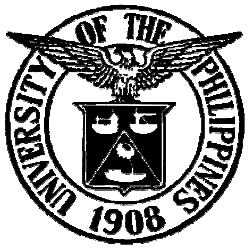


UP School of Economics
Discussion Papers

Discussion Paper No. 2009-10 November 2009

Earning trust with transparency: Performance ratings and trust in local officials in the Philippines

by
JOSEPH J. CAPUNO* and MARIA MELODY S. GARCIA**

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

Earning trust with transparency: Performance ratings and trust in local officials in the Philippines

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

Abstract

As a component of social capital, trust is deemed critical to growth and thus must be cultivated. This paper presents evidence that transparency – the public announcement of local government performance ratings– results in greater trust in local officials. The results are based on the application of propensity score matching techniques on three rounds of household survey dataset collected in a local governance project piloted in 12 local governments in the Phillippines in 2001-2003. With the public announcements, the trust may fall initially and then rise afterwards, suggesting that transparency can earn trust through time. Some policy implications are drawn.

JEL Classifications: H11, H30, C93
Keywords: Local governments, transparency, performance ratings, trust, Asia, Philippines

1. Introduction

Can transparency – or, in this case, the public announcement of local government performance ratings – result in greater trust in local officials? By answering this question, this paper contributes to the debate concerning the causes and consequences of trust, which together with the other components of social capital has been found important to economic growth and development (e.g., Knack and Keefer, 1997; Casey, 2004; Whiteley, 2007). With trust, people cooperate and undertake collective actions not otherwise possible among strangers or in infrequent social interactions. In a local context, the people who learn to trust need not invest in activities and mechanisms to monitor their local governments. In turn, the trusted officials can then perform without unnecessary encumbrances to achieve the social objectives. Thus, with trust, the costs of monitoring are minimized and more resources can be devoted to directly enhance social welfare.

To promote social welfare, transparency in governance is likewise advocated. It is found that informed voters make better choices about their government leaders or get directly involved in civic affairs (Paul, 1998; Ravindra, 2004). With active constituents, government officials become more accountable, and, can be induced to make better decisions (Jenkins and Goetz, 1999; Alatas et al., 2002). Understandably, however, office holders may be reluctant to divulge information that would rigthly or wrongly incriminate them. Voters may misunderstand the information or may be misled by political rivals who misuse the information. In the political agency model of Besley (2006), for example, trasparency can lead to lower welfare in the short run when a“bad” incumbent politician decides to abuse his position after being outed, rather than behave as a “good “politician

when his chances to another term in officer are effectively quashed . Hence, transparency can have adverse short-term effects.

In real setting, a typical term of office could present some opportunities for political agents to prove their worth and their principals to learn about it. Even when the incumbent faces a term limit, he may still want to win the support and trust of his constituents if his spouse, child or relative can run for the same office. Hence, the political agent and his principals, as it were, are playing a repeated game, from which mutual trust may evolve. Such possiblity may arise as the principals gain more information that enables them to assess more objectively the agent's performance. The need for such information is presumably more acute under a decentralized setting, where local officials are more directly accountable to their constituents.

With nearly twenty years of decentralization in the Philippines, the overall level of trust in local officials is lot less known than that in national officials, which has been regularly monitored since 1986 (e.g., Social Weather Stations, 2009). In two recent papers on Philippine communities, social capital, including trust, is shown to emanate from participation in village assemblies and frequency of meetings between officials brought about by community-driven deelopment projects (Labonne and Chase, 2008), but trust can also be hindered by existing inequalities in the project areas (Labonne, Biller and Chase 2007). This paper aims to contribute to the better understanding of evolution of trust in local officials and to recommend policies to cultivate it the Philippines and, possibly other developing countries.

The issue of whether information about local government performance can result in greater trust in local officials is investigated here using a unique dataset from a local

governance project in the Philippines. In this project, a local government performance index was introduced in 2001-2003 in twelve sites, eight of which were designated as treatment sites and the rest as control sites. To assess the impact of the index on the level of trust in officials, responsiveness of officials and level of civic participation, three rounds of household surveys were conducted. Applying propensity score matching on the household level data, the results thus provide evidence that trust could be earned, but not easily.

The local governance project is briefly described section 2, and then followed by a presentation in section 3 of the study's evaluation framework. The implementation of the propensity score matching technique is discussed in section 4. In section 5, the estimates of the impact on trust are presented, and then analyzed further in section 6. The conclusion and policy implication are given in the last section.

2. Good Governance and Local Development Project[1]

The evidence presented in this paper is based on the dataset generated under the Good Governance and Local Development (GGLD) Project of the Philippine Center for Policy Studies, a non-government organization. The aims of the GGLD Project were to develop and advocate the institutionalization of a set of performance indicators pertaining to good local governance. First formulated in 2000, the Governance for Local Development Index (GI) was piloted to ascertain thepossible influence of its public announcment on the citizens' trust in their local officials, their perception of the responsiveness of the local officials, and their level of civic participation.

