EPA-670/2-74-082
November 1974

Environmental Protection Technology Series

MEASURES OF EFFECTIVENESS FOR REFUSE STORAGE, COLLECTION, AND TRANSPORTATION PRACTICES

National Environmental Research Center
Office of Research and Development
U.S. Environmental Protection Agency

EPA-670/2-74-082
November 1974

MEASURES OF EFFECTIVENESS FOR REFUSE STORAGE, COLLECTION, AND TRANSPORTATION PRACTICES

By

MESSER ASSOCIATES, INC.
Silver Spring, Maryland 20910

Program Element No. 1DB063

Project Officer

Albert J. Klee
Solid and Hazardous Waste Research Laboratory
National Environmental Research Center
Cincinnati, Ohio 45268

NATIONAL ENVIRONMENTAL RESEARCH CENTER
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
CINCINNATI, OHIO 45268

REVIEW NOTICE

The National Environmental Research Center--Cincinnati has reviewed this report and approved its publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

FOREWORD

Man and his environment must be protected from the adverse effects of pesticides, radiation, noise and other forms of pollution, and the unwise management of solid waste. Efforts to protect the environment require a focus that recognizes the interplay between the components of our physical environment--air, water, and land. The National Environmental Research Centers provide this multidisciplinary focus through programs engaged in

- studies on the effects of environmental contaminants on man and the biosphere, and
- a search for ways to prevent contamination and to recycle valuable resources.

This report presents results of a project for that focussed on the systematic development of a set of measures and measurement tools that could be used to assess the effectiveness of solid waste storage, collection, and transportation practices. The measurement system presented is intended to support municipal decision-makers who have responsibility for such services as mixed refuse collection, street and alley cleaning, sanitary code enforcement, sanitation education, and other related activities.

Andrew W. Breidenbach, Ph.D.
Director
National Environmental
Research Center, Cincinnati

ABSTRACT

Perhaps between 75 to 80 percent of a solid waste system cost is due to storage, collection, and transportation, the remainder being attributable to disposal. Given an adequate accounting system, the monetary costs of a solid waste management system are much easier to compute than are the benefits produced and the nonmonetary cost incurred. Thus, although a community may have an accurate estimate of what it is spending upon its system, it often is uncertain as to whether or not it is receiving reasonable value in benefits returned; i.e., it has little or no idea of its "cost effectiveness."

This report presents the results of a project that focussed on the systematic development of a set of measures and measurement tools that could be used to assess the effectiveness of solid waste storage, collection, and transportation practices. The project included a pilot test of the measurement methodology in an urban community.

The measurement system presented in this report is intended to support municipal decision-makers who have responsibility for such services as mixed refuse collection, street and alley cleaning, sanitary code enforcement, sanitation education, and other related activities. It provides a model or prototype that municipal representatives can use to design effectiveness measures that are specific to their own solid waste management needs and activities.

The report includes a comprehensive list of candidate effectiveness measures along with the measurement techniques and sampling procedures needed to collect data to formulate the candidate measures. It also includes methods for combining individual measures into overall effectiveness indices.

This report was submitted in fulfillment of Contract Number 68-03-0260 by Messer Associates, Inc. under the sponsorship of the Environmental Protection Agency. Work was completed as of June 1974.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The project team wishes to extend its appreciation to the Project Officer, Dr. Albert J. Klee of the National Environmental Research Center, for his guidance and assistance during the conduct of this project. Additionally, we wish to express our gratitude to the following municipal government personnel in the City of Baltimore for their help and cooperation during the pilot demonstration phase of the project.

- Dr. Pierce Linaweaver, Director of Public Works, Department of Public Works
- Mr. R.G. Deitrich, Director of Technical Services, Department of Public Works
- Mr. George Schucker, Assistant Commissioner of Health, Director of Sanitary Services, Department of Health
- Mr. C. Edward Sachs, Director of Bureau of Environmental Hygiene, Department of Health
- Mr. Charles A. Carroll, Chief of Rat Eradication Program, Department of Health
- Mr. Neil Curran, Department of Planning
- Mr. G.L. Neff, Head of Bureau of Utility Operations, Department of Public Works
- Mr. Edward J. Moore, Chief of Division of Sanitation, Department of Public Works
- Mr. George Winfield, Technical Services, Department of Public Works
- Mr. Reuben Dagold, Bureau of Environmental Hygiene, Department of Health
- Mr. Mark Forester, Department of Planning

SUMMARY OF FINDINGS AND RECOMMENDATIONS

This report presents the results of a project that focussed on the systematic development of a set of measures and measurement tools that could be used by solid waste management agencies to evaluate the effectiveness of their solid waste operations of storage, collection, and local transportation.

The objectives of the project were to develop:

- Usable measures for assessing the effectiveness of the solid waste operations of storage, collection, and local transportation.
- Measurement techniques and sampling procedures to obtain data to formulate the measures.
- Methods for combining measurements of individual variables to yield overall effectiveness indices.

It was not the intent of this project to develop a rigid set of measures to be used by all agencies. Rather, the intent of the project was to develop a measurement system that could serve as a guide (or prototype) for various communities in the development of indicators that are specific to their own solid waste management needs and activities. In addition to providing a model that local communities could adapt and/or tailor to their needs, the project was intended to provide a mechanism that could be used by state and federal solid waste management agencies for comparing the effectiveness of the solid waste activities performed by jurisdictions under their control.

To ensure that the measures and the accompanying scheme for collecting the requisite data would be of use to solid waste managers, the project was to include a field test of the measurement methodology. This test was to take place in an urban community.

MAJOR FINDINGS OF THE PROJECT

The findings and conclusions that were developed on the basis of the data collected during the field demonstration phase of the project were of two types: findings of a general nature and findings related to the sample design.

The findings which are general in nature may be summarized as follows:

- The number of variables measured can be reduced because many of the variables were frequently found to be at their lowest value.
- Only three observational areas need to be inspected—blockfaces, alleys, and lots.
- For blockface measurements, one blockface selected at random can be used in lieu of all four blockfaces to measure the effectiveness variables.
- Observers exhibit a high degree of consistency for the "yes-no" type measurements and for "counts."
- Observers exhibit a fair degree of consistency for the more subjective-type measurements; i.e., glass, garbage, and refuse ratings. The amount of variation is less at lower scale points than at higher scale points, tending to rapidly increase and then stabilize.
- Variation among observers, however, only accounts for approximately 15 to 20 percent of total variation when measuring a tract mean.
- On the whole, observers tend to be accurate to within one-half of a scale point for the glass, garbage, and refuse ratings; a few, however, were always high or always low in their assigned ratings.
- There are a number of statistically significant correlations among the variables; however, the explained variation tends to be low.

- It is possible to develop composite measures of effectiveness, using multivariate techniques; however, the refuse rating by itself can serve as a proxy for an overall measure.

In summary, these findings indicate that an ongoing measurement system would require collection of data on only a few variables; that subjective measurements of unsightly conditions, health hazards, safety hazards, and so forth can be made, provided that one is willing to accept a small amount of inconsistency in the measurements; and that formulation of composite effectiveness indices is possible, but perhaps unnecessary.

In addition to the general findings described above, there were several findings that related to the sample design. They may be summarized as follows:

- Large differences in the mean tract ratings were found to exist between two groupings of the strata.
- The sampling plan was adequate to detect changes of one-half a point or less in the blockface ratings; larger samples would be required to detect an equivalent change in the alley ratings.
- The mean tract ratings did not appear to be biased by the day of the week when the inspection was made.

MAJOR RECOMMENDATIONS OF THE PROJECT

The recommendations stemming from this project are of two types:

- General recommendations on how to develop a measurement system that is specific to a given community.
- Detailed recommendations on how to implement an ongoing measurement system using the findings from the field test data.

The recommendations on how to develop a measurement system that is specific to a given community are as follows:

(1) Review the list of measures and measurement techniques provided in Chapter III of the report and select those measures most useful.

(2) Use the basic survey design developed for this project to obtain preliminary data on those measures that require direct observation of existing conditions.

(3) Utilize several observers in each tract and have them make the same measurements.

(4) Apply techniques similar to those used in this project to determine the appropriate sample size for an ongoing measurement system and the relevant variables for which measurements should be made.

The recommendations on how to implement an ongoing measurement system that draws upon results of the field demonstration conducted in the City of Baltimore are as follows:

(1) Collect data on blockfaces, alleys, and lots only.

(2) Sample approximately ten blocks in each census tract where the overall conditions are bad to detect a change of one-half a point or less in the blockface garbage, glass, and refuse ratings; inspect fewer blocks in areas where the overall conditions are good. Use the income level of the census tract as an initial means for classifying conditions.

(3) Inspect only one randomly selected blockface in each block; inspect all alleys and lots in the block.

(4) Utilize several observers and have them inspect different blocks in the same census tract in order to reduce the variation associated with inconsistency among observers.

(5) Periodically compare the observations of raters within the same tract to see if any of the observers are consistently high or consistently low.

(6) Make measurements only of the amount of refuse found in these areas. Use this as an indicator of overall conditions.

<u>or</u>

Make measurements of the amount of refuse, glass, and garbage found in the areas, the presence of rat signs (alleys only) and the number of bulk items (alleys only). Report the measurements separately and/or as a composite measure.

(7) Report the location of fire hazards, bulk items, abandoned vehicles, clogged basins, and other items of interest so that corrective action can be taken.

I. INTRODUCTION

In many communities, information is available to determine the costs of solid waste services; i.e., the labor and equipment costs associated with collection of mixed refuse, street sweeping and cleaning, and refuse disposal. Similarly, information is generally available to assess the operational efficiency of solid waste operations, such as tons of refuse collected per man-hour, miles of streets cleaned per man-hour, and so forth. However, information by which to evaluate the effectiveness of sanitation activities is typically lacking; i.e., information on the degree to which streets and alleys are kept free from debris so as to prevent conditions harmful to the health and safety of the public and to promote an aesthetically pleasing environment.

Opinions on sanitation conditions are usually reported by sanitation crew supervisors. Unfortunately, this is usually a sporadic and subjective process, characterized by the lack of well-defined measures that have been agreed upon, and the absence of a structured mechanism to relate this information to data on costs and operating activities. The need to clearly delineate the many variables that are affected by solid waste operations (particularly storage, collection, and transportation activities) and develop usable means of quantifying and measuring changes in these variables is the primary reason why this project was undertaken.

This chapter describes the objectives of the project and the potential management uses of effectiveness indicators. It also provides an overview of how we performed the project and how this report is organized.

OBJECTIVES OF THE PROJECT

The primary purposes of this project were to develop:

- Usable measures to assess the effectiveness of solid waste operations, with emphasis on the functions of storage, collection, and local transportation.

- Measurement techniques and sampling procedures to obtain data to formulate the measures.

- Methods for combining measurements of individual variables to yield overall effectiveness measures.

The basic thrust of the project was on the systematic development of a set of indicators and appropriate measurement tools that could be used by solid waste management agencies to evaluate the effectiveness of their solid waste operations.

It was not the intent of this project to develop a rigid set of measures to be used by all agencies. Rather, the intent of the project was to develop a measurement system that could serve as a guide (or prototype) for various communities in the development of indicators that are specific to their own solid waste management needs and activities. This is a particularly important consideration because each different solid waste management agency is likely to have its own ideas on what constitutes meaningful effectiveness indicators. The operating characteristics of the local agency and the environmental characteristics of the community that it serves may require a unique set of indicators. Furthermore, the adequacy of current information systems and the corresponding ability of individual agencies to generate measures may vary widely across locales.

In addition to providing a model that local communities could adapt and/or tailor to their needs, the project was intended to provide a mechanism that could be used by state and federal solid waste management agencies for comparing the effectiveness of the solid waste activities performed by jurisdictions under their control. By adopting a standardized reporting system that requires a sufficient degree of conformity in the types of data that are collected and in the data collection procedures, it would be possible for these agencies to compare the effectiveness of local solid waste management activities. This, in turn, would provide state and federal agencies with information to facilitate their planning, financing, and regulatory responsibilities.

To ensure that the measures and the accompanying scheme for collecting the requisite data would be of use to solid waste managers, the project was to include a field test of the measurement methodology. This test was to take place in an urban community.

MANAGEMENT USES OF EFFECTIVENESS MEASURES

The measures developed during this project provide solid waste managers with quantitatively based feedback information on how well the goals and objectives of their solid waste activities are being met, particularly storage, collection, and local transportation activities. The measures indicate whether changes may be needed in the underlying policies that govern a given mode of operation (e.g., frequency of collection, type of collection service, level of sanitary enforcement activities, etc.) or in the level and mix of resources devoted to these activities. Where a change is undertaken, they provide a means for assessing whether this change has produced the desired impact.

Effectiveness measures are particularly useful in:

- Determining those service areas most in need of corrective action.
- Assessing the impact of changes in policy and in the allocation of resources among various service areas.
- Assessing the impact of special sanitation operations (e.g., anti-litter campaigns, special cleanup campaigns, etc.).
- Preparing annual budgets and justifying budgetary appropriations for solid waste activities, particularly when additional resources are needed.
- Establishing standards for evaluating the performance of private waste collectors.
- Justifying to citizen groups the type and level of service that their area is receiving.

