Citizen Satisfaction With Urban Services: Potential Misuse as a Performance Indicator

Brian Stipak, *Pennsylvania State University*

■ Policy analysts should exercise caution in using survey data on citizen satisfaction with local services to measure governmental performance. Responses to survey items asking citizens how satisfied they are with specific local services, or asking them to evaluate service performance, may not reflect the actual service government provides. Also, difficult statistical and conceptual problems often invalidate using such data to evaluate local services. Nevertheless, subjective data like citizen satisfaction may yet have some use in policy-making.

Sample surveys of citizens provide a potentially important method of collecting data for local policy analysis. Several Urban Institute publications have recommended that local governments annually conduct citizen surveys to aid in evaluating city services.[1] Local officials will probably collect information on citizen satisfaction with city services as part of such surveys. A danger arises that policy makers may misinterpret such information, especially if they use citizen satisfaction to measure service performance. This danger stems from two problems: 1) citizen responses to satisfaction and evaluation questions may not reflect actual service performance, and 2) difficult statistical and conceptual problems complicate the use of subjective data to evaluate service performance.

Before examining these two problems, picture how local officials might use information about citizen satisfaction with a particular local service. Policy makers might compare satisfaction levels in different geographic areas, conclude that areas reporting lower overall satisfaction receive poorer quality service, and perhaps even reallocate expenditures accordingly.[2] Similarly, policy makers might compare satisfaction expressed by different types of citizens, such as citizens from different demographic groups, to investigate whether one type receives better service than another. Although such procedures may seem sensible, in general they are invalid and potentially misleading.

Problem #1: Expressed Satisfaction May Not Reflect Service Performance

Policy makers can use subjective indicators constructed from survey items as indicators of service performance only if the responses citizens make to those items are linked to the actual services government provides. In particular, whether expressed citizen satisfaction with a service is a valid performance measure depends on whether it reflects service characteristics or quality. If the characteristics or quality of the service actually provided do not affect citizen satisfaction or evaluations, policy makers cannot logically use such indicators to measure service performance.

In a study within the Los Angeles metropolitan area, I tested whether differences in the services local governments provide do affect citizen evaluations of those services.[3] That study used a data base of merged individual, census, administrative, and other data to estimate models of citizen evaluations as a function of 1) service characteristics, 2) governmental characteristics, 3) neighborhood characteristics, and 4) individual characteristics. Subjective evaluation scales were created from survey items asking citizens to evaluate police, parks and recreation, refuse collection, and other basic services. A number of indicators for each service were used to measure different types of objective service characteristics—service outputs, service inputs, administrative workloads, and related community conditions. For example, the indicators examined for the police included clearance rate, property recovery rate, per capita expenditures, per capita employees, and crime rate. The statistical results showed little evidence that service characteristics affect citizen evaluations of local services. Most coefficient estimates for the objective service indicators were not statistically significant, and were too small to be substantively important. Also, the predictive power of the objective service indicators was usually small compared to some of the other predictors.

Tables 1 and 2 display some of the results for police and parks and recreation. The numbers are the estimated change in the subjective evaluation scale for that service, measured in standard deviations, resulting from a particular change in a predictor variable.[4] For example, increasing the clearance rate 10 per cent for the seven major felonies

Brian Stipak is assistant professor of public administration, Institute of Public Administration, Pennsylvania State University. He has been a postdoctoral fellow at the Institute of Public Policy Studies, University of Michigan, and was formerly employed by System Development Corporation, where he was involved in evaluation research.

(e.g. from 20 per cent to 30 per cent) raises the evaluations citizens express an estimated one-third of a standard deviation. Increasing the number of full-time police personnel per 10,000 population by five yields an estimated improvement of less than one-tenth of a standard deviation. For parks (Table 2) the estimated effect of five additional employees per 10,000 population is a .27 standard deviation improvement, and an estimated .18 standard deviation improvement results from an expenditure increase of $10 per capita. The results for the other objective service indicators are not even statistically significant. Interestingly enough, the distance a respondent lives from the nearest park, a rough index of accessibility of park facilities, shows no effect on the respondent's evaluation of park services. Also, differences in the crime rate show no effect on evaluations of the police service.