The pilot test was conducted for over two years (2001-2003) in 12 component local government units (LGUs) of the provinces of Bulacan and Davao del Norte. Both the two provinces and the 12 pilot sites within each of them were randomly selected from clusters of highly developed and less developed areas to control for differences in levels of socioeconomic conditions and geographic factors. Bulacan and Davao del Norte were the samples drawn from the developed and less developed groups of provinces, respectively. Just north of Manila, Bulacan had a better than national average rating in the Human Development Index (HDI) and incidence of poverty in 2000. Located in the southern island of Mindanao, Davao del Norte had lower than national average ratings in both indicators for the same year (Human Development Network, 2002; National Statistical Coordination Board, 2004).

Within each province, a total six LGUs are randomly chosen, three each from the clusters of highly developed and less developed component cities and municipalities. Four of the six sites are assigned treatment sites and the rest as control sites. In Bulacan, the four treatment sites are San Jose del Monte City and the municipalities of Angat, Baliwag and Guiguinto, and the two control sites are Bustos and Plaridel. In Davao del Norte, the treatment sites are the Island Garden City of Samal, Panabo City and the towns of Braulio Dujali and Sto. Tomas, and the control sites are Tagum City and Asuncion.

The pilot sites have both similar and contrasting features. In 2001, San Jose del Monte City and the Island Garden City of Samal were newly chartered cities, while Braulio E. Dujali was a newly created municipality. Of the 12 LGU mayors at the start of the pilot test, nine were re-elected and three (Sto. Tomas, Tagum City and Bustos) were newly elected in the May 2001 local elections. The residents in the 12 areas are

predominantly Christians, although Muslims and indigenous tribes can be found in Davao Del Norte. The average fiscal revenues per capita in 2000 of the pilot LGUs in Davao del Norte were about 1,442 pesos, only about half of this was the corresposning average of the pilot LGUs in Bulacan. The relative wealth of the pilot LGUs is Davao del Norte is only nomimal: their fiscal status is due more to their higher shares in national government revenues, which are based on land area and population size, than on the dynamism of their local economies. In 2003, the average official poverty incidence rates of the pilot sites in Davao del Norte and Bulacan were 31.7 and 7.5, respectively.

There were two major activities undertaken in the pilot sites. The first major activity was the generation of the GI scores, which was accomplished in all 12 sites. The second major activity was the dissemination of the scores, which was implemented only in the eight treatment sites. These pilot activities were carried out by local partners who were contracted and provided training and logistic support. 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 (Table 1).

[Insert Table 1 here.]

With the GI scores, LGU performance is assessed along three 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

barangay (village) –level consultations and functioning of local consultative bodies. Ranging from zero (lowest) to 100 (highest), the scores in the GI indicators were calculated based on household surveys, and on official audited financial reports and the minutes of the meetings of the local consultative bodies.

Over the two-year pilot period, most of the sites experienced a decline in their GI scores, except in the two treatment sites (Guiguinto and San Jose del Monte City) and one control site (Tagum City) (Table 2). The biggest percentage drop was in Asuncion (31.1), Angat (29.3) and Sto. Tomas (22.7). The scores were announced in the treatment sites through public presentations, which were conducted at least three times a year in each treatment site. In 2002, however, an additional forum was held exclusively for key local officials. In 2001, the total numbers of participants in the public fora were 496 and 428 in Bulacan and Davao del Norte, respectively. In the following year, the corresponding totals were 565 and 596. At least 15 percent of the forum participants in each year worked for the government.

[Insert Table 2 here.]

Further, the local partners distributed posters, stickers and *komiks*, a popular reading fare using comic strips translated into Tagalog for the Bulacan areas and Bisaya for the Davao del Norte areas (Table 3). They distributed GI materials in public places like government offices, transport terminals, stores, and marketplaces. Several of them also sent the materials to households. In these dissemination activities, the local partners possibly targeted certain groups or individuals active in civic affairs. As likely as well, some people may have purposely sought out GI materials or joined GI presentations because of their work or interests. In other words, the treatment of individuals (or

exposure to GI materials) was less than random. It is therefore important to account for the selection bias in assessing the influence of the GI ratings on the citizen's declared trust in their officials

[Insert Table 3 here.]

3. Evaluation framework

Methodology

To control for the selection bias, an evaluation method commonly used in observational studies is adopted here. Using this method, inference about the treatment impact is based on the difference in the mean outcomes with and without treatment for those who actually participated in the treatment or program (Caliendo and Kopeinig, 2008). To obtain the two mean outcomes, each treated individual is matched with an untreated individual with the same pre-treatment characteristics. The difference between the two mean outcomes yields the so-called average treatment effects on the treated, which minimizes the bias in the estimation of treatment impactsince the evaluation is limited to those who were actually treated, and whose characteristics are now taken into account (Heckman et al. 1997; Wooldridge, 2002)..

