend to be higher in areas with re-elected mayors. This particular result perhaps indicates that the residents in these areas desired to validate their election choices with the GI scores.

(c) *Matching of treated with control subsamples*

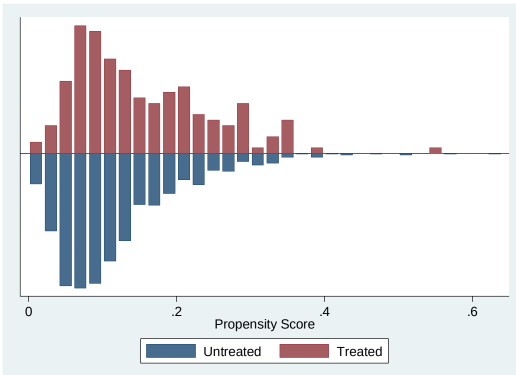
The probit regression models yield the propensity scores used to match the 178 treated individuals with subsample units from the comparison groups. For the matching to be valid, the treated individuals and matched control subsamples should have the same propensity scores along a common support. Further, the matched groups should also have the same observable characteristics (also called the balancing requirement). To check if the balancing requirement is satisfied, three methods are employed here, namely: visual inspection of the histograms of the treated and untreated observations (matched control subsamples), an analysis of the bias distribution, and the test of equality of means of the propensity scores of the treated and untreated individuals.

Since propensity scores are continuous variables, matching based on them is nearly impossible. To get around this problem, the treated individuals are matched with control sub-samples with the nearest propensity scores following an algorithm that set bounds on the acceptable differences between their respective propensity scores. The matching algorithms used here are nearest neighbor (one-to-one, random draw, and equal weights), kernel (with bandwidth of 0.06), radius (with length of 0.01) and stratification.[3] In all the matching algorithms used here, matching along a common support is imposed. (A useful discussion of the relative advantages of these matching algorithms is found in Caliendo and Kopeinig, 2008).

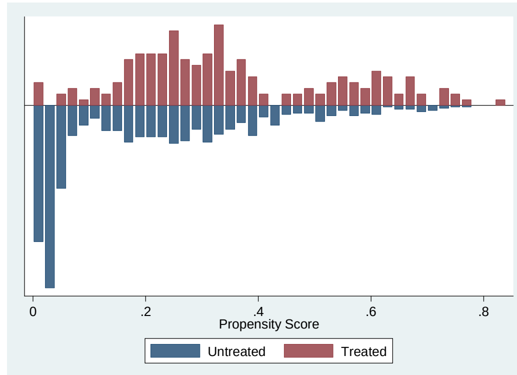
As a first check if the regression samples satisfy the balancing requirement, the histograms of the treated and untreated units are obtained. The histograms are presented in Figure 1. In panel (a) and panel (c), the bulk of the treated and untreated (matched control sub-samples) samples are found in the lower range of propensity scores (between 0 and 0.2). It can also be seen in panel (b) that the distribution of treated and untreated observations at the extreme ranges of the propensity scores are more varied than in either panel (a) or panel (c). At the lower range (0-0.2), there are more untreated than treated observations; while at the upper range (0.6-0.8), there are slightly more treated than untreated observations. In contrast, the distributions observations in the two other panels appear to be nearly unimodal. The histograms indicate that the treated individuals may not be exactly like the control sub-samples obtained from the mid-term and post-pilot surveys. But the differences appear to be insignificant when the treated individuals are compared with the pooled sub-samples from all three survey rounds.

Figure 1. Frequency distributions of the treated and matched control subsamples along common support

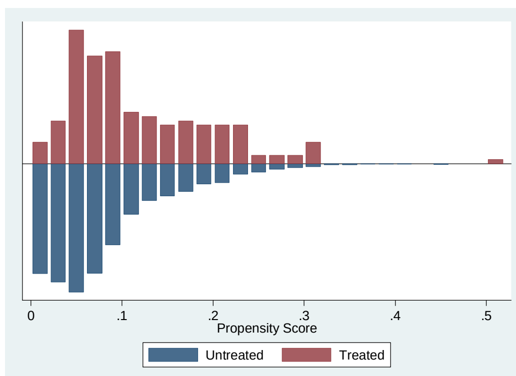
(a) Baseline year



(b) Two-year pilot



(c) All



A further test of the balancing requirement is to check whether the distance in the marginal distributions of the covariates (X) improves after matching. A measure of this distance is called standardized bias, which is defined for each covariate X as the "difference of sample means in the treated and matched control subsamples as a percentage of the square root of the average sample variances in both groups" (Caliendo and Kopeinig, 2008, p.48). The distributions of the standardized bias achieved before and after matching using the nearest one-to-one neighbor (caliper) algorithm is presented in Table 5.[4]