However, as illustrated in Figure 1 on the following page, effectiveness indicators are only one of several ways that managers have of evaluating solid waste activities. Other indicators include performance or output measures (tons of refuse collected per week, number of streets cleaned per day), and productivity or efficiency measures (tons of refuse collected per man-hour). These three

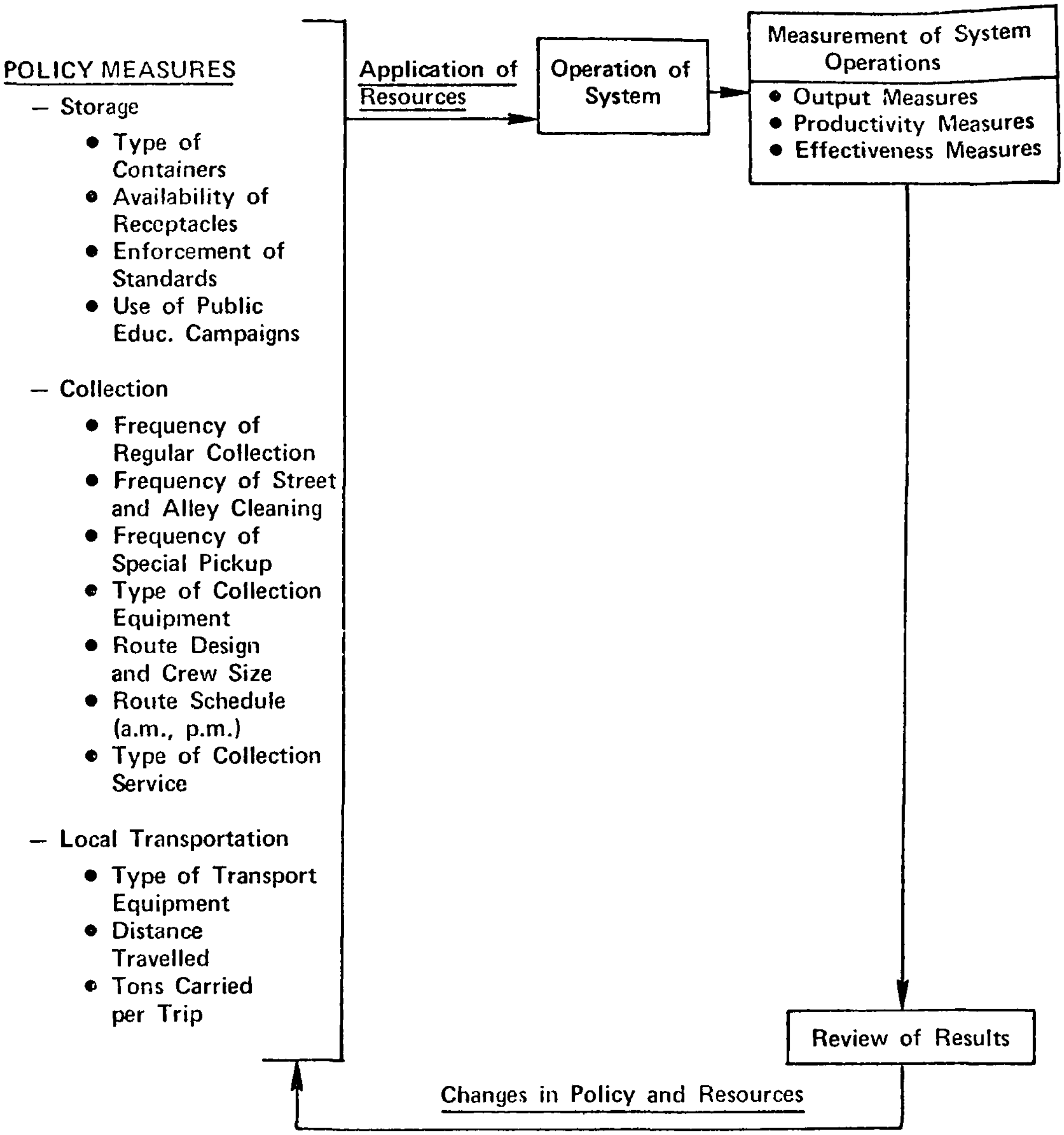


Figure 1: "Measure-Act-Measure" Mechanism for Evaluating Solid Waste Activities

types of indicators — effectiveness, efficiency, and output measures — are not unrelated. For this reason, changes in policy variables or in the level of resources that appear warranted on the basis of effectiveness measures should always be reviewed in light of their potential effect on system performance and system productivity.

A DESCRIPTION OF HOW WE PERFORMED THE PROJECT

The project was performed in several phases that consisted of the activities summarized below:

- Phase I — Collection and Review of Background Materials. A review of the literature on effectiveness measurement systems was undertaken and the solid waste managers in each of the 10 largest cities were contacted to determine the types of effectiveness measures, if any, that their communities were using. Interviews and discussions were also held with numerous persons involved in solid waste management activities to obtain their views on the relative importance and usefulness of the various measures.

- Phase II — Methodology Development. An analytical framework was developed from which a comprehensive set of effectiveness indicators was derived, along with methods for collecting and recording the data. Additionally, the various combining methods were reviewed and a procedure for developing composite measures was selected.

- Phase III — Field Testing of the Measurement Methodology. A field demonstration, designed to assess the measurement methodology and its implementation in an urban community was conducted. The City of Baltimore was used as the field test site.

- Phase IV — Analysis of the Field Test Data. The field test data were analyzed to assess the feasibility of producing the measures, the consistency or reliability of the subjective-type measurements, the correlations among the measures, and the feasibility of producing composite measures of effectiveness for solid waste activities.

ORGANIZATION OF THE REPORT

The report is organized into seven chapters and covers the activities associated with the phases listed above. The first chapter provides introductory material relevant to the project. The second chapter describes the findings that emerged from a review of background and other materials on existing methodologies designed to measure the effectiveness of solid waste operations. The development of the methodological approach and combining techniques used in the project are described in the third and fourth chapters. Details on the field demonstration comprise the fifth chapter. The last two chapters summarize the findings and recommendations that were developed on the basis of the field data.

In addition, the report includes six appendices. These provide additional information on the methodological approach, survey design, and field test data collection forms and procedures that were used in the project. They include a description of how the analysis of the field test data was performed, as well as detailed tables, charts, and graphs to support some of the findings presented in the body of the report.

II. BACKGROUND MATERIALS REVIEW

A number of measurement systems have been developed in recent years to assess the effectiveness of solid waste operations. Some of these systems or variations thereof have been implemented in several of the larger cities across the nation. This chapter describes the general types of measurement methodologies that have been proposed to collect effectiveness-type information, their respective uses in the 10 largest cities in the country, and the relationship of this project to existing measurement systems.

TYPES OF MEASUREMENT METHODOLOGIES

The measurement methodologies developed during the past few years generally include one or more of the following elements:

- Direct assessment of conditions, based upon physical measurement or direct observation by trained persons.
- Indirect assessment of conditions, based upon data obtained from records or ledgers.
- Citizen perception of conditions, as indicated in complaint records or special attitudinal surveys.

These measurement methodologies are discussed in the subsections that follow.

Direct Assessment of Conditions

Nearly all of the proposed measurement systems call for direct measurement of existing conditions, particularly along streets and alleys in the community. However, for the most part, only the overall aesthetic conditions of these areas are considered when measurements are made. Although provision is generally made within most systems for collecting information on other items of interest (e.g., bulky wastes), there is, as a rule, no attempt to include these other items as formalized measures *per se* (e.g., the number of discarded bulk items per square mile).

The existing methods for directly assessing conditions fall into one of two categories:

- Measurement of the area covered by litter
- Measurement of the volume of litter in a given area.

These methods are described in the following paragraphs.

Measurement of the Area Covered by Litter

Two techniques have been proposed for measuring the area that is covered by litter. One scheme, called a Visual Inspection System, was developed by the Urban Institute.[1] The other scheme, called a Photometric System, was developed by the American Public Works Association, in conjunction with Keep America Beautiful.[2]

The Visual Inspection System consists of a set of procedures whereby a trained inspector gives a numerical rating on a scale of 1 to 4 to the litter conditions observed on a street or alley. A set of photographs, scaled to illustrate the range of litter conditions, is used as a standard in making these ratings. Whenever the conditions on streets and alleys fall in between the conditions illustrated in the photographs, intermediate rating points of 1.5, 2.5, and 3.5 are utilized. The originators of this scheme recommend that inspectors, along with rating the cleanliness and appearance of streets and alleys, note factors such as the presence of abandoned automobiles and health and fire hazards, so that corrective action may be taken.

The Photometric System is an attempt to develop a more objective way of measuring aesthetic conditions in streets and alleys. Under this method, photographs are taken of the actual solid waste accumulations at random points along a sample of streets and alleys. The camera must be placed so that all photographs are taken at the same angle, from the same distance, and represent a 6- by 16-foot rectangular area.

After the pictures are developed, a clear plastic overlay, marked off into 96 grids, is placed over each picture. The grid overlay is designed so as to have the same perspective as the photographs. This is necessary in order to compensate for the fact that things in the foreground tend to look larger and more important than those in the background. By counting the number of grids that contain litter, the total area covered by litter can be determined. The originators of this scheme recommend that the number of grids containing litter be converted to the following six-point rating scale:

Rating	Number of Grids Containing Litter
1	0-4
2	5-10
3	11-20
4	21-30
5	31-40
6	41 or more

Both of these two procedures for assessing the area covered by litter have certain inherent problems. The visual inspection procedure relies heavily on a subjective assessment of conditions, and, as such, is subject to inspector biases and inconsistencies. The photometric procedure, while it has much to recommend it, requires that a substantially greater amount of time be spent for its implementation. Because the pictures must all be taken from the same perspective, a fair amount of set-up time is required. Second, the area photographed must be one that is free of automobiles for at least 25 to 30 feet. This often poses difficulties when it is desired to obtain a second measurement to assess changes in conditions. Third, difficulties may be encountered in the actual counting of the littered grids because: dark pieces of trash are not easily distinguishable in shadows; wet areas cause a glare to the picture, making some of the trash indistinguishable; areas around broken pavement are often discolored and appear as though they might be litter; and, there is human error in counting the number of littered squares. Fourth, the photographs are not readily interpretable without a magnifying glass.

Measurement of the Volume of Litter

As an alternative to measuring the area covered by litter, a technique has been developed for measuring the volume of litter in a given area. This technique, developed by Ralph Stone and Company, Inc., requires that all of the trash along a street or alley be swept up, carried away, and then measured.[3] This scheme, while it removes much of the subjectivity associated with measuring the overall appearance of an area, is a fairly costly scheme to implement on an ongoing basis, as well as one that requires the absence of parked cars, trucks, and other vehicles along the street.

Indirect Assessment of Conditions

A few of the proposed measurement schemes suggest utilizing information on area conditions available from records and ledgers to assess the effectiveness of solid waste activities. Some of the proposed indicators include the number of trash fires from solid waste accumulations, the level of external rat infestation, the number of missed collections, and so forth.

Formulation of these measures generally requires a fair degree of cooperation among the personnel in different municipal government agencies (e.g., health department, fire department, sanitation department, housing department, etc.). This is because the data that are needed are not available solely within the sanitation agency. Moreover, the data, when accessible, are not always comparable. For example, the sanitation department may collect and summarize information using sanitation districts as a basis, while the health and fire departments may utilize different service units (more suited to their own internal needs) in their summarization of the data. Additionally, even when the data bases are comparable, the level of summarization may be too aggregative to allow comparisons except among large geographic areas within a city.

For these reasons, implementation of a measurement system that utilizes available data will generally require a revision of an existing management information system or the development of a new information system. It will also require a system which encompasses involvement by more than just the sanitation department, because of the variety of data that would be included.

Citizen Perception of Conditions

Some of the proposed measurement schemes suggest using citizen input to assess the effectiveness of sanitation activities. Citizen input can be obtained from:

- A systematic review of complaint records on file with the sanitation agency.
- A special survey of citizens to obtain their attitudes about sanitation activities.

Complaint data are useful in pinpointing short-term or extreme problems. Long-standing deficiencies are generally less detectable from complaint data for two reasons: (1) persons having complaints may "give up" if remedial action is not initiated early on; and, (2) persons may become so accustomed to the deficiencies that they do not register complaints.

One of the problems associated with using complaint records is that generally only a handful of citizens bother to file formal complaints. Because the views of these persons may not be representative of the community-at-large, there are likely to be biases associated with using complaint data to measure the overall effectiveness of sanitation activities. A carefully designed citizen attitude survey will, as a rule, provide more accurate and reliable information about citizen perception of conditions than will volunteered complaints.

Citizen attitude surveys can be conducted by interviewing a sample of residents either in-person, over the telephone, or through the mail. The choice among survey methods depends on the desired degree of accuracy and the amount of money the community is willing to spend. The proponents of citizen attitude surveys recommend that data be gathered concurrently on citizen satisfaction with other municipal services as well as solid waste operations.

EFFECTIVENESS MEASUREMENT SYSTEMS IN THE TEN LARGEST CITIES

Summary information on the effectiveness measurement systems in use in the ten largest cities in the nation was collected at an early point in the study through telephone and personal interviews with sanitation department personnel. Where available, documentation related to the measurement systems was also obtained.