Following Heckman et al. (1997) and Becker and Ichino (2002), let the outcome for the *i*th person be Y_{i1} and Y_{i0} in the treated and untreated states, respectively. Let T_i =1 if the *i*th person received treatment and T_i =0 if not. The treatment effect for *i*th person is given by $\tau_i = Y_{i1} - Y_{i0}$ and the average treatment effect for the treated is defined as

$$\tau_{ATT} \equiv E(Y_{i1} - Y_{i0} | X_i, T_i = 1) \qquad (1),$$

where X is a vector of pre-treatment characteristics. In practice, however, the τ_{ATT} as specified in (1) cannot be directly estimated since for any person and at any given time only one state can be observed (Y_1 or Y_0) and not both. In theory, $E(Y_{i0}|X_i, T_i=0)$ can be used as an unbiased estimator for $E(Y_{i0}|X_i, T_i=1)$ under two conditions. First, treated individuals are matched with untreated individuals of the same pre-treatment characteristics. Second, after conditioning on X, the outcomes for the treated and comparison units should be independent of treatment status.

But since matching on a multi-dimensional vector X can be tedious, matching on propensity scores based on X is often used. As defined by Rosenbaum and Rubin (1983), the propensity score is the conditional probability of receiving treatment given X, i.e.,

$$p(X) \equiv Pr(T = 1|X) = E(T|X) \qquad (2).$$

The validity and ease of matching based on propensit scores is based on the following important property (Rosenbaum and Rubin, 1984):

$$Y_{i1}, Y_{i0} \perp T_i|X_i \Rightarrow Y_{i1}, Y_{i0} \perp T_i|p(X_i) \quad \forall i \qquad (3).$$

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 proposensity scores are scalars, matching can be done by pairing off treated units with control units with the same or very near scores. However, in constructing the propensity scores, it is crucial that all characteristics that could influence treatment assignment and potential outcomes simultaneously are accounted for in X. Also, matching on propensity scores should be done along common support, which is the intersection of the conditional probabilities of

the treated individuals and matched comparison units. This condition assures that each treated individual, as it were, would have the chance of not being treated.

An implication of the balancing of pre-treatment characteristics and (3) is that τ_{ATT} can now be estimated as

$$\tau_{ATT} \equiv E(Y_{i1}|p(X_i), T_i = 1) - E(Y_{i0}|p(X_i), T_i = 0) \qquad (4)$$

In practice, there are four basic matching algorithms used. In the nearest-neighbor matching, the treated unit is matched with comparator unit with the nearest propensity score. In caliper and radius matching, a tolerance level (say, 0.1) is set on the distance in propensity score to ensure that the treated unit is not matched to a closest neighbor with a very different propensity score. In kernel matching, a weighted of all matched control units is used to establish the counterfactual. Finally, in stratification matching, the common support of the propensity scores are partition into strata (say, five), over which the mean differences in outcomes between the treated and matched control units are computed. Since each basic matching algorithm involves a tradeoff between bias and efficiency, two or more or them are usually implemented to obtain a more robust estimate of τ_{ATT} (Caliendo and Kopeinig, 2008).

Data

The data used in this paper are culled from the three rounds of random household surveys conducted in the 12 pilot sites to assess the impact of GI on the degree of trust in local officials, among others. The baseline survey was conducted in April-May 2001, the second one in February- March 2002, and the last survey was in February-March 2003. In between these three surveys, the local partners undertook their assigned activities, first in June-August 2001, and then again in March-September 2002.

A uniform sampling design and instruments were used in collecting household data in all sites during the pilot period to ensure that the treated samples and the comparison units have at least the same observable covariates (Heckman and Smith 1995). In particular, 100 household respondents per site were randomly selected in each survey round. The sampling weights were calculated based on household sizes and age distribution of the local population. The interview schedule was design to elicit information on household-level socioeconomic and demographic characteristics, knowledge of the GI, trust in local officials, civic participation, and satisfaction with local government performance.

The impact variable used here is based on a trust rating that the respondents gave their local elected officials. Each survey respondent was asked to rate their trust in these officials from one to five, with one as the lowest and five as the highest. The trust ratings are then transformed into a binary trust variable with values one if the given rating is at least three and zero if the given rating is below that. With this transformation, it is important to note that "zero" does not mean that respondent distrusts the local officials, but only that respondent does not trust them highly enough. Conversely, a value equal to "one" does not mean that local officials suddenly become trustworthy, but only that there are highly trusted. The transformation of the trust rating into a binary outcome facilitates the estimation of propensity scores using probit models to match the treated individuals with control units.

From the eight treatment sites, there are a total of 178 sample individuals – 95 in 2002 and 83 in 2003 – who reported to have been exposed to GI. These are the individuals who were informed of their local government's performance by having read a

komiks, seen a poster or attended a public presentation of the GI. The treated individuals are then matched to a subsample of individuals from four sets of comparison groups. The first comparison group comprises all the 1200 baseline observations in 2001. The second comparison group constitutes the pooled control samples in 2002 and 2003. The third and fourth comparsion groups constitute the 400 control samples obtained in 2002 and 2003, respectively. When the comparsion is made between the treated samples and those in either in the baseline group or in the pooled 2002-2003 conrtol group, the implicit assumption is that there are no unobserved time-varying factors that influenced trust in officials, other than exposure to GI ratings. Of course this assumption may not hold. For one, local government officials may behave auspiciously through time as they try to win supporters for the May 2004 elections. To account for the time-varying factors, the treatment samples for 2002 and 2003 are each compared to the 400 control samples obtained in the same year. Thus, the sub-sample comparison for 2002 and 2003 will reveal possible time-trends in the level of trust in local officials.