Table 5. Distribution of the standardized bias before and after matching

	Baseline year		Two-year pilot		All	
	Unmatched	Matched	Unmatched	Matched	Unmatched	Matched
Mean bias	13.08	10.17	18.93	12.55	14.05	14.46
Median bias	13.08	11.12	18.93	12.49	9.54	13.06
Minimum bias	2.90	0.319	1.29	1.174	0.58	0.95
Maximum bias	40.82	19.08	85.69	25.22	48.97	33.28
Pseudo R-squared	0.044	0.029	0.122	0.052	0.053	0.055
Chi-square likelihood ratio	46.33	14.19	112.10	25.69	64.56	27.24
P-value	0.00	0.511	0.00	0.041	0.00	0.027
Bias on propensity score	40.80	0.30	85.70	1.9	49	0.9
Bias reduction in propensity score (in percent)	99.2		97.8		98.1	

Note: Estimates of bias based on weighted samples.

In general, the mean, median, minimum and maximum biases improve when the treated units are matched with the subsamples from the baseline survey or with those from the mid-term and post-pilot surveys. There is however some increase in the bias after matching using the pooled matched control subsamples. Caliendo and Kopeinig (2008) further suggest looking at the pseudo-R^2 and perform a likelihood ratio tests on the joint significance of all regressors in the probit model to check overall covariate balance. As can be seen in Table 5, the estimates of the pseudo-R^2 are reduced when the treated individuals are paired off with those in the baseline surveys or with those in the two impact assessment surveys. The low estimates of the pseudo-R^2 suggest no systematic differences in the distribution of covariates between the treated and the matched control groups. When the same treated individuals are matched with those in the pooled subsamples (All), the pseudo-R^2 increased a little bit. Nonetheless, the p-values of the likelihood ratio tests suggest increasing joint insignificance of the regressors after matching. Last, the matching also lead to significant reduction in the bias of the propensity score matching, ranging from 97.8 percent to 99.2 percent.

Another test of the balancing requirement performed here is a *t*-test of the equality of means of the propensity scores of the treated group and the untreated subsamples. This particular test is automatically performed in the STATA program used here (PSCORE).[5] On the whole, the three sets of test results on the balancing requirement indicate that the treated individuals are paired off with control groups that have fairly the same observable characteristics. Once this particular requirement is satisfied, it is now possible to estimate the average treatment effects on the treated.

6. Estimates of the average treatment effects

This section presents the estimates of the average treatment effects on the treated. These are approximated in terms of increase in the probability of those exposed to GI materials or presentations in joining or participating in planning, implementation, monitoring or evaluation of local programs or projects. The incremental probabilities are positive and substantial. They are also fairly robust to biases that may be due to unobserved factors.

(a) *Increase in the probability of civic engagement*

Having satisfied the balancing requirement, it is now possible to apply (2') to obtain the average treatment effects on the treated. In this case, the difference between the two mean outcomes is the increase in the probability of civic engagement due to exposure to GI materials or presentations. The estimated increases in probability of membership in local organizations or participation in civic activities (planning, implementation, monitoring or evaluation of local program or projects) are presented in Table 6.

It can be seen from the table that the increase in the likelihood of membership in local organizations due to exposure to GI could be substantial. The increments are all positive and range from 0.264 to 0.404. Each estimate is also statistically significant, irrespective of the matched control subsamples used or type of matching algorithm used. The estimated average treatment effects are generally higher the more control subsamples are matched with the treated units. This is expected since the difference in outcomes between the treated unit and the matched subsample is "magnified" (as it were) by the number of comparison units.