Even the statistically significant results for the objective service indicators are not substantially impressive. For example, the clearance rate variable shows one of the strongest affects of all the objective indicators in the entire study. Nevertheless, even half a standard deviation improvement in citizen evaluations would require a 15 per cent improvement in the clearance rate—an exceptionally difficult improvement to realize, given that clearance rates for the Los Angeles sample have a mean of 24 per cent and a standard deviation of 6 per cent. Such a huge improvement in objective performance affects citizen evaluations of police only about as much as a respondent's being in the oldest age category, compared to the youngest, or as living in an area served by the Los Angeles County Sheriff's Department (unincorporated areas or contract cities), compared to small cities with their own police departments. A clearance rate improvement equivalent to its standard deviation for the Los Angeles sample, a more reasonable municipal goal, would effect an estimated improvement of less than a fifth of a standard deviation on the evaluation scale. The efficacy of potentially achievable changes in any of the other objective service indicators to improve expressed citizen evaluations is even less, and is surpassed by effects attributable to demographics and to governmental jurisdiction. In short, large improvements in objective performance—large in the sense of what public officials could realistically hope to achieve—generally appear to have negligible effects on citizens' subjective evaluations.

The Kansas City Preventive Patrol Experiment provides experimental evidence that complements the findings of the Los Angeles study. In that experiment, the intensity of routine preventive police patrol was varied widely across areas within the city. The results showed little effect of differences in patrolling on citizen satisfaction with the police service, citizen fear of crime, or a variety of attitudes toward the police.[5] No effect was found even on citizens' perceptions of the time police spent on patrol,[6] despite some experimental areas receiving three times the normal intensity of patrol, and other areas receiving no routine patrol. Although some other existing studies do purport to provide some weak evidence of a correspondence between citizen satisfaction and service conditions,[7] those studies typically rely on bivariate analysis of city-levels data. The Kansas City experiment and the Los Angeles findings

Table 1

Estimated Independent Effects of Selected Variables on Citizen Evaluations of Police Services in Los Angeles County

Change in Predictor	Estimated Effect in Standard Deviations on Evaluation Scale[a]
Objective Service Indicators:	
Clearance rate, 7 major felonies: increase 10%	.33[b](.14)
Recovery rate for stolen property, except automobiles: increase 10%	.05 (.20)
Operating expenditures per capita for police services: increase $10	.049 (.046)
Number of full-time police personnel per 10,000 population: Increase 5	.075[b](.045)
Number of 7 major felonies committed per 1,000 population: increase 20	.032 (.050)
Demographic Characteristics of Respondent:	
Education: increase 4 years	-.016 (.048)
Income level, standardized by poverty level for family type: double	.006 (.037)
Sex: Male compared to female	.023 (.065)
Race: black compared to white	-.15 (.17)
Spanish-surname compared to white	-.02 (.11)
Age: 30-39 compared to 18-29	.11 (.09)
40-49 compared to 18-29	.33[c](.10)
50-59 compared to 18-29	.33[c](.11)
60+ compared to 18-29	.45[c](.10)
Governmental Jurisdiction:	
Unincorporated areas, compared to small cities (16-138 thou. pop.) with own police department	-.53[c](.11)
Los Angeles City, compared to small cities (16-138 thou. pop.) with own police department	-.33[c](.08)
Long Beach City, compared to small cities (16-138 thou. pop.) with own police department	-.15 (.15)
Cities which contract with L.A. County for police services, compared to small cities (16-138 thou. pop.) with own police department	-.48[c](.11)

[a]Standard errors are in parentheses.
[b]statistically significant at .05 level, 1-tail test
[c]statistically significant at .05 level, 2-tail test

clearly warn local officials not to assume automatically that citizens respond to satisfaction or evaluation questions on the basis of objective service performance.

One possible explanation why citizens may not respond to such questions based on service performance is that citizens pay little attention to those services, as long as service quality is within some adequate range. Public opinion research has generally found that citizens know little about government and public affairs, but nonetheless will express political opinions.[8] Similarly, citizens may quite willingly provide evaluations of specific local services, despite a lack of knowledge or perceptions of service quality.[9] Figure 1 illustrates the plausible hypothesis that within the range of service quality typically found in U.S. metropolitan areas, citizens pay little attention to services and fail to perceive differences in service quality; therefore, their evaluations fail to reflect differences in service quality.