4. Matching with propensity scores

The matching of each treated individual with a subssample of individuals from each comparison group is done using propensity scores. The propensity scores are derived from a probit regression model applied on a pooled dataset of treated individuals and comparison groups. The regressors in the probit model are chosen so that the treated individuals and matched sample units will have similar observable characteristics (balancing property). Then, matching is done based on propensity scores and along a common support.

The regression variables and their corresponding descriptive statistics are shown in Table 4. The outcome variables used is Trust, which takes on a value of one if the respondent's trust rating in local officials is at least three in ascending scale running from one to five, and zero if the given rating is below three. The treatment variable is Knowledge of GI, which indicates whether the individual has read a GI komiks, seen a GI poster or attended a public presentation of the GI. The other regressors pertain to the respondent's socioeconomic and demographic characteristics (income, electric bill, owner, college, age, male, household head, spouse, married, regular job, government employee). Of the total 3,600 respondents, about 26 percent at least attended college, 30 percent are male, 39 percent are household heads, 46 percent were spouses (of the household heads), 56 percent had regular jobs, and 67 percent owned the house and lot they reside in. On the average, the respondents are about 42 years old, live with about four other family members, have a combined monthly family income of about 5,500 pesos, and pay about 465 pesos per month for electricity. Also included are indicators of possible exposure to similar LGU performance measures (other index[2]), local political conditions (re-elected mayor), and residence in *poblacion* and other densely-populated barangays that may have been specifically targeted during the information campaigns. About 71 percent of the respondents live in densely populated barangays. Of the 12 mayors in the pilot sites, nine of them were re-elected in the May 2001 local elections. All of them were eligible to run for at least one more term in the May 2004 local elections.

[Insert Table 4 here.]

The regression results of the factors that account for the probability of exposure to the GI ratings are shown in Table 5. Under the Baseline column, where the 1329 observations comprise 178 treated samples and 1151 control samples from the baseline survey, the statistically significant covariates are other index (-0.228), hi-density barangay (-0.48), regular job (0.228), re-elected mayor (0.449), household head (-0.453), owner (-0.359) and the interaction term electric bill*spouse (-0.0005). In the pooled sample comprising the treated individuals and the pooled control units from 2002 and 2003, only hi-density barangay (-0.767) and re-elected mayor (1.316) are found statistically relevant. The results from the two regressions suggest possible unobserved time-varying factors that differentiate the treated individuals from those in the baseline survey. The importance of the time-varying factors is further evident in the differences in the results reflected in the last four columns of Table 5.

[Insert Table 5 here.]

In the column Year 2002, where the 95 treated units are compared with the 390 control units observed in 2002, the significant variables are hi-density barangay (-0.957), re-elected mayor (1.224), household head (0.919), and electric bill (0.0002). In contrast, the statistically variables under the column Year 2003, where the 474 observations combine the 85 treated individuals with 389 control units observed in 2003, are hi-density barangay (-0.47), college (0.503), re-elected mayor (1.622), household head (0.814), spouse (0.939), family size (0.093) and owner (-0.435). The differences in the critical covariates across regression results thus require care in comparing samples across time. Hence, the results of the comparisons of the treated units with those in the baseline surveys or in the pooled samples from 2002 and 2003 could serve as benchmarks against

which impact estimates for 2002 or 2003 can then be contrasted. Notwithstanding the differences in the regression estimates, the specification of the probit models is driven more by the desire to achieve balanced characteristics between the treated and untreated units and to obtain the propensity scores.

To verify the balancing requirement[3], the histograms and standardized biased are analyzed.The histograms of the propensity scores of the matched groups shown in Figure 1. In panel (i) of Figure 1, most of the matched observations are in the lower range of propensity scores ($0<p(X)<0.3$). In panel (ii), the common support is longer and the distribution of the matched groups wider than in panel (ii). In both panels (iii) and (iv), relatively higher proportion of the matched control units (untreated) are found in lower end of the distribution, while relatively more of the treated units are found in the upper end of the distribution.This indicates that while majority of the treated individuals with low propensity scores may be paired with more than one untreated unit, those with high propensity scores may be not be easily matched with an untreated unit of the same score. In this case, these treated units are paired with their nearest, rather than "identical", neighbor using other matching algorithms that allow for control unit with the closest propensity score (i.e., within a radius, band or strata).

[Insert Figure 1 here.]

Further, the balancing requirement is verified with the changes in the standardized bias, 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). Thus, an improvement in the standardized bias for a given covariate indicates that the matched units have

become more similar in the relevant characteristic. Table 6 shows the distributions of the standardized bias achieved before and after matching using the nearest one-to-one neighbor (caliper) algorithm.

[Insert Table 6 here.]

In all cases, the mean, minimum and maximum biases got better with matching. However, the improvements are greater when the treated units are matched with subsamples from 2002 or 2003, than with the control samples from either the baseline or combined 2002 and 2003 surveys. In nearly all cases, a lower pseudo-R^2 is achieved after matching, which suggest no systematic differences in the distribution of covariates between the treated and the matched control groups. The *p*-values of the likelihood ratio tests also suggest joint insignificance of the regressors after matching, except in the case of pooled control units from 2002 and 2003. Last, the matching reduced the bias in the propensity scores by at least 92 per cent. Thus, based on the analysis of histograms and standardized biases, it can be said the matched units have similar observable characteristics.

