M-13.

THE NEED FOR MULTIPLE INDICATORS IN MEASURING THE OUTPUT
OF PUBLIC AGENCIES

by

Elinor Ostrom
Workshop in Political Theory and Policy Analysis
Department of Political Science
Indiana University

/99.?

The need for more valid and reliable means of measuring the output of government services at all levels of government is a critical problem for policy analysts. It is particularly important at the municipal level, where such outputs and their variations have the most direct impact upon citizens, and where sophisticated measurement and evaluation capabilities are least likely to be found. This need is easily documented by reference to: **1)** Reports of service breakdowns contributing to the exodus of middle-class residents from central cities (Without valid and reliable measures of services provided, how can we begin to estimate service levels required to stem or reverse this exodus?); 2) Lawsuits contending infringement of the equal rights provisions of the constitution through discriminatory provisions of services (Without valid and reliable measures of services provided, how can such cases be equitably resolved?); and 3) Frequent calls for reorganization of governmental arrangements, particulary in metropolitan areas, in order to provide government services more efficiently (Without more accurate measures of services provided, how can we establish the probable consequences of such reforms, or measure those consequences in cases where such reorganizations are carried out?).[1]

The task of measuring the output of most government agencies is far more difficult than that of measuring the output of private firms. The output of a private firm is defined as the total quantity of any product manufactured or produced over a period of time. The output of a private firm can frequently be measured in physical terms, such as the total cars produced a year or the total loaves of bread produced a day. The number of physical units produced is an acceptable measure for most purposes.

While the output of a private firm is frequently measured in terms

of the physical units produced, the dollar value of the output is also used as a measure of output. Prices for private goods and services are established by voluntary transactions between buyers and sellers. Consumers voluntarily enter markets and decide whether and how much of a particular good to purchase from among the alternatives offered. A consumer unwilling to pay the market price for a good can be excluded from consuming its benefits. In deciding to buy or not to buy, a consumer must make some estimate of the value he will receive in the consumption process. A consumer will be willing to purchase a good only if he expects that the benefits he will receive, less the costs he will incur in consuming a good, exceed the net benefits he could obtain from using the purchase price in a different fashion. Thus, the dollar value of the output of a private firm determined by the prices at which units are sold includes a measure of both **1)** the amount of output and **2)** the expected value of the output by the consumer.

In the public sector, it is much more difficult to conceptualize and measure either the physical output of a public agency or the appropriate dollar value of the output. What is the output, for example, of a police department? In a general sense, the output of the police is the maintenance of order, the prevention and detection of crime, the enforcement of the law and the provision of emergency response services to citizens. One sees police on patrol and assumes that all individuals within some range of effects are benefited by this activity. However, the "production process" does not produce a clearly defined output and the lack of a market makes it difficult to know how much those effects are valued by those receiving them.

The outputs of many municipal agencies exhibit several characteristics of public goods. In particular, the outputs are often indivisible and potential beneficiaries cannot easily be excluded from receiving the effects of the output. Street lighting, for example, cannot be provided to one resident of a block without simultaneously providing it to others living on the block or travelling along the block. The costs of excluding some from the benefit so that those who do benefit can be charged for the benefit would be very high. Given indivisibility and nonexcludability, the problem of measuring output pervades the analysis of all public policy related to the provision of municipal public services.[3]

The indivisibilities and lack of exclusion characteristic of public goods prevents reliance upon single measures such as quantity or dollar value of output. It becomes necessary to measure several aspects of public output simultaneously. Thus, the speed of response made by a local police department is one indicator of the service level being provided by the department. Sole reliance on such a single indicator as the measure of output, however, could lead to serious problems both for policy analysts and public actors.