Basically the same qualitative results are obtained for the other outcome variable. The incremental probability of participation in civic activities is positive and statistically significant, regardless of the type of matching algorithm or comparison group used. In this case, the increases in probabilities range from 0.331 to 0.375. On the whole, the estimates of the average treatment effects on the treated indicate that information about local government performance, such as those contained in GI materials or public presentations, could motivate local residents to become more active citizens.

Table 6. Estimated average treatment effects on the treated: Impact of Gofordev Index on membership and participation.

Matching method/ Control group	Number of samples: Treated	Number of samples: Non-treated	Average treatment effects on the treated	Bootstrapped Standard Error	*t*-statistics
Member in local organization					
A. Baseline year					
Nearest 1-to-1, caliper 0.1	178	178	0.281	0.049	5.680
Nearest neighbor, random draw	178	149	0.281	0.058	4.812
Nearest neighbor, equal weights	178	149	0.281	0.051	5.486
Kernel, bandwidth of 0.06	178	1132	0.271	0.038	7.189
Radius, radius of 0.01	177	1122	0.264	0.039	6.822
Stratification	178	1132	0.273	0.039	7.082
B. Two-year pilot*					
Nearest 1-to-1, caliper 0.1	178	178	0.393	0.045	8.700
Nearest neighbor, random draw	178	138	0.354	0.048	7.434
Nearest neighbor, equal weights	178	138	0.354	0.052	6.815
Kernel, bandwidth of 0.06	178	759	0.399	0.039	10.309
Radius, radius of 0.01	177	669	0.404	0.038	10.517
Stratification	178	759	0.381	0.041	9.354
C. All					
Nearest 1-to-1, caliper 0.1	178	178	0.303	0.049	6.220
Nearest neighbor, random draw	178	161	0.309	0.049	6.369
Nearest neighbor, equal weights	178	161	0.309	0.060	5.114
Kernel, bandwidth of 0.06	178	1912	0.320	0.036	8.835
Radius, radius of 0.01	177	1902	0.317	0.036	8.711
Stratification	177	1913	0.319	0.037	8.709
Participation					
A. Baseline year					
Nearest 1-to-1, caliper 0.1	178	178	0.337	0.047	7.180
Nearest neighbor, random draw	178	149	0.331	0.047	7.076
Nearest neighbor, equal weights	178	149	0.331	0.051	6.458
Kernel, bandwidth of 0.06	178	1132	0.338	0.037	9.062
Radius, radius of 0.01	177	1122	0.331	0.038	8.770
Stratification	178	1132	0.341	0.038	8.935
B. Two-year pilot*					
Nearest 1-to-1, caliper 0.1	178	178	0.354	0.046	7.650
Nearest neighbor, random draw	178	138	0.331	0.057	5.798
Nearest neighbor, equal weights	178	138	0.331	0.053	6.208
Kernel, bandwidth of 0.06	178	759	0.371	0.042	8.825
Radius, radius of 0.01	177	669	0.375	0.040	9.454
Stratification	178	759	0.374	0.041	9.120
C. All					
Nearest 1-to-1, caliper 0.1	178	178	0.331	0.047	7.030
Nearest neighbor, random draw	178	161	0.343	0.046	7.475
Nearest neighbor, equal weights	178	161	0.343	0.053	6.510
Kernel, bandwidth of 0.06	178	1912	0.349	0.037	9.480
Radius, radius of 0.01	177	1902	0.346	0.036	9.645
Stratification	177	1913	0.345	0.037	9.275

Notes:
The average treatment effect on the treated is the difference between the mean outcomes of the treated samples and the matched controls. Estimates are based on weighted samples.
*The variable college is dropped in the propensity score estimation to satisfy balancing property.
The bootstrapped standard errors are estimated using 100 replication samples.
Nearest 1-to-1 matching algorithm is done without replacement and is derived from Leuven and Sianesi (2003) using psmatch2 while nearest neighbor, kernel, radius and stratification algorithms are from Becker and Ichino (2002).

(b) *Sensitivity to unobserved characteristics*

Recall that the validity of the propensity score matching as an evaluation technique rests on the assumption that all factors that could influence exposure to GI are all accounted for in the estimation of the propensity scores. Besides the observable characteristics, however, unobserved ones could also influence exposure to GI. While the selection bias due to observable characteristics can be minimized by satisfying the balancing requirement, the bias due to unobservable factors however cannot be directly tested. An indirect test is applied here to assess the sensitivity of the estimated average treatment effects.