5. Estimates of the effects GI ratings on trust in local officials

The impact of the GI ratings on trust in local officials are estimated here as average treatment effects on the treated using several matching algorithms. In Table 7, the treatment effects are expressed in terms of change in average probability of trust in officials. Following conventions, bootstrapped standard errors with 100 replications were generated to incorporate the potential increase in the treatment variance due to the

estimation of the propensity score, common support and ordering of observations, and thus provide statistical reliability of the estimates (Caliendo and Kopeinig, 2008).

[Insert Table 7 here.]

In Table 7, the sign and statistical significance of the impact estimates vary through time. In the top panel (Baseline), the differences in the mean probability of trust in local officials are positive and statistically significant between those who were exposed to the GI ratings and those who were not. The difference could be as much as nearly eight percentage points, which suggest that a higher level of trust by the public can be earned with information concerning local government's performance. It should be noted however that the estimated may in fact capture instead a general trend brought about by assumption of a new president in 2001 after the previous one was impeached for corruption.

To control for the possible influence of the change in the national government, the comparison is limited to 2002 and 2003, when presumably the overall political situation in the Philippines started to normalize. In the case where all treated units are compared with pooled control samples from 2002 and 2003, the impact estimates are low and unlikely to be different from zero. In only one matching algorithm (radius, 0.01) is the impact estimate (0.047) statistically significant (but at 10 percent level). In this algorithm, however, the treated unit may be compared with a control with less similar characteristics, and which may account for the differences in the level of trust rather than exposure to GI ratings per se. Thus, in this particular comparison of the treated units with the pooled control samples from the 2002 and 2003 surveys, there seems to be no

systematic differences in the level of trust across the twelve sites over the two years of public dissemination of local government performances.

Possibly, however, the pooled comparison hides a marked trend in the evolution of trust in local officials. When the comparison is limited to each pilot year, the impact estimates are statistically significant and, more interestingly, negative for 2002 and then positive for 2003. For the first year of the public announcements of the GI scores, the mean probability of trust in local officials among those who became aware of the performance ratings of their local governments was lower by at least two percentage points and as much as nearly 13 percentage points than that of those in the control areas. For the second year of the public announcements, the mean probability for the treated individuals were higher by at least eight percentage points and as much as 13 percentage points than those in the control sites. Thus, as it were, the local governments in the treatment sites, by benchmarking their performances and revealing these in public, earned the trust of their constituents.[4]

6. Discussion

The estimates of the average treatment effects on the treated suggest that transparency can lead to trust, that is, local officials by opening themselves up to public assessment of their performance can earn the trust of their constituents. The estimateas based on the comparison between the treated units and those in the baseline year are positive and statistically significant from zero. This particular comparison however may partly capture the general euphoria in 2001 when President Gloria Macapagal Arroyo was sworn to office after the then President Joseph Estrada was forced out of office due to corruption

charges. According to the Social Weather Stations, the net satisfaction ratings of President Estrada was already +9 in late 2000 and that of President Arroyo was +24 when she assumed office in early 2001. The change in administration and the fact that the mayors in the pilot site got a fresh mandate after the May 2001 elections, may have been the confounding factors whose effects are reflected in the impact estimates.

However, even if the comparison is limited to the samples observed in 2002 or 2003, the results still suggest that trust can be earned, although difficult initially. For 2002, when the performance ratings were first announced, the impact estimates were negative. Two reasons could account for this. First, the treated individuals, who might have been initially optimistic about their local governments, could have been disappointed when they got to know about the assessed performances of their local governments. Alternatively, some of them, after getting exposed to GI, got convinced of their own fears or feelings about how the general population assesses their local government.

However, whether or not these fears or suppositions are justified, the "negative" reaction of those informed provides a basis for the generally-perceived reluctance of local governments to subject themselves to performance ratings. Some officials argue that those who were exposed to the GI materials or similar ratings schemes may not have fully understood the information provided them. Other officials contend that the information may have been misrepresented by rivals.

Notwithstanding the deterioration in the level of trust in 2002, the results for 2003 show however that the trust levels can also improve with transparency. In that year, the trust level improved significantly for those who were informed of the performance rating much more than what can be expected from them otherwise. What this could mean is that

the informed population, regardless of their initial disappointment with their local governments, nevertheless will learn to trust them in time. The improvement in the level of trust came at a time when the GI scores dropped in all but one treatment site and two control sites between 2001 and 2002. This means that local officials can earn the trust of their constituents if they remain transparent even though their performance may be declining. In this case, the public disclosure of the decline in performance earns the people's trust precisely because the act is politically costly and therefore credible. Consequently, then, the people learn to value the merit of being transparent independent of the information.

7. Concluding remarks

In sum, the public dissemination of information about local government performance can influence the people's trust in their local government officials. However, the level of trust may initially fall and then only later rise as the people learn to appreciate transparency in itself. Thus, besides promoting accountability, transparency thus may be desired to enhance the overall level of trust in local officials. And to the extent that trust is critical to development, the findings here suggest some policy inputs to prome a more balanced in-country growth under decentralization..