The F.B.I. Crime Index has too frequently been utilized as a single measure of the output of police agencies.[4] The problem of widespread non-reporting of crimes can be corrected to some extent by supplementing official F.B.I, crime reports with other indicators of victimization derived from household sample surveys.[5] Further indicators of the output of police agencies can be obtained by asking receipients of police service about their specific experiences with and evaluations of local police services. The ratio of warrants issued to warrants applied for can be used as a comparative indicator of the performance of police departments

served by the same prosecutor.[6] The geographic distribution of the sale of private protection devices can be used as a negative indicator of the sense of safety felt by citizens and their lack of reliance on the police for primary protection.

Employing any single indicator of public agency output as the basis for structuring internal incentive systems can lead to pathologies. Jerome Skolnick has graphically described one such pathology resulting from the excessive reliance upon "the clearance rate" by the Oakland Police Department as a measure of performance.[7] Detectives in the department were led to treat the most active and confirmed burglars, once arrested on a charge, as a potential resource of great value for future promotions. If the detective were able to offer a suspected burglar a sufficient reward in the form of significantly reduced charges, he might convince him to confess to a large number of crimes--thus "clearing" them all with a single arrest and a reduced charge.

Policy analysts can fall into the same trap by reliance upon single indicators. Far too many articles examining factors affecting the output of public agencies have utilized as their sole measure of output an absolute or relative input quantity such as total public expenditures or per capita public expenditures. The use of expenditures figures by social scientists as the sole measure of output has legitimized the growing reliance of the courts on similar indicators. Recently, an important series of cases has been argued before the courts in which the basic issue has been the unequal distribution of public goods and services to particular neighborhoods based on alleged racial, ethnic, or economic discriminations.[8] Most of the cases focusing on the provision of educational

services have utilized dollar resources per child as their measure of

output.[9] James Coleman and others have challenged the adequacy of simple

dollar expenditures levels as measures of the output of educational

systems.[10]   In several studies of police performance in the Indianapolis,

Chicago, Grand Rapids, St. Louis, and Nashville metropolitan areas, the

output of the police measured by a series of multiple indicators of per-

formance has not been positively associated with average per capita ex-

penditure levels by city.[11] Excessive reliance upon expenditure levels

as the sole indicators of output may lead the courts into the position

of finding the most wasteful (or most graft-ridden) cities providing the

highest levels of output to their citizens.

Such methodological traps can be mitigated by the conscious develop-

ment and reliance on multiple indicators of output derived wherever possible

from multiple modes of data collection. Researchers at the Urban Institute

have pioneered in this regard. They have explored a wide variety of poten-

tial indicators to be used simultaneously in any effort of measure output.[12]

The Workshop in Political Theory and Policy Analysis at Indiana

University has recently embarked on a research effort for the RANN Division

of the National Science Foundation. The major focus of this project is the

development of multi-mode measurements of two municipal services: street

lighting and street repair. For street lighting, we developed a method to

utilize a precision light-meter to record the level of night lighting on side-

walks and streets facing a particular block face.   For street repair, we

developed a mechanical device, called the Residential Street Roughness Indi-

cator, to measure the roughness of, a street.   In addition, we also devel-

oped an observation form and procedure which can be used by trained observers

to record specific data about various aspects of street condition.

In addition to the physical mode of data collection used for each of the service areas, we also utilized two other modes: a survey of respondents, and indicators derived from agency records. A pretest of our survey instrument was administered to respondents living in seven small neighborhoods in Indianapolis during 1974. A total of 326 respondents were interviewed in this test. Citizens were asked a series of questions to elicit their perceptions of street lighting and road repair, their evaluations of the same, and their preferences for differing levels of these services. Agency records provided our third mode of data collection. Unfortunately, we found this mode of data collection to be the most difficult of all. Agency records were so fragmentary that few consistent indicators could be developed. For road repair we were able to code the frequency of complaints directed to the Indianapolis Department of Transportation concerning the roads facing the respondents in our survey. For street lighting, we were not able to code much more than the frequency and pattern of street lights shown on agency maps.