Table 2

Estimated Independent Effects of Selected Variables on Citizen Evaluations of Park and Recreation Services in Los Angeles County

Change in Predictor	Estimated Effect in Standard Deviation on Evaluation Scale[a]
Objective Service Indicators:	
Distance to nearest park: increase .3 mile	-.024 (.030)
Operating expenditures per capita for park and recreation services: increase $10	.18[b](.09)
Number of full-time park and recreation personnel per 10,000 population: increase 5	.27[b](.10)
Demographic Characteristics of Respondent:	
Education: increase 4 years	-.076 (.044)
Income level, standardized for poverty level of family type: double	.001 (.036)
Sex: male compared to female	-.072 (.062)
Race: black compared to white	-.25 (.16)
Spanish-surname compared to white	-.20[b](.11)
Age: 30-39 compared to 18-29	.28[c](.09)
40-49 compared to 18-29	.14 (.10)
50-59 compared to 18-29	.29[c](.10)
60 + compared to 18-29	.35[c](.10)
Governmental Jurisdiction:	
Unincorporated areas, compared to small cities (16-138 thou. pop.)	-.22[c](.10)
Los Angeles City, compared to small cities (16-138 thou. pop.)	-.37[c](.07)
Long Beach City, compared to small cities (16-138 thou. pop.)	.51[c](.15)

[a]Standard errors are in parentheses.
[b]statistically significant at .05 level, 1-tail test
[c]statistically significant at .05 level, 2-tail test

Figure 1

ONE POSSIBLE FUNCTIONAL FORM OF THE RELATIONSHIP BETWEEN EXPRESSED CITIZEN SATISFACTION AND SERVICE QUALITY

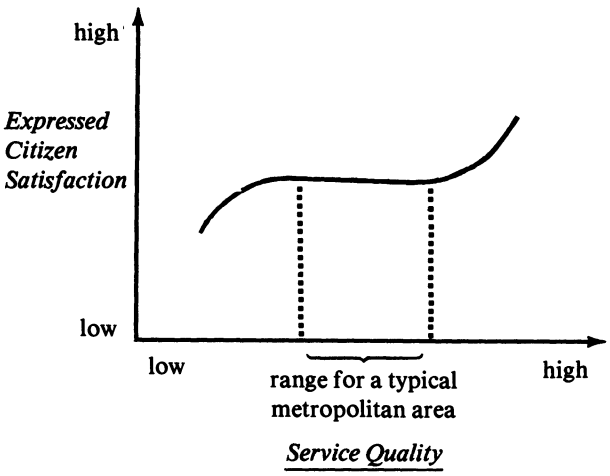

Figure 1 presents an interpretation which can account for the findings of the Los Angeles study mentioned above without arguing that under no conditions do citizen evaluations reflect service quality. If service becomes excessively bad or remarkably good—i.e., sufficiently divergent from citizens' expectations—service quality may become salient enough to affect evaluations. For example, citizens may have little interest in a service like street repair, paying little attention to service performance if quality is above minimum level. However, if large potholes in the streets make driving difficult and unpleasant, the low quality of streets may become conspicuous enough to affect evaluations of street repair obtained by citizen surveys.

Problem #2: Complications in Analyzing Subjective Indicators

Even if we assume there is a strong link between the actual service government provides and citizen responses to evaluation and satisfaction questions, analyzing those responses for performance evaluation faces complex conceptual and statistical problems. It is not valid simply to compare overall levels of satisfaction in different geographic areas and attribute those differences to differences in service performance. For example, a policy maker might find that a higher percentage of citizens in one service area say they are dissatisfied with refuse collection than in another service area. The policy maker cannot validly infer

that this difference in overall satisfaction reflects a difference in actual performance, because of both conceptual and statistical complications.

The conceptual complications arise from the unknown form of the relationship between the subject measure and the dimension(s) of service quality upon which citizens base their expressed satisfaction or evaluations. Ideally, from the analyst's point of view, there would always be an increasing monotonic relationship between the subjective indicator and the quality dimension. That is, the subjective indicator would always increase with increases in actual quality. In that case, the analyst could use the subjective indicator to rank the quality of service individuals or groups receive.