The review of measurement systems in use revealed the following:

- Only two of the cities—New York and Washington, D. C. —have implemented formalized procedures for systematically assessing the effectiveness of their solid waste operations. A third city, Baltimore, is planning to implement such a system in the near future. In all three cases, the assessment is based on the Visual Inspection System. All three measurement systems are funded by sources outside the sanitation department.

- In the remainder of the seven cities that were contacted—Cleveland, Chicago, Dallas, Detroit, Houston, Los Angeles, and Philadelphia—there are no formalized procedures for measuring the effectiveness of sanitation operations. Crew foremen and supervisors periodically inspect areas subsequent to collection and make a qualitative assessment of conditions, but no quantitative measures are utilized.

The effectiveness measurement systems for New York City, Washington, D. C., and Baltimore, are described below.[4,5,6]

New York City

Visual inspection procedures for rating the overall cleanliness of designated areas are being used in the City of New York. This system, called Project Scorecard, utilizes the techniques developed by the Urban Institute, with some modifications. Scorecard utilizes an eight-point scale, where 1.0 is immaculately clean and 4.5 is filthy. The scale progresses at .5 intervals to reflect intermediate levels of cleanliness.

Under the Scorecard system, an inspector rates both sides of the street and the adjoining sidewalks, but not the alleys. The location of abandoned cars and bulk trash items are noted. These are reported to the sanitation district superintendent responsible for the area so that remedial action can be taken.

The streets and sidewalks are assigned separate ratings, which are then averaged together to form a single rating for each block inspected. The block averages are subsequently combined and a new composite average is produced for what is called a "strip" (a linear set of blocks). These strips form the basic unit on which comparisons are made between sanitation districts. The ranking of sanitation districts, based on the cleanliness ratings, has resulted in minor competition among the various district superintendents to keep their areas as clean as possible, as none of them want to appear at the bottom of the list.

At the time that the information on Project Scorecard was obtained, the system was operational on Manhattan's Lower East Side and in areas designated by the City as needing concentrated cleaning attention. Inspections were being made several times weekly along all streets where the system was operational. Plans, however, were underway for expanding Scorecard's operations to permit inspection of a sample of streets in the entire city.

Washington, D. C.

Washington, D. C., has reinstituted a Visual Inspection System for assessing the appearance of streets and alleys within the city. A nine-point scale is used, with 1.0 representing the cleanest areas and 5.0 representing the dirtiest areas. The scale progresses at half-point intervals to reflect conditions in between the two extreme values.

The inspectors rate one side of the street, both sides of alleys, conditions on public property, and conditions on private property. Information is also recorded about dead animals, abandoned vehicles, clogged catch basins, bulky waste, fire hazards, and a number of other items of interest.

The entire city is inspected four times a year, and the information is summarized by census tract, by collection route, and by service area. In "clean" areas of the city, 20 to 30 percent of the blockfaces (the area along one side of the block from the center line of the street to the curb) and adjacent public and private ways are inspected. These are selected on the basis of random sampling techniques. In the "not so clean" areas, 80 to 100 percent of the blockfaces and the adjacent public and private ways are inspected. All alleys in the city are inspected, independent of whether the area is classified as clean or not.

The unique aspect of the Washington, D. C., system is that it is one of the few systems where the information that is collected is assimilated into an overall rating to produce an "Environmental Rating" for an area. A 100-point rating scale is used. Conditions on both public property and private property are included in formulating the index. The various index components and relative weights are as follows:

Item or Condition	Weight (in terms of points assigned)
Overall Blockface or Alley Cleanliness Rating	40
Blockface or Alley Cleanliness Rating with Leaves	10
Unsightly Conditions on the Public Way	10
Unsightly Conditions Beyond the Public Way	20
Presence of Special Items (e. g., dead animals, abandoned vehicles, bulk, etc.)	20
TOTAL	100

Based on the weighting scheme shown above, a composite rating is developed for each blockface or alley that is inspected. The individual composite ratings are averaged to give an overall rating for the area under consideration.

Baltimore

Baltimore is currently in the process of implementing a solid waste effectiveness measurement system. The system combines effectiveness measures with information on solid waste activities and costs, in an attempt to identify the types and costs of programs and resource combinations that have the greatest effectiveness. A visual inspection process on cleanliness and appearance of streets and alleys, along with a procedure to obtain information on sanitation violations and citizen complaints, constitute the effectiveness measurement system in Baltimore. The effectiveness measures will form part of a comprehensive sanitation management information system.

Baltimore has experimented with the Photometric System to produce cleanliness ratings and has decided that the visual inspection procedure is the more efficient method to use for their purposes for the following reasons. First, they experienced difficulties in finding the designated areas free from parked vehicles at the time the photographs were taken and in interpreting the photographs, even with a magnifying glass. Second, they found the method could not be applied during periods of inclement weather. Thirdly, they felt that the increased degree of accuracy (which the Photometric System provides) was not sufficient enough to justify the high costs associated with operating this type of system.[7]

GENERAL DEFICIENCIES IN EFFECTIVENESS MEASUREMENT SYSTEMS

The measurement methodologies presented in this chapter have proven useful in solid waste decision-making, as evidenced by their implementation and continued use in several cities. However, there are certain weaknesses in these measurement systems that necessitated the comprehensive investigation undertaken during this project. These weaknesses may be summarized as follows:

- The major thrust of these systems is on solid waste cleaning and collection activities as illustrated by the emphasis upon measures to rate cleanliness and appearance of streets and alleys. The storage function is considered only to a limited extent in that indicators on the condition of cans are included in the Baltimore and Washington measurement systems. No attention at all is given to the transportation function.

- The primary focus of these systems is on measuring the appearance of the area. Although provision is generally made for collecting ancillary information, routine measurement of factors such as health and safety hazards is generally not part of the measurement system. The potential impact of solid waste systems upon these factors may be considerable; their effects may, in fact, be opposite to their effects upon the more easily measured variables such as street cleanliness and appearance. Therefore, a comprehensive measurement system must include indicators for these factors.

- The systems generally do not provide a means for combining the individual measures into an overall effectiveness rating. The schemes being implemented in New York and Baltimore, provide indices of overall street and alley cleanliness, but that is the extent to which individual measures are combined. In Washington, D. C., where a composite rating of effectiveness has recently been developed, cleanliness factors receive 80 percent of the total weight. Without usable methods for combining the various solid waste indicators into overall measures of effectiveness, it is difficult to judge the true status of a community's solid waste operations or to establish comparisons with other communities.

- The systems are essentially "parochial" in that they were designed for specific communities to serve specific purposes. Although guidelines are generally provided for the extention of these systems to other cities, the selection of the various measures and the measurement techniques is influenced by the persons who originally developed the measurement methodologies.

This project, while building upon the available measurement systems, attempted to overcome the deficiencies cited above. Measures related not only to collection and cleaning activities but also to storage and transportation activities were considered. A fairly comprehensive set of indicators, encompassing health hazards, safety hazards, fire hazards, etc., was developed and applied during a field test. A method for combining disparate indicators into single effectiveness measures was developed and implemented using the data collected during the field survey. Finally, a set of procedures was designed to assist communities that want to develop their own unique measurement schemes, rather than adopt existing ones.

III. DEVELOPMENT OF MEASURES AND MEASUREMENT TECHNIQUES

There were two major aspects to the methodology development phase of the project—(1) the development of individual measures and measurement techniques for assessing the effectiveness of the solid waste operations of storage, collection, and local transportation, and (2) the development of analytical procedures for combining individual measures to produce an overall measure of effectiveness. This chapter describes how the measures and the measurement techniques were developed and provides a complete list of the "candidate" effectiveness measures, many of which were investigated during the pilot test phase of the project. The following chapter describes the development of the combining methods.

The procedures utilized in developing a comprehensive set of effectiveness measures may be summarized as follows:

- Definitions were developed for the solid waste operations included in the project.
- The scope of the study was defined relative to the various generic types of effects that could be measured.
- An analytical framework, consisting of measurement categories and indicators matched to the solid waste operations under study, was developed.

The remaining paragraphs discuss these procedures in greater detail.

TERMS RELEVANT TO THE PROJECT WERE DEFINED

Solid waste materials are those unused or unwanted materials that result from normal community activities and have insufficient content to be free flowing. The development of policies and procedures for controlling the generation, storage, collection, transport, separation, processing, recovery, and disposal of these materials is the responsibility of those federal, state, and local officials who oversee and operate the solid waste management system. This project focussed specifically on the effectiveness of policies and procedures related to the storage, collection, and local transport of solid waste materials.

The components of the solid waste management system that this project focussed on were defined as follows:

- Storage Operations — The methods by which solid waste products are discarded by individuals, households, and establishments, in either a loose or a contained form, prior to collection/removal by public or private haulers.
- Collection Operations — The methods associated with the gathering or accumulating of solid waste products from various storage points for the purpose of transporting them to a central waste depository site; i. e., a transfer station, reclamation center, or disposal site.
- Local Transportation Operations — The methods associated with the conveyance of solid waste products from a final collection point to the closest transfer station, reclamation center, or disposal site.

Storage practices thus reflect the mode in which solid waste materials initially present themselves to the system. This may be in a containerized form (from households, commercial establishments, etc.) or in a non-containerized form (such as street litter, bulk waste, etc.). Collection activities include: regular mixed refuse collection, bulk or special item pickup, and street and alley cleaning. Transportation practices refer solely to activities associated with the conveyance of solid waste products to the closest depository site. Long haul transportation via rail, truck, and so forth, was not part of this project.

These solid waste activities were considered relative to municipal solid waste products only; i. e., materials discarded by households, apartment complexes, and commercial concerns, and materials found in public areas such as streets, sidewalks, alleys, vacant lots, parks, etc. The types of solid waste materials likely to be generated by these sources include:

- garbage (not discharged into the sewer system)
- rubbish (both combustible products, such as paper, leaves, wood; and non-combustible products, such as metal, glass, plastic, and so forth)
- ashes
- street refuse (street sweepings, dirt, leaves, catch basin dirt, street litter, and so forth)
- bulky wastes
- abandoned vehicles
- construction and demolition waste.

ONLY THE EFFECTS OF SOLID WASTE ACTIVITIES ON THE PUBLIC WERE CONSIDERED

Conceptually speaking, there are two distinct types of effects which may be expected to result from the operation of a solid waste system, namely:

- Effects which are external to system operating procedures
- Effects which are internal to system operating procedures.

Effects external to the system refer to the benefits and damages which are sustained by the public as a result of solid waste system operations. An example of this type of effect is unsightliness caused by the presence of spilled or scattered refuse subsequent to mixed refuse collection.

Effects internal to the system refer to the benefits and damages that are sustained by the system itself as a result of a given type of solid waste operation or procedure. An example of this type of effect would be injuries suffered by waste collectors.

This project concerned itself only with the former type effect; i.e., effects external to the system.

A GENERAL FRAMEWORK WAS DEVELOPED

Once the focus of the study had been clearly defined, a general framework was developed from which a "candidate" set of effectiveness measures could be derived. The framework consisted of:

- Major measurement categories that corresponded to the objectives of a solid waste system.
- Qualitative indicators (or descriptors), matched to the solid waste operations under study, for each measurement category.
- Quantitative measures (or variables) for each indicator.
- Measurement techniques for each measure.

An overview of the general framework is presented in Figure 2 on the following page. The elements of the framework are described in the paragraphs that follow.

Several Major Categories for Measuring Effectiveness Were Developed to Reflect the Objectives of a Solid Waste Management System

In general, effectiveness measures relate to the degree to which goals and/or objectives are being met. Thus, in order to properly assess the effectiveness of the solid waste system, it is necessary to define its goals and/or objectives in a precise manner.