Preliminary data analysis has been initiated. One of the first questions we have addressed is "What is the relationship between citizen perceptions of service levels and our unobstrusive measures of service levels?" This is a particularly important question since many policy analysts are hesitant to rely at all upon citizen reported evaluation of output due to the assumed inaccuracy of citizen perceptions of service levels. Early analysis does not provide a completely uniform picture of accurate perception across all indicators. However, the more specific and concrete the referant to which our questions were addressed, the

more likely a high level of association exists between unobstrusive meas-
ures and citizen perceptions of service levels.

For example, in general citizens were quite accurate in their per-
ceptions concerning a number of specific aspects concerning the condition
of street repair on their block face. Citizens accurately reported the
type of street surface, the presence or absence of curbs, the condition
of their curbs, the presence of surface disintegration, and the presence
of potholes.[15]

Sue Carroll of the Workshop staff developed a roughness scale for
each block face included in our study and for each quadrant of a block on
which respondents lived.[16] The scale was composed of individual items
coded for each quadrant of a block including amount of surface disinte-
gration, number and size of potholes, presence or absence of cracks,
presence or absence of bumps, and presence or absence of utility cuts.
Each observer was also asked to rate each blockface as being 'Very
rough," "fairly rough," "fairly smooth," and "very smooth." These
observer ratings were strongly related to the "roughness scale" for
both a quadrant (gamma = .94) and for the block face as a whole (gamma =
.97). Given these high coefficients, the roughness scale derived from
the individual items coded by observers would appear to have at least
some face validity.

When respondents' perceptions of the roughness of the street on their
block were then associated with the quadrant and blockface roughness scale,
the measure of association between them is fairly strong (gamma = .76 for
both scales). As shown on Table I, some variation occurred across various
control variables. Those persons with more than a high school diploma.

Table 1*

MEASURES OF ASSOCIATION BETWEEN CITIZEN PERCEPTIONS OF ROUGHNESS AND SCORES ON BOTH THE QUADRANT AND THE BLOCKFACE ROUGHNESS SCALES

	Sex			Education			Age				Length of Residence		
	Entire Sample	Females	Males	Less Than High School Graduate	High School Graduate	More Than High School Graduate	Less Than 30 yrs. old	30-45 yrs. old	46-60 yrs. old	More Than 60 yrs. old	Less Than 1 yr.	1-5 yrs.	More Than 5 yrs.
Association Between Perceptions and Scores on Quadrant Roughness Scale	gamma = .76	.74	.78	.73	.69	.85	.68	.61	.87	.93	.66*		.89
	N=(247)	(142)	(102)	(82)	(77)	(80)	(80)	(66)	(57)	(44)	(126)		(121)
Association Between Perceptions and Scores on Blockface Roughness Scale	gamma = .76	.71	.81	.72	.71	.80	.67*		.89*		.67*		.87
	N=(319)	(183)	(132)	(99)	(106)	(106)	(190)		(129)		(160)		(159)

CONTINUED

Length of Blockface		
Short (<650 ft.)	Medium (650-900 ft.)	Long (>900 ft.)
.81	.85	.44
(74)	(79)	(74)
Yule's Q= .91**	Yule's Q= .69**	Yule's Q= .69**
(100)	(101)	(118)

* NOTE: Because of the nature of the distributions, there were too few respondents to compute gamma for these categories separately. Therefore, two categories have been combined.

** NOTE: Even by combining these two categories, there were so few respondents with scale scores placing them into the "very rough" or "fairly rough" scale categories that gamma could not be computed. Therefore, the scale was dichotomized into "rough" and "smooth" and Yule's Q was computed.

those over 45, those who have lived on a block more than five years, and those living on medium to short blocks tended to be more "accurate" in their perceptions of road roughness. Initial data analysis with scores produced by the Residential Street Roughness Indicator device has produced consistent findings with those reported on here. There is a high association among all of these individual modes of data collection concerning the level of road roughness--individual respondent perception, the roughness scale computed from observer coding of street conditions, and the output from the mechanical device for measuring road roughness.