. . . large improvements in objective performance—large in the sense of what public officials could realistically hope to achieve—generally appear to have negligible effects on citizens' subjective evaluations.

However, monotonicity may not exist, even if the individual citizens are knowledgeable about the services the researcher asks about and respond according to their knowledge and perceptions. Different respondents may base their subjective assessments on different aspects of the service, depending on what aspects of performance they perceive as most salient or important. Different respondents may have different expectations of what service quality should be, and apply different standards in evaluating performance. Some respondents may respond by comparing perceived performance to a specific desired level of service, perhaps because of their assessments of the marginal costs and benefits of improving performance. Such respondents will express lower satisfaction beyond a certain performance level, whereas other respondents may express greater and greater satisfaction the higher the perceived performance. In short, overall satisfaction levels, such as the percentage from an area which expresses high satisfaction or low satisfaction, result from some unknown mixture of different perceptions, expectations, and types of evaluation processes. Because of these complications, high levels of subjective performance do not necessarily imply that in any other sense service performance is higher.[10]

In order to deal with these conceptual difficulties, analysts using subjective data to evaluate service performance must make a number of strong assumptions, often of questionable validity. The necessary assumptions will depend on the particular type of analysis, as well as the standards of rigor the analysts apply, since some analysts may undertake an analysis not defensible according to strict standards of rigor and logic if that is the only analysis possible. However, almost all reasonable analyses require assuming at least 1) that most respondents base their responses on the same aspects of service performance, which they perceive fairly accurately, and 2) that most respondents respond monotonically, i.e., that most respondents will express greater satisfaction the higher the perceived performance. Comparing average satisfaction levels requires the additional assumption that the subjective data approximate interval measurement. Also, in comparing group averages for respondents from different areas or demographic categories, any effort to rank-order the groups makes sense only if the groups have comparable intra-group variation. An assumption is thus required that the distributions (about the group means) of quality of service received by members of different groups are similar, or alternatively, that members within each group receive approximately the same quality of service.

The statistical complications which plague analyses of satisfaction data for performance evaluation basically result from the nonexperimental nature of the research. Samples of citizens are not equivalent to experimental groups in the laboratory. Citizens are not randomly assigned to different "treatments" consisting of different service levels in different areas of the city. Rather, citizens from different areas often vary on a number of individual characteristics, such as race and income, which may affect satisfaction independent of the level of service. Also, the different "treatments" (different service levels) found in different areas are not "administered" with all other factors held constant, because areas vary on other neighborhood characteristics besides the quality of governmental services.

Because of these nonexperimental conditions, variations in citizen satisfaction can result from factors other than service performance. For example, the type of citizens living in one area may tend to have especially high expectations for service performance, and therefore tend to express lower satisfaction at a given level of actual performance than do citizens in other areas. To take another example, citizens living in neighborhoods with dilapidated housing may tend to generalize their dissatisfaction with housing conditions to all aspects of the neighborhood, including local governmental services. Comparisons of satisfaction levels in different areas within a city may therefore reveal differences that are merely artifacts of these types of factors, and that cannot validly be interpreted as differences due to the quality of service government provides.

The policy analyst can try to avoid such artifactual findings by employing more sophisticated statistical techniques than simple comparisons of overall satisfaction levels across areas. One approach is to estimate an individual-level multiple regression equation which regresses respondent satisfaction on individual-level variables and a set of dummy variables corresponding to the service areas being examined. The purpose of that type of multivariate statistical procedure is to remove the effects of factors other than service performance, so that the analyst can examine the independent effect of differences in service performance.

To illustrate how policy analysts could use such procedures, I will take a very simple case. Assume that the analyst has an approximately interval-level scale, Y, of citizen satisfaction with a local service, created from a number of survey items. Also assume:

1) Expressed satisfaction, Y, with a particular service increases linearly with some continuous, unmeasured dimension, Q, of the quality of that service.
2) Y is also influenced by some characteristic of the respondent, X, which has an additive effect. For example, X might be a dummy variable corresponding to the respondent's race (e.g. black vs. white), if one racial group is known to have a negative or positive evaluative bias relative to the other.
3) Y has a stochastic component, ε, which is unrelated to Q, X, or the service area in which the respondent resides.