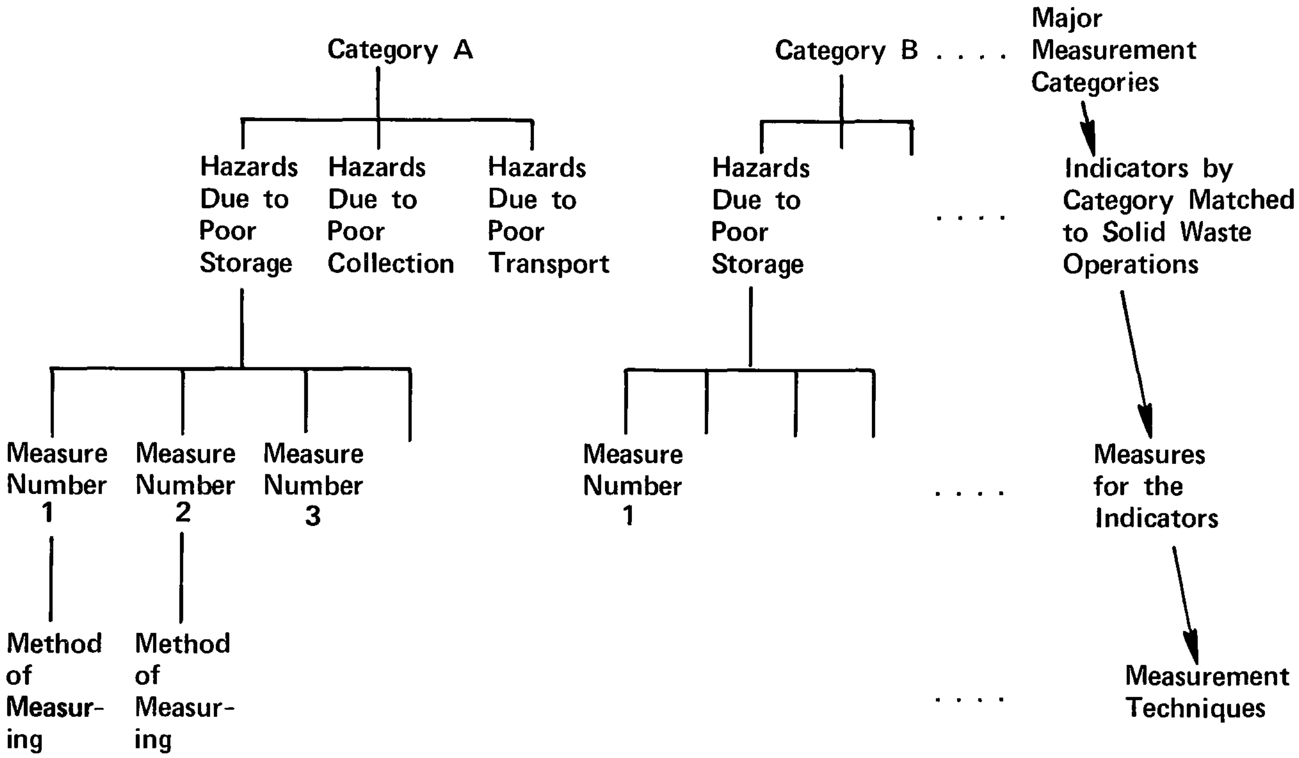


Figure 2: Overview of the General Framework for Developing Candidate Effectiveness Measures

Since the project was not meant to deal with a specific type of solid waste system, but rather with solid waste systems in general, the statement of objectives had to be broad enough to apply to any solid waste system, independent of its physical components. The statement of objectives that was developed may be summarized as follows:

> The objective of a solid waste management system is to provide for the operation of a waste handling system with a level of service sufficient to—
>
> (1) meet the community needs in terms of handling and disposing of all unused or unwanted solid waste materials that are discarded within the community
>
> (2) maintain a safe, healthful, and aesthetically pleasing community; i.e., counter those effects which would result from an unmanaged or improperly functioning solid waste system.

Based on this statement of objectives, it can be inferred that an effective solid waste system is one which:

- meets the community's storage and collection needs
- mitigates against conditions that might cause deterioration of public health
- acts to prevent accumulations of solid waste materials which might cause injury to the public
- strives to minimize deterioration of the appearance of the community due to street litter and solid waste materials
- acts to prevent offensive odors caused by inadequate storage or infrequent collection of solid waste materials.

Thus, the following effectiveness categories emerge from the statement of objectives:

- Meeting the Community's Storage and Collection Needs
- Safety of the Public
- Public Health
- Aesthetic Conditions
  - — Appearance
  - — Odor

These effectiveness categories relate directly to the reasons for performing solid waste collection and transportation activities; for example, keeping the streets clean and preventing public health and safety problems from occurring. They pertain to solid waste systems in general.

In order to accomplish the objectives stated above, each solid waste system sets its own ordinances, standards, and policies with respect to its storage, collection, and transportation operations. These standards and procedures generally relate to the types of allowable containers and to the frequency of collection and cleaning operations. Thus, the degree of compliance with these standards constitutes another major category against which effectiveness can be measured. That is, if plastic bags as outside storage containers are prohibited, then measurement of the compliance level for this could be important with respect to the overall effectiveness of the solid waste system.

In addition, there are features of the actual solid waste system in use (i.e., its equipment, labor use, and other methods and procedures) that are either closely related to meeting the overall objectives or are constraints (restrictions) on how the job of storing, removing, and transporting refuse should be accomplished. Some of these features may cause inconvenience or discomfort to the public. Collection truck noise levels and induced traffic congestion because of collection scheduling are examples of these types of operating system features.

Thus, for a specific solid waste system, two additional categories for assessing effectiveness may be set forth:

- Compliance with System Standards
- Inconvenience/Discomfort to the Public Because of System Operating Procedures.

Table 1 shown below summarizes the major measurement categories that were developed for this project.

Table 1. SUMMARY OF MAJOR CATEGORIES FOR MEASURING EFFECTIVENESS

CATEGORIES OF EFFECTIVENESS RELATED TO NATURE OF SOLID WASTE SYSTEMS IN GENERAL:

- Public Health
- Public Safety
- Appearance of Community
- Odor Within Community
- Satisfaction of Community's Storage and Collection Needs

CATEGORIES OF EFFECTIVENESS RELATED TO SOLID WASTE SYSTEM UTILIZED:

- Compliance With Standards
- Inconvenience/Discomfort to Public

Indicators of Effectiveness Matched to Solid Waste Operations Were Developed for Each Category of Effectiveness

For each of the categories of effectiveness listed in Table 1, a number of indicators were developed. These indicators were designed to provide a qualitative statement of conditions from which quantitative measures (or variables) could subsequently be developed. That is, they were used to describe the conditions that we would be attempting to measure.

In developing these indicators, an attempt was made to match effectiveness categories with the specific solid waste operations under study; namely, storage, collection, and transportation. This was done in order to facilitate the formulation of quantitative measures that related specifically to one or more of the operations.

A summary of the indicators of effectiveness, matched to solid waste operations, is shown in Table 2 on the following page.

Quantitative Measures and Measurement Techniques Were Developed for Each of the Indicators of Effectiveness

For each effectiveness indicator, measures (or variables) were developed for which data would be gathered. This initial list of variables was subsequently reviewed by an in-house team of analysts, and a number of measures were eliminated from further consideration because they did not meet one or more of the evaluative criteria set forth.

The evaluative criteria used to screen the initial list of variables included an assessment of each measure's:

- Validity—Does the measure indicate what it purports to?
- Accuracy—Will the measurement reflect a "true" picture of the measured conditions?
- Reliability (or) Consistency—Will independent observers come up with similar measurements of the same phenomena?

Table 2. INDICATORS OF EFFECTIVENESS FOR EACH MEASUREMENT CATEGORY, MATCHED TO SOLID WASTE SYSTEM OPERATIONS

	Measurement Categories	Indicators for Each Category Matched to Solid Waste System Operatons				
		Storage Operations	Collection/Cleaning Operations			Transportation Operations
			Regular Collection	Special Collection	Cleaning	
Related to Nature of Solid Waste Systems in General	PUBLIC HEALTH	● Conditions hazardous to health in storage areas/or private areas	● Widespread health hazards throughout neighborhood storage areas ● Health hazards in waste collection area		● Health hazards on the public way ● Health hazards on vacant lots and public areas	
	PUBLIC SAFETY	● Conditions hazardous to safety in storage areas/or private areas	● Widespread safety hazards throughout neighborhood storage areas		● Safety hazards on the public way ● Safety hazards on vacant lots and public areas	
	APPEARANCE OF COMMUNITY	● Unsightliness of storage areas/or private areas ● Presence of bulk items in storage areas/or private areas	● Spilled or scattered refuse *subsequent to collection* ● Spilled or scattered Refuse *prior to collection*	● Presence of abandoned bulk items ● Presence of abandoned automobiles	● Unsightliness of public way *subsequent to cleaning* ● Unsightliness of vacant lots and public areas ● Clogged drain basins	● Spilled or scattered refuse during transport
	ODOR WITHIN COMMUNITY	● Offensive odors in waste storage areas	● Widespread odors throughout storage areas			
	SATISFACTION OF NEEDS	● "Storage-collection capacity" relative to "storage-collection needs"				
		● Capacity deficiencies in waste storage areas	● Widespread capacity deficiencies throughout storage areas			
Related to Specific Aspects of Solid Waste System Utilized	COMPLIANCE WITH STANDARDS	● Compliance with storage requirements	● Compliance with mixed refuse collection standards	● Compliance with special pickup standards		
	INCONVENIENCE/ DISCOMFORT TO PUBLIC		● Inconvenience due to type of collection service ● Noise from collection ● Traffic congestion from collection scheduling ● Property damage from collection activities			● Traffic congestion during transport to deposit site ● Air pollution from poorly maintained vehicles

- Usefulness to Solid Waste Managers — Will the measure assist decision-makers in making comparisons among areas and/or in analyzing trends over time?

Measurement techniques were then designed for each of the resultant measures. These related to: (1) the type of data that would be needed to formulate a given measure; (2) the procedures and methods that could be used to obtain the requisite data; and (3) the suggested frequency with which the data should be collected and the measures formulated.

The full set of candidate effectiveness measures and their respective measurement techniques is presented in tabular form in Table 3 on pages 30 through 36. This list is designed to indicate the relevant measurement category, solid waste activity, and effectiveness indicator associated with each measure. The table is organized as follows:

- Measurement Category—Indicates the general category of effectiveness to which the measure relates.
- Solid Waste Activity — Delineates the solid waste activity (storage, regular collection, special collection, cleaning, transportation) to which the measure relates.
- Indicators of Effectiveness — Designates the effectiveness indicator to which the measure relates.
- Measures of Effectiveness —Indicates the form of the measure itself.
- Measurement Techniques — Specifies what would be measured, how the data would be collected, and how frequently measurements should be made.
- Type of Measure—Indicates the source of data by generic category: direct observation of existing conditions, special measurement apparatus, available records and ledgers, or a household survey.

Table 3. CANDIDATE EFFECTIVENESS MEASURES AND MEASUREMENT TECHNIQUES BY MEASUREMENT CATEGORY, INDICATOR, AND ACTIVITY

MEASUREMENT CATEGORY – PUBLIC HEALTH

Solid Waste Activity	Indicators of Effectiveness	Measures of Effectiveness for Each Indicator	MEASUREMENT TECHNIQUES			Type of Measure[b]
			Measurement Elements	Data Collection Methods	Frequency of Collection[a]	
Storage	Conditions hazardous to health in storage areas/or private areas	–Average rat count for storage areas –Percent of storage areas where rat count exceeds given threshold	Rat count for each stroage area inspected	Inspection of storage areas using rat count apparatus	Q	SM
		–Average fly count for storage areas –Percent of storage areas where fly count exceeds given threshold	Fly count for each storage area inspected	Inspection of storage areas using fly count apparatus	Q	SM
		–Percent of storage areas found to contain health hazards	Presence of health hazards in waste storage area 6 = yes, 7 = no **Types of Health Hazards** –Flies, insects –Signs of rodents –Garbage not in metal container –Garbage in a metal container without tight lid –Decaying animals –Other health hazard	Inspection of storage areas using a checklist to denote specific health hazards observed	Q	DO
		–Average health hazard rating for storage areas –Percent of storage areas where health hazard rating exceeds a given threshold	Health hazard rating assigned to each storage area inspected. 3 point scale (based on above data)-- 1 = no health hazards observed 2 = minor health hazards observed (no signs of insects, flies, or rodents) 3 = major health hazards observed (signs of insects, rodents, and flies)			
Regular Collection	Widespread health hazards throughout neighborhood storage areas	–Percent of blocks where more than "X" percent of the storage areas are found to contain health hazards	This is a derived measure based on data gathered from inspection of several waste storage areas in a number of blocks			
	Health hazards in waste collection area	–Number of containers per block set out the night before collection which pose potential health hazards	Number of containers in waste collection areas which are of a nature that rats and/or animals can gain access and which are put out the night before collection	Inspection of waste collection areas	Q	DO
Cleaning	Health hazards on the public way (streets and alleys)	–Percent of inspections found to contain health hazards on the public way (inspection unit = blockface or alley)	Presence of health hazards (as defined above) on streets and alleys. 6 = yes; 7 = no	Inspection of streets and alleys using checklist to denote specific health hazards observed	Q	DO
	Health hazards on vacant lots and public areas	–Percent of inspections where health hazards are found on lots or public areas (inspection unit = blockface or alley)	Presence of health hazards (as defined above) on lots and public areas. 6 = yes; 7 = no	Inspection of lots and public places using checklist to denote specific health hazards observed	Q	DO

See footnotes at end of table.

Table 3 (continued)

MEASUREMENT CATEGORY – PUBLIC SAFETY						
Solid Waste Activity	**Indicators of Effectiveness**	**Measures of Effectiveness for Each Indicator**	**MEASUREMENT TECHNIQUES**			**Type of Measure** [b]
			Measurement Elements	**Data Collection Methods**	**Frequency of Collection** [a]	
Storage	Conditions hazardous to safety in storage areas/or private areas	—No. of fires caused partially by improper storage of combustible solid waste per 1000 persons	Number of fires attributable to solid waste accumulation	Review of fire department records	A	R
		—Percent of storage areas found to contain safety hazards	Presence of safety hazards in waste storage area: 6 = yes, 7 = no **Types of Safety Hazard** —Broken glass —Barbed wire —Refrigerator with door intact —Combustible waste sufficient to cause a fire —Other safety hazards	Inspection of storage areas using a checklist to denote specific safety hazards observed	Q	DO
		—Average safety hazard rating for storage areas —Percent of storage areas where safety hazard rating exceeds a given threshold	Safety hazard rating assigned to each storage area inspected: 3 point scale (based on above data) -- 1 = no safety hazards observed 2 = minor safety hazards observed (i e., broken glass, barbed wire) 3 = major safety hazards observed (i.e., refrigerator with door intact, waste sufficient to cause fire)			
Regular Collection	Widespread safety hazards throughout neighborhood storage areas	—Percent of blocks where more than "X" percent of the storage areas are found to contain safety hazards	This is a derived measure, based on data gathered from inspection of several waste storage areas in a number of blocks			
Cleaning	Safety hazards on the public way (streets and alleys)	—Percent of inspections found to contain safety hazards on the public way (inspection unit = blockface or alley)	Presence of safety hazards (as defined above) on streets and alleys: 6 = yes; 7 = no	Inspection of streets and alleys using a checklist to denote the specific safety hazards observed	Q	DO
	Safety hazards on vacant lots and public areas	—Percent of inspections where safety hazards are found on lots or public areas (inspection unit = blockface or alley)	Presence of safety hazards (as defined above) on lots and public areas: 6 = yes; 7 = no	Inspection of lots and public places using a checklist to denote the specific safety hazards observed	Q	DO

See footnotes at end of table.