Although the levels of association are not, in general, as high as in the case of street conditions, statistically significant correlations between citizen perceived streetlight brightness levels on their blockface and data from a precision protoelectric meter were found. Further, a distinct pattern emerged between the strength of association and the size of the interval on either side of a respondent's house over which light-meter readings were averaged. Correlations reach a maximum when meter readings are averaged over intervals relatively proximate to a respondent's home and decline as the meter readings are averaged over widening intervals. The lowest correlation is between citizen perceptions and light-meter readings averaged over the entire blockface. Indeed, for some subsets of the sample, this correlation was not statistically significant. Thus, citizens appear to show a pronounced tendency to perceive blockface street-light brightness conditions in terms of the brightness levels relatively proximate to their own homes. When asked specifically about conditions proximate to their homes, citizens are more accurate still. We also found that citizens who had lived on their block for more than 10 years, who had

a high school or better education, or who lived on relatively short blocks showed a higher than average degree of accuracy.

As further analysis proceeds, we hope to address the question of which measures can be used most effectively and economically by public officials or public interest groups in measuring the output of these two municipal services. Where we have multiple modes of data collection for the same attribute, such as we have for road roughness, we can begin to answer the question of which mode is most "cost effective" in providing a reliable and valid indicator of that attribute. It would appear, for example, that the observation form developed to measure several attributes of road repair can be used to generate a valid measure of road roughness which is highly associated with both citizen perceptions of roughness and with a mechanical recording of road roughness. Given the considerable economy of administering an observation form as compared to either a survey of residents or the use of the Residential Street Roughness Indicator, one would have to judge the observation form to be the most cost effective mode of data collection concerning road roughness of those studied.

FOOTNOTES

1. See Robert A. Bish and Vincent Ostrom, <u>Understanding Urban Government</u> (Washington, D.C.: American Enterprise Institute, 1973).

2. Paul A. Samuelson, "The Pure Theory of Public Expenditure," <u>The Review</u> of <u>Economics and Statistics</u> 36 (November 1954), pp. 387-389.

3. William J. Baumol, <u>Welfare Economics and the Theory of the State</u> (Cambridge, Massachusetts: Harvard University Press, 1962), and Mancur Olson, <u>The Logic</u> of <u>Collective Action</u> (Cambridge, Massachusetts, Harvard University Press, 1965).

4. For a discussion of the loss of information about the large volume of service calls and activity levels resulting from a police department's primary reliance upon the F.B.I. Index Crimes, see Roger B. Parks, <u>Measurement</u> of <u>Performance</u> in <u>the Public Sector</u>: A <u>Case Study of the Indianapolis Police Department</u> (Bloomington, Indiana: Indiana University, Department of Political Science, Workshop in Political Theory and Policy Analysis, 1971). See also Elinor Ostrom, "Institutional Arrangements and the Measurement of Policy Consequences: Applications to Evaluating Police Performance," <u>Urban Affairs Quarterly</u> 6 (June 1971), 447-476.

5. See President's Commission on Law Enforcement and Administration of Justice, <u>The Challenge of Crime</u> in a <u>Free Society</u> (Washington, D. C.: Government Printing Office, 1967).

6. This indicator of performance is utilized in Dennis Smith and Elinor Ostrom, "The Effects of Training and Education on Police Attitudes and Performance: A Preliminary Analysis," in Herbert Jacob, (ed.) **The** <u>Potential for Reform</u> of <u>Criminal</u> **Justice** (Beverly Hills: Sage Publications, 1975). For a discussion of other measures of police performance see Elinor Ostrom, "On the Meaning and Measurement of Output and Efficiency in the Production of Urban Police Services," <u>Journal of Criminal Justice</u> 1 (June 1973), pp. 93-111.

7. Jerome Skolnick, <u>Justice Without Trial</u>: <u>Law Enforcement</u> in <u>Democratic Society</u> (New York: John Wiley & Sons, 1967).