First, the local dissemination of the performance ratings is critical to building trust. In contrast, many of the ratings systems introduced to assess LGU performance in the Philippines are disseminated locally or announced only in national or regional media where only the scores of the top-performers or "best practices" are divulged (Capuno 2007). While such methods of dissemination may also inform the concerned constituents,

they may not encourage further trust in local officials precisely because only the favorable outcomes are announced.

Second, the performance ratings should be institutionalized instead of being undertaken irregularly or as a one-shot activity. As the results here suggest, the regular announcements of the ratings help educate the people to trust the “messanger” irrespective of the “message”. Since the information has to be revealed first before it can be digested, the revelation itself should be politically credited regardless of its implications. In this case, the local officials may still get credit, possibly in terms of trust of their constituents, for fessing up to their faults or failures.

Notes

1. This section, section 2 and section 4 draw from Capuno and Garcia (forthcoming).
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 sensitivity of the results to possible biases induced by unobserved factors is also checked here using the Mantel and Haenszel (MH) statistic proposed by Becker and Caliendo (2007a, 2007b). The MH test 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 the same observed covariates. Except in the case where comparison is made for the combined samples in 2002 and 2003, the results obtained under the assumption that no unobservable factors systematically influenced exposure to GI ratings are found robust. However, in general we cannot reject the hypothesis at the 5% level of significance that a hidden factor that increases the odds ratio to 1.5 could either lead to an underestimate or overestimate of the average treatment effects on the treated. While these MH test do not provide direct evidence of bias of unobserved factors, but only how much bias they need to induce to undermine the initial impact estimates conditioned on observed covariates, they suggest caution in the reliability of the estimated average treatment effects.

References

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.

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

Becker, S.O and Caliendo, M. (2007b) Sensitivity analysis for average treatment effects. *The Stata Journal*,7, 1, pp.71-83.

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

Besley, T. (2006) *Principled agents? The political economy of good government*. Oxford, UK: Oxford University Press.

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.

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

Capuno, J. J. and M. M. S. Garcia (forthcoming) Can information about local government performance induce civic participation? Evidence from the Philippines. *Journal of Development Studies.*

De Dios, E. S. (2007) "Tlocal politics and local economy", in A. Balisacan and H. Hill (eds.), *The Dynamics of Regional Development: The Philippines in East Asia*. Cheltenham, UK: Edward Elgar Publishing. pp. 157-203.

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).

Jenkins, R. and Goetz, A.M. (1999). Accounts and Accountability: Theoretical Implications of the Right-to-Information Movement in India. *Third World Quarterly*, 20(3), pp. 603-622.

Knack, S., Keefer, P. (1997) Does Social Capital Have An Economic Payoff? A Cross-Country Investigation. *Quarterly Journal of Economics*, 112(4), pp. 1251-1288.

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.

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

Paul, S. (1998) Making voice work: The report card on Bangalore's public service. World Bank Policy Research Working Paper 1921,World Bank, Washington, DC.

Ravindra, A. (2004) An assessment of the impact of Bangalore Citizen Report Cards on the performance of public agencies. ECD Working Paper 12, Operations Evaluation Department, World Bank, Washington, DC.

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.

Social Weather Stations. “Net satisfaction ratings of presidents of the Philippines, May 1986 to June 2007.” http: //w.w.w.sws.org.ph/ind-pres.html. Accessed 23 July 2009.

Whiteley, P. (2000). Economic growth and social capital. *Political Studies*, *48*, pp. 443–466.

Wooldridge, J. (2002) *Econometric Analysis of Cross Section and Panel Data* (Cambridge: MIT Press).

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.

Table 2. GOFORDEV Index, 2001 and 2002

Pilot sites	Scores		
	2001	2002	Percent change
Bulacan			
Angat	41	29	-29.3
Baliwag	61	60	-1.6
Bustos	44	35	-20.5
Guiguinto	68	72	5.9
Plaridel	48	44	-8.3
San Jose Del Monte City	52	63	21.2
Davao del Norte			
Asuncion	61	42	-31.1
Braulio E. Dujali	79	76	-3.8
Island Garden City of Samal	60	58	-3.3
Panabo City	58	55	-5.2
Sto. Tomas	44	34	-22.7
Tagum City	60	64	6.7

Table 3. 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.

Table 4. Variable definitions and descriptive statistics

Variable	Definition	Obs.	Mean	Std. Dev	Min.	Max
Trust	1= if respondent ranked 3, 4 or 5 his trust with government officials in a scale of 1 to 5, where 1 is the lowest and 5 is the highest; 0=otherwise	3600	0.833	0.373	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	41.838	14.847	18	90
Age*age	Squared age	3598	1970.801	1390.410	324	8100
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	5.194	2.250	0	28
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
Government employee	1=if the respondent is a government employee or worker; 0=otherwise	3600	0.063	0.243	0	1
Electric bill	Average monthly electric bill for the last six months (in pesos)	3509	464.847	658.690	0	20000
Income, ln	Natural logarithm of monthly family income	3557	8.610	0.942	4.143	12.429
High density barangay	1=if resident in highly populated barangays (village); 0=otherwise	3600	0.713	0.452	0	1
Male	1=if the respondent is male; 0=otherwise	3600	0.306	0.461	0	1
Household head	1=if the respondent is the household head; 0=otherwise	3600	0.387	0.487	0	1
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
Spouse	1=if the respondent is the spouse of the household head; 0=otherwise	3600	0.461	0.499	0	1
Electric bill * spouse	Interaction between average monthly electric bill for the last six months (in pesos) and spouse	3509	193.161	385.449	0	7000
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