Table 3 (continued)

MEASUREMENT CATEGORY – APPEARANCE OF COMMUNITY

Solid Waste Activity	Indicators of Effectiveness	Measures of Effectiveness for Each Indicator	MEASUREMENT TECHNIQUES			Type of Measure[b]
			Measurement Elements	Data Collection Methods	Frequency of Collection[a]	
Storage	Unsightliness of storage areas/or private areas	—Average appearance rating for storage areas —Percent of storage areas where appearance rating exceeds a given threshold	Appearance rating assigned to each storage area inspected: 7 point scale	Visual inspection of storage areas, using photographs as a reference for selecting the appropriate rating	Q	DO
	Presence of bulk items in storage areas/or private areas	—Percent of storage areas found to contain abandoned or discarded bulk items	Presence of abandoned or discarded bulk items in storage area: 6 = yes; 7 = no	Inspection of storage areas	Q	DO
Regular Collection	Spilled or scattered refuse *subsequent to collection*	—Percent of blockfaces containing spilled or scattered refuse subsequent to collection (where curbside collection is performed) + Percent of alleys containing spilled or scattered refuse subsequent to collection (where alley collection is performed)	Presence of spilled or scattered refuse subsequent to collection: 6 = yes; 7 = no	Inspection of waste collection areas	Q	DO
Special Collection	Presence of abandoned bulk items	—Percent of inspections where abandoned or discarded bulk items are observed (inspection unit = blockface or alley)	Presence of abandoned or discarded bulk items on streets, alleys, lots, and public areas: 6 = yes; 7 = no	Inspection of streets, alleys, lots, and public places	Q	DO
	Presence of abandoned automobiles	—Percent of inspections where abandoned automobiles or trucks are observed (inspection unit = blockface or alley)	Presence of abandoned automobiles or trucks on streets, alleys, lots, and public areas: 6 = yes; 7 = no	Inspection of streets, alleys, lots and public places	Q	DO
Cleaning	Unsightliness of public way *subsequent to cleaning*	—Average litter for streets and alleys —Percent of streets and alleys where litter rating exceeds a given threshold	Appearance rating assigned to each street or alley inspected: 7 point scale	Visual inspection of streets and alleys, using photographs as a reference for selecting the appropriate rating	Q	DO
		—Number of unsolicited complaints from citizens about the appearance of their community per 1000 persons	Number of complaints from residents about the appearance of their community	Review of sanitation department records	M	R
		—Index of citizen satisfaction with the appearance of their community	Citizen attitudes toward general appearance of their community: 4 point scale	Survey of a representative sample of residents to ask about appearance of the community	A	S
	Unsightliness of vacant lots and public areas	—Average litter rating for vacant lots and public areas	Appearance rating assigned to each vacant lot or public place inspected: 7 point scale	Visual inspection of lots and public places using photographs as a reference for selecting the appropriate rating	Q	DO
	Clogged drain basins	—Number of drain basins which are clogged per basin inspected	Presence of clogged drain basin: 6 = yes; 7 = no	Inspection of streets and alleys	Q	DO
Transportation	Spilled or scattered refuse during transport	—Percent of collection fleet likely to cause spillage while in transport	Number of open trucks in the collection fleet	Review of records on truck inventory	A	R

See footnotes at end of table.

Table 3 (continued)

MEASUREMENT CATEGORY — ODOR

Solid Waste Activity	Indicators of Effectiveness	Measures of Effectiveness for Each Indicator	MEASUREMENT TECHNIQUES			Type of Measure[b]
			Measurement Elements	Data Collection Methods	Frequency of Collection[a]	
Storage	Offensive odors in waste storage areas	—Percent of storage areas found to contain offensive odors	Presence of odor in waste storage area: 6 = yes; 7 = no	Inspection of storage areas	Q	DO
Regular Collection	Widespread odors throughout storage areas	—Percent of blocks where more than "X" percent of the storage areas are found to contain offensive odors	This is a derived measure, based on data gathered from inspection of several waste storage areas in a number of blocks			
		—Number of unsolicited complaints from citizens about the existence of offensive odors per 1000 persons	Number of complaints from residents about offensive odors in their community	Review of sanitation department records	M	R

See footnotes at end of table.

Table 3 (continued)

MEASUREMENT CATEGORY – SATISFACTION OF NEEDS						
Solid Waste Activity	Indicators of Effectiveness	Measures of Effectiveness for Each Indicator	MEASUREMENT TECHNIQUES			Type of Measure[b]
			Measurement Elements	Data Collection Methods	Frequency of Collection[a]	
Storage and Regular Collection	"Storage – collection capacity" relative to "storage – collection needs"	–Amount to which the combined storage-collection capacity falls short of the "true" storage-collection need per block (in pounds of refuse)	Amount of refuse set out for collection (and picked up) per week as compared with the amount of refuse generated in a given block	Review of sanitation department records; review of census records	Q	R
		–Percent of blocks where the combined storage-collection capacity falls short of the "true" storage-collection needs	This is a derived measure, based on data gathered on a number of blocks			
Storage	Capacity deficiencies in waste storage areas	–Percent of storage areas found to contain an inadequate number of containers	Indications that there are an insufficient number of containers in the waste storage area: 6 = yes; 7 = no –Overflowing containers –Refuse piled on ground –Containers in poor condition	Inspection of storage areas	Q	DO
Regular Collection	Widespread capacity deficiencies throughout storage areas	–Percent of blocks where more than "X" percent of the storage areas contain an inadequate number of containers	This is a derived measure, based on data gathered from inspection of several waste storage areas in a number of blocks			

See footnotes at end of table.

Table 3 (continued)

MEASUREMENT CATEGORY — COMPLIANCE WITH STANDARDS

Solid Waste Activity	Indicators of Effectiveness	Measures of Effectiveness for Each Indicator	MEASUREMENT TECHNIQUES			Type of Measure[b]
			Measurement Elements	Data Collection Methods	Frequency of Collection[a]	
Storage	Compliance with storage requirements	—Percent of storage areas having containers which do not comply with the regulations	Presence of containers which do not comply with the regulations: 6 = yes; 7 = no	Inspection of storage areas	Q	DO
Regular Collection	Compliance with mixed refuse collection standards	—Percent of blockfaces where mixed refuse remains uncollected for one or more days (where curbside collection is performed) + Percent of alleys where mixed refuse remains uncollected for one or more days (where alley collection is performed)	Presence of uncollected mixed refuse along streets or alleys on the day subsequent to pickup: 6 = no, 7 = yes	Inspection of waste collection areas	Q	DO
		—Average delay time in meeting regular pickup schedules for alleys/blockfaces	Amount of lapsed time between actual and scheduled pickup of mixed refuse	Comparison of advertised collection schedule with dispatchers log	M	R
		—Number of unsolicited complaints from citizens about delays in pickup of mixed refuse per 1000 persons	Number of complaints from residents about delays in the pickup of mixed refuse	Review of sanitation department records	M	R
Special Collection	Compliance with special pickup standards	—Percent of instances in which there was a delay of one day or more in the pickup of bulky items	Presence of uncollected bulk items on the day subsequent to pickup: 6 = yes; 7 = no	Inspection of waste collection areas	Q	DO
		—Average delay time in the pickup of bulky items	Amount of lapsed time between actual and scheduled pickup of bulky items	Comparison of advertised collection schedule with dispatcher's log	M	R
		—Number of unsolicited complaints from citizens about delays in pickup of special or bulk items per 1000 persons	Number of complaints from residents about delays in the pickup of bulk refuse	Review of sanitation department records	M	R

See footnotes at end of table.

Table 3 (continued)

MEASUREMENT CATEGORY – INCONVENIENCE/DISCOMFORT TO PUBLIC

Solid Waste Activity	Indicators of Effectiveness	Measures of Effectiveness for Each Indicator	MEASUREMENT TECHNIQUES			Type of Measure [b]
			Measurement Elements	Data Collection Methods	Frequency of Collection [a]	
Regular Collection	Inconvenience due to type of collection service (curb, alley, etc.)	–Average amount of time spent per household per month preparing refuse for collection	Amount of time spent per household per month in preparing refuse for collection	Survey of a representative sample of residents	A	S
	Noise from collection	–Percent of collection areas where noise from collection exceeds a given threshold	Noise measurement taken at each waste collection area inspected	Inspection of waste collection areas using noise measurement apparatus	A	SM
		–Number of collection miles taking place during early morning hours as a percent of total collection miles	Time of day that collection takes place along linear segments of the collection route	Review of dispatcher's log	M	R
		–Number of collection stops taking place during early morning hours as a percent of total collection stops	Time of day that collection takes place at each collection area			
		–Number of unsolicited complaints from citizens about noise caused by refuse collection activities per 1000 persons	Number of complaints from residents about noise from collection activities	Review of sanitation department records	M	R
	Traffic congestion from collection scheduling	–Number of miles of major and secondary arterial roads where refuse collection is performed during peak hours –Amount of peak hour time during which refuse collection is taking place along major and secondary arterial roads	Time of day that collection takes place along each major and secondary arterial road in a given collection route	Review of dispatcher's log	M	R
	Property damage from collection activities	–Number of reported instances of property damage caused by collection equipment or collection personnel per 1000 persons	Number of reported incidents of property damage caused by collection personnel or collection equipment	Review of sanitation department records	M	R
		–Total dollar value of property losses caused by collection equipment or collection personnel	Dollar value of individual claims made against the sanitation department for property damage due to collection activities	Review of sanitation department records	M	R
Transportation	Traffic congestion during transport to deposit site	–Amount of peak hour time during which collection fleet is enroute (to or from) central deposit source	Total transport hours which occur during peak hours of the day	Review of dispatcher's log	M	R
	Air pollution from poorly maintained vehicles	–Percent of vehicles where air pollution rating exceeds a given threshold	Air pollution measurement taken for each collection vehicle inspected	Inspection of collection vehicles using air pollution testing devises	A	SM

[a] M = monthly; Q = quarterly; A = annually.

[b] R = records; S = household survey; DO=direct observation; SM= special measurement apparatus.

IV. DEVELOPMENT OF ANALYTICAL METHODS FOR COMBINING COMPONENT VARIABLE MEASUREMENTS TO PRODUCE AN OVERALL EFFECTIVENESS MEASURE

Besides developing effectiveness measures and measurement techniques, the methodological approach to this project included:

- A review of the available methods for combining component variable measurements into overall measures.
- The development of analytical procedures to be used in this project to formulate solid waste effectiveness indices.

These aspects of the methodology development are discussed in this chapter.

RELEVANT CONSIDERATIONS IN FORMULATING EFFECTIVENESS INDICES

The problem to be addressed is one of selecting a decision model that appropriately combines a set of multidimensional variables into a unidimensional measure (index) that correctly classifies the effectiveness of a community's storage, collection, and transportation activities. The salient aspects to be considered include:

- Types of combining models that are appropriate
- Procedures for formulating these models and determining the weights for the component variables that comprise the index
- Methods for achieving a consensus when more than one evaluator or decision-maker is involved.

These points are discussed in the following paragraphs.

Available Combining Models

Fundamental to the development of an index is an underlying model that reflects the relative importance that a decision-maker attaches to the variables that comprise the measure. There are several decision models that can be used to derive an index of effectiveness. These include the linear, conjunctive, and disjunctive models, as described below.

- ### The Linear Model

  The linear model produces an index that is an additive sum of the component variables. For example, if health, safety, and appearance are the principal indicators in assessing solid waste system effectiveness, the values of the variables associated with these indicators are weighted to reflect their relative importance and then added together to obtain an overall score. This score is the index value for the given values of the variables.

- ### The Disjunctive Model

  The disjunctive model produces an overall measure that can be used to classify a solid waste system into one of two categories—effective or ineffective. In this model, if one of the indicators (e.g., appearance) has an extremely high value (where high represents a favorable condition), then the solid waste system is deemed effective, regardless of the values of the other variables.

- ### The Conjunctive Model

  The conjunctive model also produces an overall measure that can be used to classify a solid waste system into one of two categories—effective or ineffective. However, in this model, high values on one or more variables will not compensate for low values on the other variables. Rather, all of the individual component variables must meet certain minimum thresholds in order for the system to be classified as effective. For example, if the health variable did not meet the standard, the system would be ineffective, even if it were meeting (or surpassing) the standards set for the other relevant variables.