The basic model therefore can be written:

$$Y = \alpha + \beta Q + \gamma X + \varepsilon \qquad (1)$$

For simplicity of presentation assume that Q is approximately the same for all respondents from the same service areas within a city:

$$Q = \begin{cases} q_1 \text{ if respondent is from service area 1} \\ q_2 \text{ if respondent is from service area 2} \\ q_3 \text{ if respondent is from service area 3} \end{cases}$$

Equation (1) can now be rewritten by replacing Q with dummy variables, allowing area 3 to be the reference:

$$S_1 = \begin{cases} 1 \text{ if respondent is from service area 1} \\ 0 \text{ other wise} \end{cases}$$

$$S_2 = \begin{cases} 1 \text{ if respondent is from service area 2} \\ 0 \text{ other wise} \end{cases}$$

Equation (1) therefore becomes:

$$Y = \alpha + \beta q_3 + (\beta q_1 - \beta q_3)S_1 + (\beta q_2 - \beta q_3)S_2 + \gamma X + \varepsilon$$

Finally, let $\alpha^* = \alpha + \beta q_3$, $\lambda_1 = \beta q_1 - \beta q_3$, and $\lambda_2 = \beta q_2 - \beta q_3$, yielding:

$$Y = \alpha^* + \lambda_1 S_1 + \lambda_2 S_2 + \gamma X + \varepsilon \qquad (2)$$

The analyst can now estimate equation (2), requiring only data for X and Y, and knowledge of which respondents live in which service areas. Even though Q remains unobserved, the estimates for λ_1 and λ_2 allow comparison of the relative service quality in the service areas. Since area 3 is the reference, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are the estimated differences in Y attributable to living in area 1 and area 2, respectively, compared to area 3. Thus, to compare area 1 or area 2 to area 3, we compare $\hat{\lambda}_2$ to $\hat{\lambda}_2$. These comparisons rank the service areas, as well as show the relative magnitude of the differences. Stated more formally, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ allow construction of estimates of a linear transformation of the Q values for the areas. Because these estimates were obtained by estimating equation (2), the possible confounding effects of X have been removed, and the estimates correctly reveal (subject to sampling fluctuation) the areas' ranking and relative differences in service quality.

In contrast, the means for Y within each area may yield misleading information about the ranking and relative differences in service quality. This potential to mislead results from the possible confounding effects of X, which can formally be seen by taking the expectations of the sub-sample means:

$$E(\bar{Y}_1) = E(Y|S_1 = 1) = \alpha + \beta q_1 + \gamma E(X|S_1 = 1) + E(\varepsilon)$$
$$E(\bar{Y}_2) = E(Y|S_2 = 1) = \alpha + \beta q_2 + \gamma E(X|S_2 = 1) + E(\varepsilon)$$
$$E(\bar{Y}_3) = E(Y|S_1 = S_2 = 0) = \alpha + \beta q_3 + \gamma E(X|S_1 = S_2 = 0) + E(\varepsilon)$$

Whenever X is related to the area a respondent is from, then $E(X|S_1 = 1)$, $E(X|S_2 = 1)$, and $E(X|S_1 = S_2 = 0)$ are not all equal. Therefore, the differences in the X values will affect the sub-sample means and be confounded with the effects of the q_i. Attributing differences in $\bar{Y}_i$ solely to differences in Q can consequently lead to erroneous interpretations.

As the above example illustrates, average expressed satisfaction levels can mislead whenever factors (besides service performance) that affect satisfaction are related to the areas being compared. Only under specific conditions do these statistical problems disappear:

1) Service performance is the only determinant of satisfaction.
2) No determinants of satisfaction, other than service performance, are related to the geographic areas being compared.
3) The determinants of satisfaction, other than service performance, cancel each other and have no net effect on satisfaction within each area.