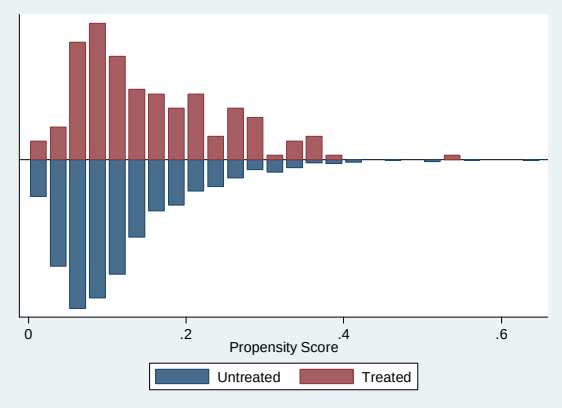
Table 5. Probit model of the probability of knowledge of Gofordev Index
Dependent variable: Knowledge of Gofordev Index (GI)

Explanatory variables	Baseline		Years 2002-2003		Year 2002		Year 2003	
	Coeff	Robust Std. error	Coeff	Robust Std. error	Coeff	Robust Std. error	Coeff	Robust Std. error
Other Index	-0.228[c]	0.120	0.156	0.125	0.181	0.177	0.112	0.189
Hi-density barangay	-0.48[a]	0.120	-0.767[a]	0.129	-0.957[a]	0.177	-0.47[b]	0.183
Income, ln	-0.065	0.071	-0.083	0.06	-0.124	0.077	-0.047	0.106
College	-0.187	0.133			-0.232	0.198	0.503[b]	0.203
Regular job	0.228[c]	0.124	0.073	0.133	0.101	0.184	0.112	0.192
Re-elected Mayor	0.449[a]	0.117	1.316[a]	0.118	1.224[a]	0.162	1.622[a]	0.203
Electric bill * Spouse	-0.0005[c]	0.0003	-0.0004	0.0002	-0.0004	0.0003	0.00007	0.0004
Age	-0.023	0.023	-0.0003	0.022	0.038	0.029	-0.037	0.035
Age * Age	0.0003	0.0002	-0.00002	0.0002	-0.0004	0.0003	0.0002	0.0004
Male	0.075	0.18	0.241	0.158	0.211	0.21	0.308	0.241
Household Head	-0.453[c]	0.232	-0.241	0.221	-0.919[a]	0.293	0.814[a]	0.311
Spouse	-0.133	0.256	0.352	0.249	-0.029	0.319	0.939[b]	0.375
Family size	0.005	0.024	0.028	0.03	-0.015	0.044	0.093[b]	0.040
Electric bill	0.0002	0.0001	0.0001	0.00008	0.0002[b]	0.0001	-0.0004	0.0002
Owner	-0.359[a]	0.125	-0.086	0.133	0.179	0.184	-0.435[b]	0.184
Married	-0.059	0.166	-0.242	0.175	-0.231	0.239	-0.302	0.277
Government job	0.281	0.198	0.205	0.203	0.267	0.276	0.223	0.304
Constant	0.32	0.779	-0.476	0.706	-0.28	0.923	-1.22	1.102
Pseudo R^2	0.085		0.249		0.309		0.271	
Likelihood ratio	52		193		125		91	
P-value	0.0000		0.0000		0.0000		0.0000	
No. of observations	1329		954		480		474	

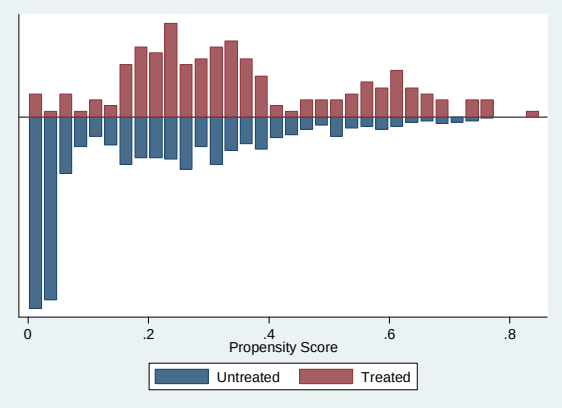
Notes:
The results are obtained using sampling weights.
[a] significant at p<0.01, [b] significant at p<0.05, [c] significant at p<0.10.
*The variable college is dropped in the comparison group Year 1 and Year 2 to satisfy the balancing property in the propensity score estimation.