Of these three decision models, the linear model is the one most frequently employed when a mathematical formula for combining multidimensional attributes is desired. However, there is evidence to support the contention that individuals, when faced with the task of mentally combining potentially contrasting effects to formulate decisions, use non-linear decision rules, similar to those exemplified by the conjunctive and disjunctive models.[8]

It should be noted that in the strict sense in which these models are defined, neither the disjunctive nor the conjunctive model produces an index in the same manner as the linear model. The former two models provide indices that denote "acceptability" or "non-acceptability." The latter model, on the other hand, produces an index that has an ordered set of values, ranging from some minimum level to a maximum level. Thus, the linear model can be used to rank neighborhoods. That is, if one neighborhood has an index value of I_1 and another neighborhood has an index value of I_2, then I_1 larger than I_2 implies, by the index, that the first neighborhood is "better off" than the second neighborhood.* All that can be said when the conjunctive or disjunctive models are used to compare neighborhoods is that certain ones are "good" and others are "bad." If two neighborhoods are both good or both bad, there is no way to select the one that is better or the one that is worse.

On the other hand, the conjunctive and disjunctive methods explicitly require the notion of "good" or "bad" in their application, while the linear model does not. Thus, index values derived from the latter model do not in themselves provide sufficient information by which to classify a neighborhood as good or bad. To assess the status of a neighborhood using a linear model, it is necessary to determine a threshold value for the index such that if the index exceeds the threshold, the neighborhood is classified as good. Otherwise, it is classified as bad.

The discussion thus far has focussed on the similarities and dissimilarities among the three combining models, based on the strict sense in which they have been defined. However, if mathematical functions are used to approximate these various models,

---

* This assumes that the variables are scaled so that the larger the value of the index, the better it is.

they all take on similar properties; that is, they are all continuous functions that yield to the ordinal assumption, whereby rankings can be made, and they all require explicit statements as to what determines "goodness" or "badness" in the resultant index values.

For the linear model, there is a mathematical formulation explicit in its definition. For the other two models, mathematical approximations are required for their representation. Appendix A indicates functional relationships that can be used to provide mathematical approximations for the models and illustrates geometrically the models and their approximations.

Procedures for Formulating an Index Using the Models

There are two methods that can be used to formulate indices that are based upon the decision models described above. These are:

- Mathematical methods
- Judgmental methods.

Mathematical methods are based on statistical relationships among the measured variables. Judgmental methods, on the other hand, utilize value judgments or opinions of one or more individuals, knowledgeable about the activities being assessed, to formulate a composite measure. Each of these methods is described below.

Mathematical Methods for Formulating an Index

Fundamental to the use of mathematical methods in formulating an index is the notion that statistically significant relationships can be established among the measured variables. Where the actual status of that which one is trying to classify is known or can be estimated (e.g., a global judgment about the overall conditions in a census tract, neighborhood, etc.), it can be used as the dependent variable and the profiles (as reflected in the measurements made on health, safety, appearance) can be used as independent variables. The relationship between the dependent variable and the independent variables can then be estimated, utilizing mathematical functions that approximate the linear, disjunctive, and conjunctive models. The model that seems to "best fit" the data is selected. The parameters that characterize the function become weights and reflect the relative importance of each component variable in assessing the overall effectiveness of solid waste operations. This technique has been employed in medical studies where the dependent variable is the presence or absence of a disease, and the independent variables consist of symptoms, pathological signs, and clinical findings prior to the time the actual diagnosis is made.

Where there is no a priori way to characterize the dependent variable, Principal Components Analysis may be used as an alternative statistical procedure for combining variables. This procedure is a method for reducing the number of variables in such a fashion as to lose as little information as possible.

Judgmental Procedures for Formulating an Index

As an alternative to the mathematical procedure, the opinions of "experts" may be solicited in such a manner as to construct a composite index of effectiveness. To form indices corresponding to the linear, conjunctive, or disjunctive models, the following minimal information must be obtained from the experts:

(1) For the linear model, the judges must be questioned so as to elicit the weights w_1, w_2, ... w_i for the variables x_1, x_2, ... x_i that are to be measured.

Generally, the weights are normalized so that $\sum w_i = 1$. The index of effectiveness (E) is represented as:

$$E = w_1 x_1 + w_2 x_2 + \ldots + w_i x_i$$

(2) For the conjunctive model, the judges must be questioned in such a manner as to elicit the "tolerable" thresholds t_1, t_2, ... t_i for the variables x_1, x_2, ... x_i. A neighborhood is considered "good" if all variables exceed their respective threshold values; i.e.,

$$x_i > t_i \text{ for all } i$$

It is considered "bad" if at least one variable fails to exceed its threshold value; i.e., if

$$x_i \leq t_i \text{ for any } i$$

(3) For the disjunctive model, the judges must be questioned in such a manner as to elicit the thresholds or standards s_1, s_2 ... s_i for the variables x_1, x_2, ... x_i. A neighborhood is considered good if at least one variable exceeds its respective threshold; i.e.,

$$x_i > s_i \text{ for any } i$$

It is considered bad if none of the variables exceed their respective thresholds; i.e., if

$$x_i \leq s_i \text{ for all } i.$$

There are several techniques that can be used to assist the expert in expressing his (or her) preferences in an ordered manner, and, in turn, facilitate development of a composite measure.[9] These include:

- Rating
- Ranking
- Forced Decisions
- Decision Alternative Ratio Evaluation (DARE).

Rating involves assigning weights to all of the variables directly on a scale from zero to one, where the weights reflect the relative importance of the variables. For example, the expert might assign .5 points to the health variable, .3 points to appearance variable, and so on. Ranking involves ordering the variables in terms of their relative desirability. This technique, however, does not permit the expert to state the strength of his preferences.

The third method, called "forced decisions," is based on pairwise comparisons. When using this technique, a decision is made for each pair of variables as to which is the more important one. A score of "one" is assigned to the preferred variable; a score of "zero" is assigned to the other variable. After all pairs of variables have been assessed, the scores for the individual variables are totalled by variable. Weights are then determined on the basis of the number of points a given variable receives relative to the total points overall.

The fourth method, called Decision Alternative Ratio Evaluation (DARE), requires the expert to assess a series of variables in terms of how much more important one variable is than any other. The expert is asked to take the variables two at a time and assign a numeric value that reflects how strongly he prefers one variable over the other. For example, the expert might feel that the public health variable is three times as important as the public safety variable; he might find the public safety variable only half as important as the appearance variable; and so forth. This technique thus allows the expert to quantitatively order his preferences. When the appropriate algorithim is applied, these stated preferences can be converted into an overall measure, where each variable receives a weight based on its importance relative to the other variables under consideration.

All of the above procedures belong to the general category of decision weighting models. They are generally used in conjunction with developing an overall measure, based on the linear model. An alternative technique for formulating judgmental indices is the Delphi Method, a Feedback and Reassessment process.[10] This will be discussed further in the following subsection.

Methods for Achieving a Consensus

Whenever more than one individual is involved in the decision-making process, there arises what has been termed the "consensus problem." That is, individuals are likely to disagree on the relative importance and, in turn, the weights they assign to the variables. There is, unfortunately, no universally applicable criterion for handling this problem, because there is no theoretical basis for making the interpersonal comparisons that are needed in order to appropriately combine individual preference patterns. Any change in the method of combining the variables will affect some persons favorably and others adversely, and there is no a priori way of weighting the net results. This does not mean that interpersonal comparisons should not be made. Indeed, often they must be made. Rather, no general formula can be invented that handles all such problems satisfactorily.[11,12]

Among the proposed solutions are the following:[13]

- Procedures or criteria for weighting the individual opinions
- Procedures for reconciliation of individual desires into a collective opinion.

The procedures for mechanically combining individual opinions include both simple averages and weighted averages of individual preferences. The simple average approach is generally employed when there is no reason to believe that any one individual's preferences should be given more weight than another. The weighted average approach is applicable when dealing with expert opinion and there is reason to believe that each individual has a specific area of expertise. In this case, it is desirable to weight the opinions or preferences of certain individuals more heavily than others.

The procedures for reconciliation of individual desires into a collective opinion include Group Decision techniques and Feedback and Reassessment techniques. Group Decision techniques involve meeting as a group to discuss the matter, with a view to ultimately arriving at a satisfactory compromise among the divergent views. Feedback and Reassessment techniques involve querying each individual separately, and then providing feedback on the opinions of all individuals. Individuals are then asked to reconsider their initial assessments based on the group feedback.

The Delphi Method is the most well known form of the Feedback and Reassessment method. Questions are asked privately and anonymously of each person. The distribution of responses is summarized in a statistical fashion. The statistical summarization is then provided as feedback information to each person, who is then asked to reconsider his position in light of the majority response. Where an individual's opinion deviates substantially from the group norm, he is asked to justify his reason for holding this position. In a sense, he is asked to rate his expertise on the question. This has the effect of causing persons without strong preferences to move toward the views held by the majority of the group, while allowing those with extremely strong preferences to retain their positions on the matter. The process continues in a similar fashion for several rounds, until fairly close accord is reached among the participants.

THE ANALYTICAL PROCEDURES USED IN THIS PROJECT

The procedures utilized in this project to develop overall measures of effectiveness draw upon the techniques described above. These procedures may be summarized as follows:

(1) The field data gatherers were asked to assign a separate composite (or global) rating to each area they surveyed at the time they collected data on the effectiveness variables. This composite rating was to reflect their subjective assessment of the area and to take into consideration only the component variables; i.e., factors not related to the study, such as the color of the house, were to be ignored.

(2) The relationship between the composite rating and the component effectiveness variables was then examined, using regression techniques. Functional forms corresponding to the three decision models; i.e., linear, conjunctive, and disjunctive, were utilized to test the nature of the relationship. The results were compared to determine the strength of the relationship and the relative predictability of each of the three models.

The following paragraphs further describe these procedures.

Description of the Composite (or Global) Rating

The individuals who collected the data were asked not only to provide information from which the individual effectiveness measures could be formulated; but also, to provide an overall assessment of the area. A semantic differential approach was used in this regard. This is a self-reporting technique in which an individual is asked to directly evaluate an attitudinal object, and to indicate his opinion along a scale where the endpoints have opposite meanings, such as:

most favorable ___: ___: ___: ___: ___: ___: worst

The respondent marks the scale according to how closely he feels one adjective or the other describes his impression of the object.

For purposes of this project, the field team members were instructed to use a scale of one to eight in making their assessments. A value of "one" was assigned to the most favorable overall conditions and a value of "eight" to the worst overall conditions. Intermediate points were used to reflect conditions in between the two values.

The data gatherers were instructed to ignore those factors not related to the study in making these ratings (e.g., the color of the house, the socio-demographic characteristics of the neighborhood, and so forth). They were requested to form their overall opinion solely on the basis of the data they were collecting on the individual effectiveness measures.

Assessment of the Relationship Among Variables

The relationship between the global ratings and the corresponding component variables was examined in light of the three decision models. The models were tested using regression analysis techniques. The mathematical expressions used to formulate the models were of the type illustrated in Appendix A.

A mathematical approach was utilized in preference to a judgmental approach for the following reasons:

- The difficulties associated with implementing judgmental techniques when a sizeable number of variables is to be assessed.

- The difficulties associated with converting qualitative statements into statistical terms when more than one judge (or evaluator) is involved.

The opinions of the officials in the health, sanitation, and planning departments in the pilot test city were, however, solicited in conjunction with another aspect of the study. A summary of the results of this survey is presented in Appendix B.

In examining the relationship between the composite rating and the component variables, the data collected in the one census tract that all observers visited were used. The data from this tract were pooled across all raters; that is, simple average values were computed for the global ratings and for the component variables. The resultant equations were compared in terms of their correlation coefficients, and index values were developed, using the regression parameters as weights.

In addition, the relationship between the overall rating and the component variables was tested separately for selected raters (data gatherers). This was performed in order to compare the types of decision models the various raters used with the models that resulted from the pooled data.

V. THE FIELD DEMONSTRATION

This chapter describes the methods that were used to assess the measurement system, particularly its implementation in an urban community. The City of Baltimore was used as the test site for the field demonstration.

The field demonstration consisted of:

- An evaluation of the usefulness of the various candidate effectiveness measures by potential users of the measurement system.
- A field test of the measurement techniques in selected areas of the city.

To assess the usefulness of the candidate effectiveness measures, the list of measures, illustrated in Table 3 on pages 30 through 36, was presented to a group of city government representatives in the test city. The group included representatives from the sanitation, health, and planning departments. They were asked to separately assess each variable in terms of how useful it would be to them in their decision-making needs. The results of this survey are summarized in Appendix B.

To assess the measurement techniques, a field test was performed, during which data were collected in ten areas of the city. It lasted for a two-week period and included a three-day training session in which ten persons were instructed in how to make the measurements and record the resultant data on a set of data collection forms.

The remainder of this chapter provides additional information on the field test. It covers the following points:

- The purpose of the field test
- The principal focus of the field test
- The survey design for the field test

- The activities performed in preparation for the field test
- The conduct of the field test.

PURPOSE OF THE FIELD TEST

The purpose of the field test was to obtain information by which to evaluate:

- The feasibility of producing the individual and composite measures (in terms of time, cost, and difficulty involved).
- The consistency (or reliability) with which subjective measurements are made; i.e., the extent to which independent observers come up with similar measurements of the same phenomena.
- The extent to which different measures are likely to be highly correlated with one another; i.e., if clean alleys are highly correlated with clean storage areas, it may not be necessary to measure both variables in future studies of this type.
- The degree to which individual variables vary in different sections of the city.
- The feasibility of developing composite effectiveness indices.