If any one of these conditions applies, the analyst can forgo more complex multivariate statistics, and simply compare marginal distributions or averages for the different areas. However, these assumptions are probably unrealistic in most situations. In general a preferable approach is to apply multivariate statistical procedures, using the respondent as the case for analysis and including all important respondent-level variables, as in equation (2).[11]

The same statistical issues arise when comparing groups of citizens defined in ways other than geographic area of residence. For example, say a policy analyst compares expressed satisfaction with a particular service for black and white sub-samples. Imputing overall differences in satisfaction to differences in service performance different

racial groups receive can mislead whenever other factors affecting satisfaction are related to race. To illustrate, assume that level of education and income strongly affect expressed satisfaction, and that blacks in the city tend to be less educated and of lower income; thus, the effects of education and income would be confounded with the effects of service performance in the satisfaction the groups express. Only when one of the three conditions listed above (for groups from geographic areas) applies to the non-geographic groups do these statistical problems disappear. Otherwise, the analyst must use multivariate statistical procedures to remove the effects of the confounding variables.

. . . within the range of service quality typically found in U.S. metropolitan areas, citizens pay little attention to services and fail to perceive differences in service quality . . .

Unfortunately, multivariate statistics cannot always solve the difficulties of nonexperimental inference. Important respondent-level variables, such as differences in individuals' expectations for service performance, may go unmeasured. The analyst can then proceed only by assuming that other individual-level variables for which data are available, such as demographic variables, correlate highly with the unmeasured variables and can therefore serve as proxy variables in the analysis. Even more seriously, differences in service performance may be confounded with other differences in the service areas, such as neighborhood characteristics, that affect expressed satisfaction. Whenever important individual-level or other (e.g., neighborhood-level) variables go unmeasured and cannot be accounted for in the statistical analysis, multivariate procedures cannot remove these threats to the validity of using subjective data for performance evaluation.

Similar analytical problems can arise when time-series data are used to assess changes in service quality. For example, a policy analyst might compare expressed citizen satisfaction with a service at two different time points to assess the impact of program changes implemented in the interim.[12] The validity of this over-time comparison rests on the assumption that no important changes have occurred in other factors affecting satisfaction. This assumption is analogous to the assumption, when analyzing cross-sectional data for one time point, that no determinants other than service quality are related to the areas being compared. Over long time-periods this assumption will usually be untenable, but over short periods typically of interest to decision makers it may often be reasonable.

Recommendations

To measure the quality of service performance, policy makers should not rely heavily on survey items asking citizens how satisfied they are with particular local services, or asking citizens to evaluate particular local services. The meaning of such indicators is not clear, and some prior research suggests that responses to satisfaction or evaluation items may not accurately reflect the actual service government provides. Also, the potential analytical difficulties are great, and are not always soluble. Using satisfaction data to assess service performance may therefore mislead more than enlighten.

Whenever analysts do use subjective data to evaluate service performance, they should remain aware of the conceptual and statistical complications. Assumptions which underlie the analysis should be stated explicitly in reports to decision-makers. For example, what assumptions about how citizens responded to the survey items are necessary to support the inferences made about service performance? Also, overall satisfaction measures—such as average satisfaction or the percentage satisfied or dissatisfied—should be compared only under special conditions, as described earlier. Otherwise, the analyst should if possible employ multivariate statistical procedures to avoid artifactual conclusions.

More specific subjective indicators than general satisfaction with a service may offer greater potential for evaluating service performance. Even if data from general satisfaction or evaluation questions are 1) closely linked to actual performance, and 2) subjected to sophisticated statistical analysis, they still may confound different aspects of performance in one indicator. Responses to vague satisfaction or evaluation questions probably reflect at best some unknown mixture of different aspects of service provision. Whenever some of these aspects can be measured by more objective questions, perhaps the role of attitude or opinion items should be restricted to measuring more specific performance characteristics too intangible to measure in other ways.

. . . citizens living in neighborhoods with dilapidated housing may tend to generalize their dissatisfaction with housing conditions to all aspects of the neighborhood, including local governmental services.

For example, a vague survey item asking citizens how satisfied they are with police services may reflect a wide variety of factors, such as whether the respondent was 1) a victim of a crime recently, 2) stopped by local police recently, and 3) if he was stopped, how the police treated him. Information concerning the first two factors could be obtained directly by factual questions. An evaluation item which specifically asks the respondent how politely he was treated might be used to measure the third factor. Measuring these different factors separately, rather than confounding them in one general subjective indicator, probably would provide more useful information for evaluating service performance.