8. Ralph S. Abascal, "Municipal Services and Equal Protection: Variations on a Theme by Griffin V. Illinois," <u>Hastings Law Journal</u> 20 (May 1969), pp. 1367-1391; Dennis R. Anderson, "Toward the Equalization of Municipal Services: Variations on a Theme by Hawkins," <u>Journal</u> of <u>Urban Law</u> 50 (November 1972), pp. 177-197; Michael P. Schumaecker, "Equal Protection: The Right of Equal Municipal Services," <u>Brooklyn Law Review</u> 37 (Spring 1971), pp. 568-587.

9. See San Antonio Independent School District v. Rodriguez, 41 U.S.L.W. 4407 (1973); Sweatt v. Painter, 339 U.S. 629 (1950); Serrano v. Priest, 96 Cal. Reptr. 601 (Sup. Cr. Cal., 1961); U.S. v. Jefferson County Board of Education, 380 P. 2d. 385 (5th cir., 1967); Hombson v. Hansen, 269 F. Supp. 401 (D.D.D., 1967).

10. James Coleman, "The Concept of Equality of Educational Opportunity," Harvard Educational Review 38 (Winter 1968), pp. 7-22.

11. Elinor Ostrom, William H. Baugh, Richard Guarasci, Roger B. Parks, Gordon P. Whitaker, Community Organization and the Provision of Police Services (Beverly Hills: Sage Publications, 1973); Elinor Ostrom and Roger B. Parks, "Suburban Police Departments: Too Many and Too Small," in Louis H. Masotti and Jeffrey K. Hadden (eds.) The Urbanization of the Suburbs (Beverly Hills: Sage Publications, 1973); Samir T. IsHak, Consumers' Perception of Police Performance, Consolidation vs. Deconcentration. The Case of Grand~Rapids, Michigan Metropolitan Area, (Bloomington, Indiana: Indiana University, Department of Political Science, Workshop in Political Theory and Policy Analysis, 1975); Bruce D. Rogers and C. McCurdy Lipsey, "Metropolitan Reform: Citizen Evaluations of Performance in Nashville-Davidson County, Tennessee," Publius (Spring 1975).

12. See Andrew J. Boots, Grace Dawson, William Silverman, Harry P. Hatry, Inequality in Local Government Services: A Case Study of Neighborhood Roads (Washington, D.C.: The Urban Institute, no date); Donald M. Fisk, Harry P. Hatry, Kethleen Hudak, Kenneth Webb and Robert Fiore, Measuring the Effectiveness of Local Government Recreation Services (Washington, D.C.: The Urban Institute, 1972); Richard E. Winnier and Harry P. Hatry, Measuring the Effectiveness of Local Government Services: Transportation (Washington, D.C.: The Urban Institute, 1973).

13. See Richard Rich, "The Development of a Technique for the Physical Measurement of Residential Street Lighting," (Bloomington, Indiana: Indiana University, Department of Political Science, Workshop in Political Theory and Policy Analysis, Research Report Number 5).

14. See Richard Rich, "The Development of the Residential Street Roughness Indicator as a Mode of Measurement for the Study of Municipal Services," (Bloomington, Indiana: Indiana University, Department of Political Science, Workshop in Political Theory and Policy Analysis, Research Report Number 6).

15. See Sue Carroll, "An Analysis of the Relationship Between Citizen Perceptions and Unobstrusive Measures of Street Conditions," (Bloomington, Indiana: Indiana University, Department of Political Science, Workshop in Political Theory and Policy Analysis, Research Report Number 10).

16. Ibid.

17. See Vernon Greene, "An Analysis of the Relationship Between Citizen Perceptions and Physical Measures of Street Lighting," (Bloomington, Indiana: " Indiana University, Department of Political Science, Workshop in Political Theory and Policy Analysis, Research Report Number 7).