FOCUS OF THE FIELD TEST

As illustrated in Table 3 on pages 30 through 36, four types of data were used to formulate the candidate effectiveness measures. Some measures utilized existing data records and ledgers; for example, the number of trash fires per 1,000 persons, the number of complaints about schedule delays per 1,000 persons, and so forth. Some required the use of special instruments; for example, the noise from collection trucks. A third category of measures was that which required a household survey; for example an index of citizens' attitudes about the appearance of streets and alleys in their community. The majority of the measures, however, were of a nature that required direct observation of existing conditions; for example, the number of storage areas containing health hazards.

The field test focussed on the latter type of measures. It was originally envisioned that data from records and ledgers would also be included. This did not prove feasible for the following reasons:

- It would have required manual tabulation based on source materials, many of which were not centrally located.
- The method for recording complaint data in the pilot test city was such that real complaints were often indistinguishable from "requests for service."
- The sanitation department itself was in the process of changing over to a borough system, thereby making data on collection routes and collection schedules difficult to obtain.

Because of these factors, an attempt to formulate measures based on recorded data would have required designing a management information system that, on one hand, may have only limited applicability and, on the other hand, was beyond the scope of the project.

The four measures that required the use of special measurement apparatus—rat counts, fly counts, noise, and vehicle pollution—were excluded because of feedback received from city government representatives in the test city during the design phase of the study. These officials were of the opinion that measurement of flies and rats using special measuring devices was no more reliable or accurate than similar measurements made visually by trained inspectors.* It was, therefore, decided to adopt the latter

---

* These comments are not based on scientific experiments designed to compare the various measurement techniques; rather, they reflect the professional opinion of several management officials in the health department of the test city.

technique in making these types of measurements. Noise and vehicle pollution were not directly measured because the city government managers, in their evaluation of the effectiveness measures, rated them low in terms of usefulness. Additionally, it was felt that specialized measures of these types might be too cumbersome and costly for an urban community to adopt on an ongoing basis.

A Citizen Attitude Survey was not undertaken because this technique has been tested on a number of occasions and found to be feasible, if properly designed and if the city is willing to commit the needed resources. A fair amount of work in this area has been done by the Urban Institute, Washington, D. C.[14]

SURVEY DESIGN FOR THE FIELD TEST

Having defined the scope of the pilot test as one that would concentrate on measures that require visual inspection of existing conditions, a sampling plan was developed. The sampling plan was based on the concept of sampling from among and within strata. Each stratum represented a set of census tracts, having certain things in common. Since no census tract was split between strata, each stratum represented a unique set of census tracts.

There are a number of alternative schemes by which the strata could be defined. The scheme used for the field test defined five strata as follows:

- Dirty Stratum—Those census tracts the sanitation and health department personnel defined as particularly dirty.
- Model Cities Stratum—Those census tracts contained within the Model Cities areas of the city.
- Income Stratum No. 1—Those census tracts where the average family income in 1970 was less than $9,000.
- Income Stratum No. 2—Those census tracts where the average family income in 1970 was between $9,000 and $11,999.
- Income Stratum No. 3—Those census tracts where the average family income in 1970 was $12,000 or more.

To eliminate any overlap among strata, census tracts belonging to the Dirty Stratum were classified first, followed by tracts belonging to the Model Cities Stratum. The group of census tracts that remained were then classified into one of the three income-related strata.

The Dirty Stratum was included to ensure that a maximum amount of variation would be observed in the individual measures. The Model Cities Stratum was included because this area was receiving more frequent service (with respect to regular and bulk collection and street and alley cleaning) than the rest of the city.

Using information available from the 1970 Census of Housing, two census tracts were randomly selected from each stratum, and ten census blocks were selected at random from each tract.[15] Thus, altogether 100 blocks were chosen for observation during the field test.

The sampling plan called for ten observers to perform the data gathering. They were to form five teams of two observers each. Each team member was to record his (or her) observations separately. The 100 blocks were divided up among the five teams so that:

- All observer teams inspected all census tracts, but surveyed different blocks within the tracts.
- Each census tract was inspected on each day of the week.

Thus, the basic survey design consisted of a set of interpenetrating samples, the details of which are more fully described in Appendix C.

The basic survey design was augmented by having all observers inspect all blocks in one tract, termed the "common tract." These inspections were made daily, at approximately the same time, by all teams. The intensive inspection of one census tract was added to the basic survey design in order to facilitate the subsequent analysis of the measurement methodology, particularly the reliability and accuracy of the measurements and the development of effectiveness indices.

ACTIVITIES PREPARATORY TO THE FIELD TEST

Several activities were performed prior to the actual conduct of the field test to facilitate the collection of data. These included:

- Conducting a pre-survey of the units to be sampled
- Designing data collection forms and procedures
- Developing training aids and other related materials.

Each is briefly described below.

Pre-Survey of the Census Blocks in the Sample

Members of the project team visited each block that was to be surveyed during the field test. This was done for two reasons:

- To prepare block maps that could be used by the field personnel
- To confirm the fact that a selected block actually contained homes.

The block maps were developed in order to assist the field personnel in locating the correct block and in recording data. The maps indicated the shape of each block, the street names, and the location of the alleys. The streets and alleys were subsequently assigned numeric identifiers that were used by the observers when recording information. A replica of a block map is shown on the following page in Figure 3.

Several of the selected census tracts were in areas where urban renewal was underway. In some instances, all of the homes on a given block had been torn down (or condemned) since the 1970 census was taken. Where this was the case, the block was removed from the sample and a replacement block from the same census tract was selected.

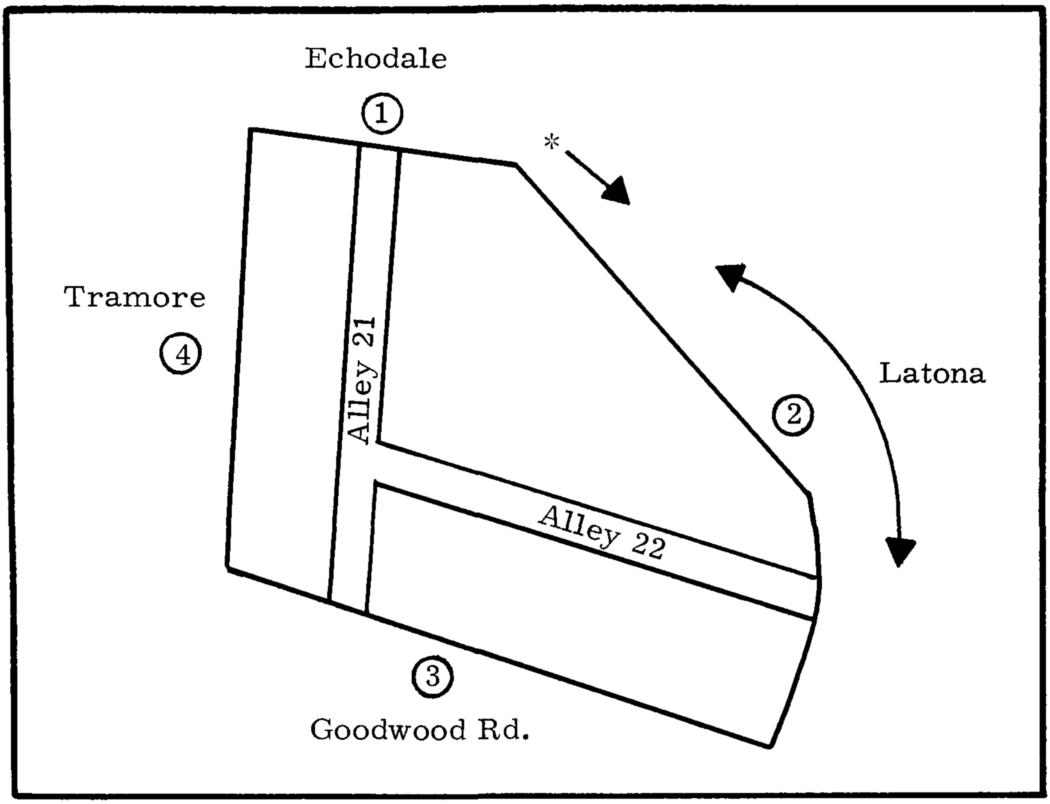


Figure 3: Replica of a Block Map That Was Prepared for the Field Test

Development of Data Collection Forms and Procedures

A set of five data collection forms was developed to facilitate the recording and subsequent computer processing of the data. These forms, replicas of which are provided in Appendix D, were designed to capture information for those measures noted in Table 3 as requiring direct observation of conditions. A detailed description of the data collection procedures was prepared in the form of an instruction booklet that was given to all field observers. Appendix D provides a summary of these procedures.

Development of Training Materials and Other Aids

A training program was developed to familiarize the field personnel with the measurement system concepts, the data collection forms, and the field test procedures. The training materials consisted of:

- An instruction booklet explaining the recording procedures for each of five data collection forms
- Opaque projector slides
- Color transparencies (slides)
- A filmstrip
- Maps
- Handouts.

CONDUCT OF THE FIELD TEST

The field test was conducted during the first two weeks in December. It included a three-day training session in which the ten participants received instructions on how to locate the sample blocks, make the required measurements, and complete the data collection forms. The remainder of the time was spent collecting the data from the ten census tracts that were to be surveyed.

The Training Session

The training session included both structured and situational type instruction. That is, visual aids and other materials were used to explain what was meant by each item on the data collection forms and how the various items should be completed. Special attention was given to clarifying the subjective-type measurements that required use of rating scales; namely, measurements of the amount of glass, garbage, and refuse. This was all done in a classroom-type setting.

The field observers were then split up into small groups, sent out to a sample block, and asked to complete the data forms by themselves. Each group was accompanied by a member of the project team, who answered questions about the recording procedures. The field observers, however, were requested to seek answers for themselves, using their instruction booklets, prior to asking the project analyst.

Following the on-site training, the group was reassembled. They were asked to rate the conditions that were presented in a series of slides. The slides focussed on situations where the recording procedure required the use of a rating scale. After each slide, the assigned ratings were discussed and clarified.*

The Field Procedures

Subsequent to the training session, the ten observers were split into five teams of two people each, to begin the actual collection of data. Each person was asked to complete the data collection forms by himself (or herself), independent of the other team member. That is, they were to refrain from discussing what they recorded before, during, or after completion of the forms.

The field personnel assembled each morning and were assigned from four to six new blocks to survey that day, including several blocks in the "common" tract where all raters went each day. Using the census tract and block maps that were provided, field team members drove to the sample block, parked their cars, and conducted the inspection on foot.

For each block, they collected data on the following six observational areas:

- Blockfaces—The area from the center of the street up to and including the curb and gutter, extending from any corner of a street to an adjacent corner.
- Private Ways—Sidewalks and front yards bordering blockfaces.
- Alleys—Passageways, usually 5 to 10 feet wide, extending into or through the interior of a block.

---

* It was as a result of these discussions that the initial scale, which was a four-point scale, was revised to a seven-point scale to allow for intermediate measurement points.

- Backyards—Areas in the rear of a structure, bounded by the property lines of adjacent structures, and, in most cases, by an alley.

- Storage Areas—Areas (external to the structure) that normally serve as the location for the refuse containers.

- Lots and Public Places—Open areas, having no structures.

Figure 4 on the following page provides a graphical display of these six areas.

Blockfaces, alleys, and private ways were inspected in segments of approximately 100 feet. Information was recorded separately for each segment. Team members were asked to mutually agree as to the specific boundaries of the area being included in each segment, to ensure the comparability of the responses of the team members. Approximately eight storage and backyard areas were inspected in each block. The field personnel were given instructions on how to randomly select the areas for inspection. All lots in the block were inspected.

Types of Information Collected and Recording Procedures

The types of information that were collected for each of the six observational areas listed above are summarized in tabular form on page 60. The table also indicates the recording method associated with each measurement element.