Policy makers might profitably use data from general satisfaction and evaluation items for purposes other than performance evaluation. Community attitudes themselves may affect service provision: for example, widespread feelings of dissatisfaction with police may lower citizen cooperation with law enforcement personnel. Variations in satisfaction across time, geographic areas, or demographic groups may consequently affect performance of the police functions, and affect the choice of the best standard operating procedures for law enforcement personnel to use. The sophisticated policy maker, in short, may find some use for general satisfaction or evaluation data, but will exercise caution in making inferences about the quality of service performance.

Notes

1. See Kenneth Webb and Harry P. Hatry, *Obtaining Citizen Feedback: The Application of Citizen Surveys to Local Governments* (Washington, D.C.: Urban Institute, 1973), and Louis H. Blair and Alfred I. Schwartz, *How Clean is our City?* (Washington, D.C.: Urban Institute, 1972).
2. Webb and Hatry, *op. cit.*, pp. 20-22, suggest such a procedure. Similarly, Shin infers from aggregate satisfaction data for different localities the relative service quality the localities receive, and goes on to recommend redistributing service resources. See Doh C. Shin, "The Quality of Municipal Service: Concept, Measure and Results," *Social Indicators Research*, Vol. 4 (May 1977), pp. 218-220, 225.
3. Brian Stipak, *Citizen Evaluations of Urban Services as Performance Indicators in Local Policy Analysis*, Ph.D. dissertation, University of California, Los Angeles (1976).
4. The statistics given in Tables 1 and 2 are estimated partial regression coefficients and their standard errors, appropriately transformed to show the estimated impact of the changes listed in Tables 1 and 2 in standard deviation units of the evaluation scale. The results in each table are based on more than one regression equation, because of data availability for the policy variables and other reasons. For a complete description of model specification and estimation results the technical reader should consult Stipak, *op. cit.* The statistics in Tables 1 and 2 were calculated based on Stipak, *op. cit.*, Tables 5.1, G.1, G.2, and G.10.
5. George L. Kelling, Tony Pate, Duane Dieckman, and Charles E. Brown, "The Kansas City Preventive Patrol Experiment: A Summary Report," in Gene V. Glass, ed., *Evaluation Studies Review Annual* (Beverly Hills: Sage, 1976), pp.631-637.
6. *ibid.*, p. 637.
7. E.g. Howard Schuman and Barry Gruenberg, "Dissatisfaction with City Services: Is Race an Important Factor?" in Harlan Hahn, ed., *People and Politics in Urban Society* (Beverly Hills: Sage, 1972).
8. For a discussion of information levels and opinion formation in the general public see Philip E. Converse, "Public Opinion and Voting Behavior," in Fred I. Greenstein and Nelson W. Polsby, eds., *Handbook of Political Science* (Reading, Mass.: Addison-Wesley, 1975), pp. 79-83. For a discussion of meaningless responses to attitude questions see Philip E. Converse, "Attitudes and Non-Attitudes: Continuation of a Dialogue," in Edward R. Tufte, ed., *The Quantitative Analysis of Social Problems* (Reading, Mass.: Addison-Wesley, 1970).
9. See Brian Stipak, "Attitudes and Belief Systems Concerning Urban Services," *Public Opinion Quarterly*, Vol. 41 (Spring 1977), pp. 50-51, for a discussion of processes of local political attitude formation in the absence of strong perceptions.
10. For a more detailed discussion of these complications see Brian Stipak, "Are There Sensible Ways to Analyze and Use Subjective Indicators of Urban Service Quality?", *Social Indicators Research* (forthcoming). Also available as Discussion Paper No. 121, Institute of Public Policy Studies, University of Michigan (1978).
11. Stipak, *ibid.*, presents simulation examples which contrast interpretations based on multivariate statistics, as opposed to interpretations based on sub-sample means. The underlying statistical issues are essentially problems of specification error and resulting statistical bias, a topic which is discussed in the econometrics literature.
12. E.g., see Harry P. Hatry, Richard E. Winnie, and Donald M. Fisk, *Practical Program Evaluation for State and Local Government Officials* (Washington, D.C.: Urban Institute, 1973), p. 102.