Proposed by Becker and Caliendo (2007), the Mantel and Haenszel (MH) test statistic is a statistical test of the effect of the possible unobserved variable on the odds ratio of being a participant and non-participant in the treatment. When there is no hidden bias (due to the unobserved variable), the odds ratio (represented by Γ) is one when matching is conditioned on observed covariates. With hidden bias, the odds ratio could increase or decrease depending on the nature of the selection. The MH test statistic would indicate whether the change in the odds ratio is significant enough, and therefore the selection bias due to the unobservable factors is substantial enough, to undermine the results of the initial matching analysis (performed under the null hypothesis of no hidden bias).

The results of the MH tests contained in Table 7 are presented in terms of Mantel-Haenszel bounds on the estimated Γ. The figures indicate the critical level of the odds the ratio obtained with an imputed hidden bias big enough to reject the null (of no hidden bias) at 5-percent significance level. Under the column member, the figures in the first row (and third row) indicate that whatever bias that could arise from possible missing covariate has lead to more than doubling of the ratio of the odds of being member and not being a member to reject the initial hypothesis that such bias does not exist. In the second row, where the comparison units are derived from the two impact assessment surveys, the hidden bias should lead at least lead to quadrupling of the odds ratio to invalidate the initial estimates of the average treatment effects on membership status.

Table 7. Critical level of hidden bias (Mantel-Haenszel bounds)

Control group	Member	Participation
Baseline year	2.3 – 2.4	3.2 – 3.3
Two-year pilot	4.6 – 4.7	3.6 – 3.7
All	2.6 - 2.7	3.1 – 3.2

Notes:
Estimates derived using weighted samples.
The figures are Mantel-Haenszel bounds at 5-percent significance level.
Sensitivity tests made using the nearest 1-to-1 matching algorithm without replacement.

Based on the same tests, the average treatment effects on participation status are likewise found to be robust to selection bias due to unobserved characteristics. The MH bounds are all around 3, which mean that the selection bias should triple the odds ratio in favor of participation to undermine the null hypothesis (of no hidden bias). Thus, the overall results of the MH tests that the estimates of the average treatment effects are fairly robust to unobserved characteristics.

7. Concluding remarks

In summary, the estimated average treatment effects on the treated obtained through propensity score matching technique are significant and fairly robust to the choice of comparison groups, matching algorithm and to possible bias induced by unobserved characteristics. The estimated increase in the probabilities of joining a local organization or participating in civic activities among those exposed to GI materials or public presentations, where assessments of local government performance were presented, are positive and substantial, ranging from 28 to 40 percentage points.

The overall findings therefore affirm the critical value to residents of providing them relevant information about their local governments. The particular set of information contained in the GI materials pertain to general assessment of all residents of their local government's achievements in terms of providing for their needs for basic public services and the extent to which local officials confer with the general population. The GI materials also include information on actual fiscal outlays for basic services, which would suggest the relative priorities of the local governments. When all these information are communicated using the local vernacular, they also serve to educate the readers or listeners. In the process, the informed individual gains a better understanding or their local governments. With information about the extent to which others share her views and sentiments, she is able to assess better her chances of influencing local public decisions. Her chances of course improve by being more active in civic affairs.

In closing, the findings lend support to the claim that greater transparency in governance can lead to greater accountability since at least some of the informed citizens will undertake the appropriate action. Promoting transparency and accountability is particularly important under decentralization when local governments are expected to provide frontline services. To make the LGUs effective and efficient service providers, an objective and well-disseminated performance benchmarking mechanisms should be adopted.

Notes

1. The same sampling design was employed in both the impact assessment surveys and the household surveys conducted to generate the GI scores. Further, the questions in the four-page household questionnaire used for the GI results were also asked in the seven-page interview questionnaire used in the impact assessment surveys. The latter questionnaire had additional questions on participation, satisfaction with the performance of local officials, trust of local officials and others. The impact assessment surveys were also undertaken with local academic institutions (but different from those contracted as local partners).

2. The other performance measures or indicator systems that preceded the GOFORDEV Index in the pilot areas are the Minimum Basic Needs, Clean and Green Awards, Galing Pook Awards and Human Development Index. The first two are promoted by the national government, and the last two by two national civil society organization supported by donor agencies.