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

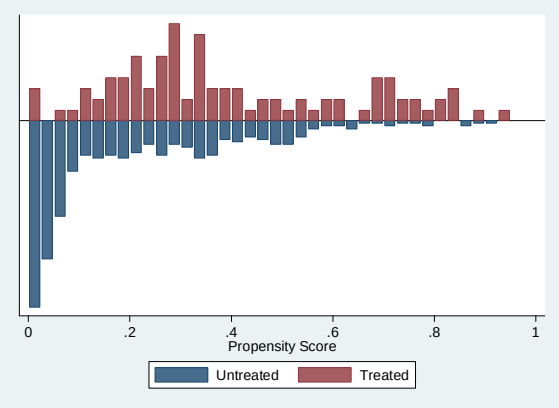
i. Baseline



ii. Years 2002-2003



iii. Year 2002



iv. Year 2003

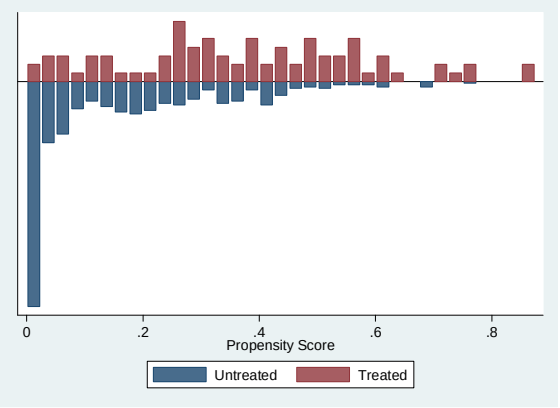


Table 6. Distribution of the standardized bias before and after matching

	Baseline		Years 2002-2003		Year 2002		Year 2003	
	Un-matched	Matched	Un-matched	Matched	Un-matched	Matched	Un-matched	Matched
Mean bias	13.081	10.170	19.536	13.349	21.737	9.970	26.486	8.939
Median bias	8.425	11.122	14.208	12.488	18.777	7.336	21.111	6.541
Minimum bias	2.901	0.319	1.293	1.174	1.292	0.000	1.868	1.044
Maximum bias	40.821	19.082	85.692	35.356	91.801	27.899	116.743	23.767
Pseudo R-squared	0.044	0.029	0.127	0.061	0.152	0.037	0.206	0.3
Chi-square likelihood ratio	46.33	14.19	116.72	30.2	72.52	9.44	90.56	6.92
P-value	0.000	0.511	0.000	0.011	0.000	0.853	0.000	0.96
Bias on propensity score	40.8	0.3	85.7	1.9	91.8	7.3	116.7	9.2
Bias reduction in propensity score (in %)	99.2		97.8		92		92.1	

Note: Estimates of bias based on weighted samples.

Table 7. Estimated average treatment effects on the treated: Impact of Gofordev Index on trust in local officials

Matching method/ Control group	Number of samples		Average treatment effects on the treated	Bootstrapped Standard Error	*t*-statistics
	Treated	Non-treated			
A. Baseline year					
Nearest 1-to-1, caliper 0.1	178	178	0.073[b]	0.039	1.851
Nearest neighbor, random draw	178	149	0.079[b]	0.045	1.743
Nearest neighbor, equal weights	178	149	0.079[b]	0.044	1.808
Kernel, bandwidth of 0.06	178	1132	0.033	0.027	1.211
Radius, radius of 0.01	177	1122	0.021	0.028	0.757
Stratification	178	1132	0.033	0.028	1.174
B. Years 2002-2003					
Nearest 1-to-1, caliper 0.1	178	178	0.034	0.037	0.921
Nearest neighbor, random draw	178	138	0.039	0.048	0.818
Nearest neighbor, equal weights	178	138	0.039	0.04	0.974
Kernel, bandwith of 0.06	178	759	0.027	0.031	0.884
Radius, radius of 0.01	177	669	0.047[c]	0.032	1.485
Stratification	178	759	0.027	0.032	0.84
C. Year 2002					
Nearest 1-to-1, caliper 0.1	92	92	-0.109[b]	0.053	-2.052
Nearest neighbor, random draw	95	67	-0.126[a]	0.049	-2.552
Nearest neighbor, equal weights	95	67	-0.126[a]	0.052	-2.426
Kernel, bandwith of 0.06	95	353	-0.039	0.044	-0.889
Radius, radius of 0.01	83	271	-0.022	0.053	-0.411
Stratification	95	353	-0.047	0.045	-1.043
D. Year 2003					
Nearest 1-to-1, caliper 0.1	82	82	0.073[c]	0.051	1.446
Nearest neighbor, random draw	83	57	0.084[c]	0.062	1.363
Nearest neighbor, equal weights	83	57	0.084[c]	0.058	1.456
Kernel, bandwidth of 0.06	83	295	0.113[a]	0.040	2.836
Radius, radius of 0.01	76	257	0.124[b]	0.054	2.278
Stratification	83	295	0.131[a]	0.042	3.137

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 bootstrapped standard errors are estimated using 100 replication samples.
Matching using nearest 1-to-1 algorithm is done without replacement using psmatch2 (Leuven and Sianesi 2003), and the other matching algorithms (nearest neighbor, kernel, radius and stratification) using pscore (Becker and Ichino 2002).
[a]Significant at 1% level.
[b]Significant at 5% level.
[c]Significant at 10% level.
The bootstrapped standard errors are estimated usinng 100 replications.