3. To achieve matching and satisfy the balancing requirement, two STATA programs are used, namely: PSCORE (Becker and Ichino, 2002) and PSMATCH2 (Leuven and Sianesi, 2003). The PSCORE program is able to perform five types of matching algorithms (nearest neighbor with random draw or equal weights, kernel, radius and stratification). It also automatically checks if the balancing requirement is satisfied by performing a test of the equality means on the propensity scores of the treated units and the matched control sub-samples. The PSMATCH2 program performs the nearest neighbor one-to-one matching algorithm. It also checks for the bias distribution before and after matching to see if the balancing requirement is satisfied. In the implementation of PSMATCH2 here, only the regression samples that satisfied the test of equality of means in the PSCORE are used. Moreover, in both implementations of the PSCORE and PSMATCH2 programs, a matching along common support is imposed.

4. The standardized bias is computed for each of the covariates found in the three probit models. In general, a reduction in the standardized bias is observed after matching. Detailed results available from the authors.

5. The PSCORE first divides the treated and matched control subsamples into blocks (based on the range of propensity scores) and performs the *t*-test. If the test is failed for a particular block, it is subdivided intro smaller blocks on which the test is performed again. This procedure is iterated until the test is passed with the minimum number of blocks. The number of blocks used in the tests ranges from 4 to 7. Detailed results of the t-tests are available from the authors.

References

Agrawal, A. and Gupta, K. (2005) Decentralization and Participation: The Governance of Common Pool Resources in Nepal's Terai. *World Development*, 33(7), pp. 1101-1114.

Alatas, V., Pritchett, L. and Wetterberg, A. (2002) Voice lessons: local government organizations, social organizations and the quality of local governance, World Bank Policy Research Working Paper, 2981, The World Bank, Washington, D.C.

Alesina, A. and La Ferrara, E. (2000) Participation in heterogeneous communities. *The Quarterly Journal of Economics*, 115(3), pp. 847-904.

Bardhan, P. and Mookherjee, D. (2002) Decentralization and accountability in infrastructure delivery in developing countries. *Economic Journal*, 116(508), pp. 101-127.

Blair, H. (2000) Participation and accountability at the periphery: democratic local governance in six countries. *World Development*, 28 (1), pp. 21-39.

Becker, S. O and Caliendo, M. (2007). Mhbounds- Sensitivity analysis for average treatment effects. Institute for the Study of Labor Discussion Paper No. 2542, Bonn, Germany.

Becker, S. O. and Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *The Stata Journal* 2 (4), pp. 358-377.

Best M. and Dustan L. (2006) The civic participation of visible minority Canadians: a literature review, in: *For the communities within: diversity and exclusion in Ottawa project*. Social Planning Council of Ottawa, Canada.

Blackburn, J and De Toma, C. (1998) Scaling-down as the key to scaling up? The role of participatory municipal planning in Bolivia's Law of Popular Participation, in: J.Blackburn and J.Holland (eds) *Who Changes? Institutionalizing Participation in Development*, (London: Intermediate Technology Development Group (ITDG) Publishing), pp. 103-105.

Callahan, R., Wilkinson, L. and Muller, C. (2008) School context and the effect of ESL placement on Mexican-origin adolescent's achievement. *Social Science Quarterly* 89(1), pp. 177-198.

Caliendo, M. and Kopeinig, S. (2008) Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys* 22(1), pp.31-72.

Campos, J.E. and Hellman, J.S. (2005) Governance Gone Local: Does Decentralization Improve Accountability?, in: R. White and P. Smoke (eds) *East Asia Decentralizes: making Local Government Work* (Washington, DC: The World Bank), pp. 237-252.

Capuno, J. J. (2007) The quality of local governance and development under decentralization, in: A. M. Balisacan and H. Hill (eds) *The Dynamics of Regional Development: The Philippines in East Asia*, (Chetelham, Gloucester: Edward Elgar Publishing Limited), pp. 204-244.

Dehejia, R. H. and Wahba, S. (2002**)** Propensity Score-Matching Methods for Nonexperimental Causal Studies. *Review of Economics and Statistics* 84(1), pp.151-161.

Department of the Interior and Local Government (2002) *The State of Philippine Cities: A Self-Assessment by the Cities based on the Local Productivity and Performance Management System*. Quezon City, Philippines: Department of the Interior and Local Government.

Gaspart, F., Jabbar, M., Mélard, C., and Platteau, J.P. (1998) Participation in the construction of a local public good with indivisibilities: an application to watershed development in Ethiopia. *Journal of African Economics*, 7(2), pp. 157-184.

Goetz, A.M. and Gaventa, J. (2001). Bringing Citizen Voice and Client Focus into Service Delivery. IDS Working Paper, 138, Institute of Development Studies, University of Sussex.

Heckman, J. J., Ichimura, H. and Todd, P. E. (1997) Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies* 64(4), pp. 605-654.

Heckman, J. J. and Smith, J. A. (1995) Assessing the case for social experiments. *Journal of Economic Perspectives*, 9(2), pp. 85-110.

Human Development Network (2002) *Philippine Human Development Report 2002*.Manila: Human Development Network and United Nations Development Programme.

Isham, J., Narayan, D. and Pritchett L. (1995) Does participation improve project performance? Establishing causality with subjective data. *The World Bank Economic Review*, 9(2), pp. 175-200.

Jennings, R. (2000). Participatory development as new paradigm: the transition of development professionalism, in*: Conference on Community based reintegration and rehabilitation in post-conflict settings*. Washington, DC

Knoke, D. (1988) Incentives in collective action organizations. *American Sociological Review*, 53(3), pp. 311-329.

Leuven, E. and Sianesi, B. (2003) PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing,, http://ideas.repec.org/c/boc/bocode/s432001.html.

Lowndes V., Pratchett L., and Stoker G. (2001a) Trends in public participation: part 1-local government perspectives. *Public Administration*, 79 (1), pp. 205–222.

Lowndes V., Pratchett L., and Stoker G. (2001b) Trends in public participation: part 2 - citizens' perspectives. *Public Administration,* 79 (2), 445–455.

Moe, T. M. (1980) A calculus of group membership. *American Journal of Political Science*, 24(4), pp. 593-632.

National Statistical Coordination Board (2004) Philippine Statistical Yearbook 2004. Makati City, Philippines: National Statistical Coordination Board.

Olson, M. (1965) The *Logic of Collective Action* (Cambridge: Harvard University Press).

Ohmer, M. (2007) Citizen participation in neighborhood organizations and its relationship to volunteers' self- and collective efficacy and sense of community. *Social Work Research,* 31(2), pp. 109-120.

Organization for Social Science Research in Eastern and Southern Africa - Rwanda Chapter. (2006) Citizen Report and Community Score Cards. Organization for Social Science Research in Eastern and Southern Africa.

Pierre, J. (1998) Public consultation and citizen participation, in: B. Peters and D. Savoie (eds) *Taking Stock*, (London: McGill-Queens University Press), pp. 137 – 186.

Platteau, J.P. (2000) *Institutions, Social norms, and Economic Development* (London: Harwood Academic Publishers).

Platteau, J.P. and Gaspart, F. (2003) The Risk of Resource Misappropriation in Community-Driven Development. *World Development*, 31(10), pp. 1687-1703.

Putnam, R. D. (1993). *Making Democracy Work: Civic Traditions in Northern Italy.* Princeton, NJ: Princeton University Press.

Rosenbaum, P. and Rubin, D.B. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, pp. 41-45.

Rosenbaum, P. and Rubin, D. B. (1984) Reducing the bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* 79, pp. 516-24.

Schugurensky, D. (2003) Civic participation: on active citizenship, social capital and public policy. *Canadian Diversity*, 2(1), pp. 10-12.

Trujillo, A. J., Portillo, J. E. and Vernon, J. A. (2005) The impact of subsidized health insurance for the poor: Evaluating the Colombian experience using propensity score matching. *International Journal of Health Care Finance and Economics* 5, pp. 211-239.

Weinberg, K. and Jütting, J.P. (2001) Women's participation in local organizations: conditions and constraints. *World Development*, 29(8), pp. 1391-1404.

World Bank (2001) *Philippines: Filipino Report Card on Pro-poor Services*, Washington, DC.