Essentially, there were four major types of recording procedures:

- Seven-Point Rating Scales—These were used to rate the amount of garbage, glass, and refuse observed in each of the six observational areas. The following codes were to be assigned:

1 = None observed

3 = Minor amount observed

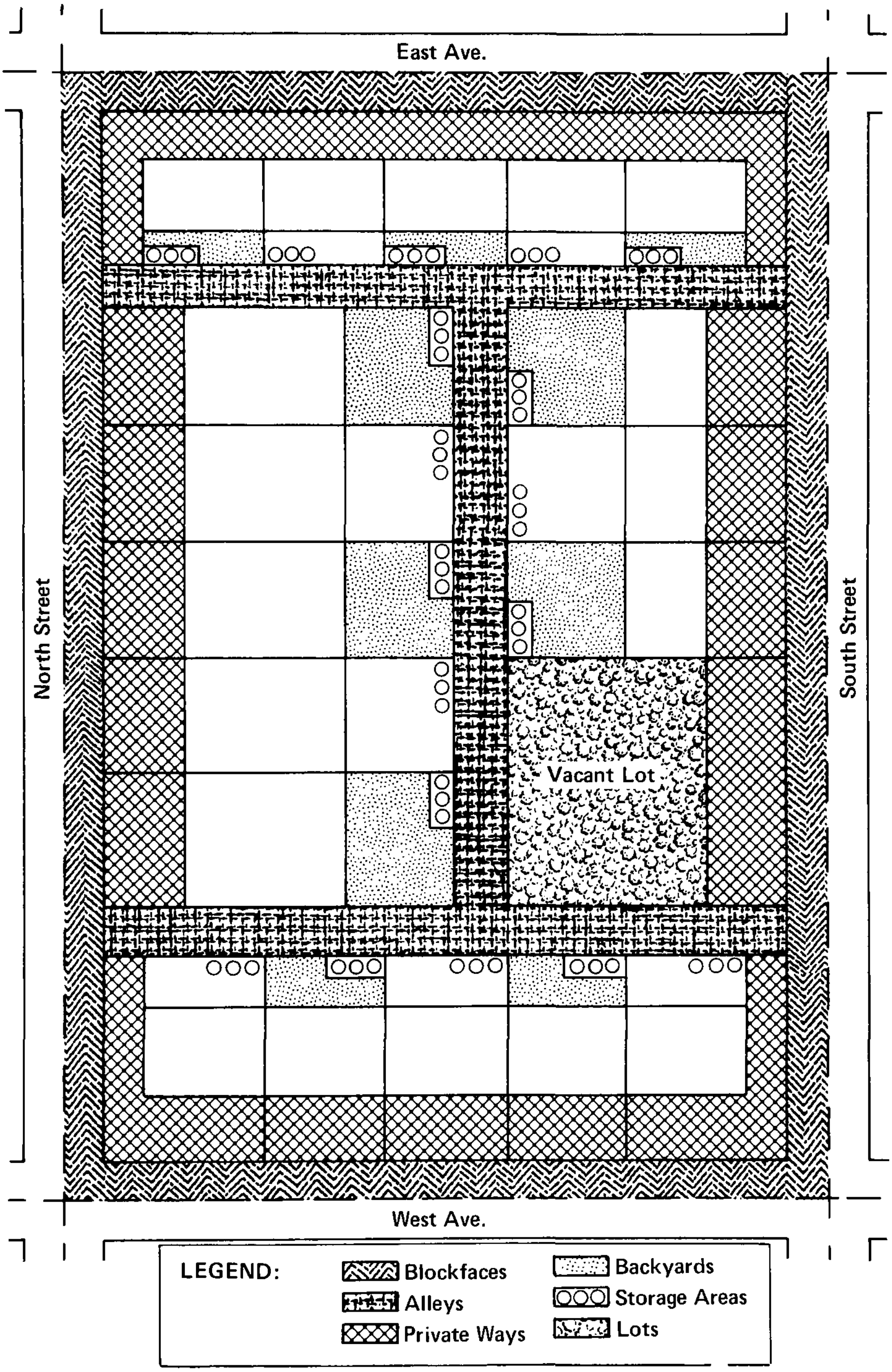


Figure 4: A Sketch of the Six Observational Areas That Were Inspected in Each Survey Block

Table 4. SUMMARY OF THE INFORMATION COLLECTED AND THE RECORDING PROCEDURES BY TYPE OF MEASUREMENT

Type of Measurement	Information Collected	Observational Area[a]	Recording Method
Health Hazards	Garbage Rating	all	7-point scale
	Rat Indicators	all	Yes/no
	Insect Indicators	Storage areas	Yes/no
	Dead Animals	all	Number observed
Safety Hazards	Glass Rating	all	7-point scale
	Fire Hazards	all	None/minor/major
	Refrigerator With Door	all	Yes/no
Unsightly Conditions	Refuse Rating	all	7-point scale
	Discarded Bulk Items	all	Number observed
	Abandoned Vehicles	all	Number observed
	Clogged Drain Basins	Blockfaces, alleys	Number observed
Odor	Odor	Storage areas	Yes/no
Satisfaction of Needs	Containers	Storage areas	Number observed (by size)
Compliance with Standards	Improper Containers	Storage areas	Yes/no
	Non-complying Containers		Number observed
Overall Assessment	Composite Rating	all	8-point scale

[a] This refers to the six areas for which data were collected: blockfaces, alleys, private ways, backyards, storage areas, and lots.

5 = Moderate to heavy amount observed

7 = Substantial amount observed

Intermediate values of 2, 4, and 6 were used when in between conditions were observed.

- Yes/No Codes—These indicated the presence or absence of a given condition.
- Counts—These reflected the number observed of a given item.
- Composite Rating Scale—This scale, ranging from 1 to 8, was used by the observer to indicate his (or her) subjective evaluation of the conditions observed. Code 1 was used to indicate the most favorable overall condition. Code 8 was used to indicate the worst overall condition. The scale progressed at one-point intervals to indicate conditions in between the two extremes.

In addition, there was one measurement that utilized a recording procedure other than those listed above; namely fire hazards. This was recorded utilizing one of the following three descriptors: none, minor, or major.

VI. FINDINGS AND CONCLUSIONS

This chapter presents the major findings and conclusions that were developed based on the field demonstration. These findings are of two types:

- General findings
- Findings related to the sample design.

GENERAL FINDINGS

The general findings of this project may be summarized as follows:

- The number of variables measured can be reduced because many of the variables were frequently found to be at their lowest value.
- Only three observational areas need to be inspected—blockfaces, alleys, and lots.
- For blockface measurements, one blockface selected at random can be used in lieu of all four blockfaces to measure the effectiveness variables.
- Observers exhibit a high degree of consistency for the "yes-no" type measurements and for "counts."
- Observers exhibit a fair degree of consistency for the more subjective-type measurements; i.e., glass, garbage, and refuse ratings. The amount of variation is less at lower scale points than at higher scale points, tending to rapidly increase and then stabilize.
- Variation among observers, however, only accounts for approximately 15 to 20 percent of total variation when measuring a tract mean.

- On the whole, observers tend to be accurate to within one-half of a scale point for the glass, garbage, and refuse ratings; a few, however, were always high or always low in their assigned ratings.
- There are a number of statistically significant correlations among the variables; however, the explained variation tends to be low.
- It is possible to develop composite measures of effectiveness, using multivariate techniques; however, the refuse rating by itself can serve as a proxy for an overall measure.

These findings are discussed in more detail in the paragraphs that follow. The analysis plan on which many of the findings are based is provided in Appendix E. Detailed tables, charts, and graphs that support the findings are presented in Appendix F.

The Number of Variables Measured Can Be Reduced

When a frequency distribution of each of the measurements was reviewed, it became apparent that many of the variables were frequently at their lowest value. By lowest value is meant the absence of the condition being measured. In the case of the rating scales, the lowest value was 1, while for counts, it was 0.

Table 5 on the following page illustrates this point. It shows for each variable the percent of time when the lowest value did not occur. The table indicates that only the following variables are likely to exhibit variation:

- Refuse Rating
- Glass Rating
- Garbage Rating
- Rat Indicators in Alleys, Storage Areas, Backyards, and Lots
- Bulk Items in Alleys, Storage Areas, Backyards, and Lots.

Table 5. DISTRIBUTIONAL CHARACTERISTICS OF THE VARIABLES

Variable	% of Observations Where Lowest Value Did Not Occur	
	All Tracts in Sample [a]	Common Tract Only [b]
Refuse Rating	65	89
Glass Rating	45	85
Garbage Rating	19	37
Rat Indicators in Alleys, Storage Areas, Backyards, and Lots	17	71
Bulk Items in Alleys, Storage Areas, Backyards, and Lots	15	28
Clogged Drain Basins	5	6
Odors	4	c
Rat Indicators in Blockfaces and Private Ways	4	11
Bulk Items in Blockfaces and Private Ways	2	3
Fire Hazards	2	2
Insects	1	c
Abandoned Vehicles	1	0
Dead Animals	1	1
Refrigerator With Door	0	0

[a] This is based on several thousand observations.

[b] This is based on slightly less than a thousand observations and reflects data on blockfaces, private ways, and alleys only.

[c] Data not available.

This would tend to suggest that only these variables need to be measured.

For comparative purposes, the table presents information for the common tract by itself (i. e., the tract where all observers went) as well as for all tracts combined. The common tract was the tract that exhibited the worst conditions overall. Thus, it is interesting to note that the same five variables listed above were also the predominant ones in this tract as well in terms of the frequency with which they exceeded their minimum levels.

More detailed information on the distribution of the measurements is provided in Appendix F, Tables F-1 through F-14 and Figures F-1 and F-2. These tables and figures present the frequency of occurrence of each variable for all the sample tracts combined and for the common tract by itself.

Only Three Observational Areas Need to Be Inspected—Blockfaces, Alleys, and Lots

There is a sufficient degree of correlation among similar measurements when compared across the six observational areas to suggest that data need be collected only for the following observational areas:

- Blockfaces
- Alleys
- Lots.

Conditions observed along private ways were found to be related to conditions observed along blockfaces. Shown at the top of the following page are the correlation coefficients and explained variation for the three measurements that exhibited the most sensitivity; namely, refuse, glass, and garbage ratings.

	Blockfaces and Private Ways	
Variable	Coefficient of Correlation (R)	Explained Variation (R^2)
Refuse Rating	.94	.88
Glass Rating	.96	.92
Garbage Rating	.80	.64

Conditions along alleys were found to be related to conditions observed in storage areas and backyards. The correlation coefficients and the explained variation between the refuse, glass, and garbage ratings for alleys and storage areas, and for alleys and backyards, are as follows:

	Alleys and Storage Areas	
Variable	Coefficient of Correlation (R)	Explained Variation (R^2)
Refuse Rating	.72	.52
Glass Rating	.79	.63
Garbage Rating	.66	.44

	Alleys and Backyards	
Variable	Coefficient of Correlation (R)	Explained Variation (R^2)
Refuse Rating	.84	.71
Glass Rating	.84	.71
Garbage Rating	.88	.78

It was difficult to assess the degree of correlation between lots and other observational areas, because of the small number of lots in the total sample. Inspection of the data, however, reveals that lot characteristics do tend to differ from those of the other five observational areas.

For Blockface Measurements, One Blockface Selected at Random Can Be Used in Lieu of All Four Blockfaces to Measure the Effectiveness Variables

During the pilot test, observers collected data on all four sides of each block that they inspected. The findings indicate that this level of detail is generally not needed to accurately estimate a tract mean rating for refuse, glass, or garbage. Rather, it would be sufficient to inspect only one blockface of each block in the sample.

In deriving this conclusion, mean refuse, glass, and garbage ratings were computed for each of the ten census tracts. One set of ratings was based on all four blockfaces in each block that was sampled. Another set of ratings was based on only one randomly selected blockface in each of the blocks. The mean values were compared using a t-test. The results, shown in Tables 6 through 8, indicate that, in general, there is no statistically significant difference in the mean values computed using the two methods.

Observers Exhibit a High Degree of Consistency for the "Yes-No" Type Measurements and for "Counts"

The raters were in agreement 96 percent of the time or better for those measurements that required either an assessment of the presence or absence of a given condition or a count of the number of similar type items that were present. Table 9 on page 72 shows the extent of agreement by variable for these types of measurements.

The item of particular significance on this table is the level of agreement among observers when making measurements of rat indicators, since this particular variable is probably the most subjective of those on the list, and, hence, likely to contain higher amounts of rater variation.

Table 6. BLOCKFACE MEAN GARBAGE RATINGS FOR CENSUS TRACTS — RATINGS BASED ON ALL BLOCKFACES IN A BLOCK VERSUS RATINGS BASED ON ONE RANDOMLY SELECTED BLOCKFACE PER BLOCK

Census Tracts	Mean Garbage Rating		Absolute Value of the Difference Between the Two Ratings	Significant Difference Established at	
	Based on All Blockfaces in the Sample Blocks	Based on One Randomly Selected Blockface in Each Sampled Block		90% Confidence Level	95% Confidence Level
1	1.43	1.25	.18	No	No
2	1.67	1.91	.24	No	No
3	1.02	1.00	.02	No	No
4	1.15	1.13	.02	No	No
5	1.06	1.05	.01	No	No
6	1.67	2.22	.55	Yes	No
7	1.04	1.05	.01	No	No
8	1.05	1.07	.02	No	No
9	1.23	1.41	.18	No	No
10	1.30	1.11	.19	Yes	Yes

Table 7. BLOCKFACE MEAN GLASS RATINGS FOR CENSUS TRACTS — RATINGS BASED ON ALL BLOCKFACES IN A BLOCK VERSUS RATINGS BASED ON ONE RANDOMLY SELECTED BLOCKFACE PER BLOCK

Census Tracts	Mean Glass Rating		Absolute Value of the Difference Between the Two Ratings	Significant Difference Established at	
	Based on All Blockfaces in the Sample Blocks	Based on One Randomly Selected Blockface in Each Sampled Block		90% Confidence Level	95% Confidence Level
1	2.53	2.14	.39	Yes	No
2	3.17	3.29	.12	No	No
3	1.55	1.35	.20	No	No
4	2.28	2.20	.08	No	No
5	1.62	1.35	.27	Yes	No
6	2.93	3.09	.16	No	No
7	1.71	1.91	.20	No	No
8	1.31	1.33	.02	No	No
9	2.01	1.82	.19	No	No
10	3.81	4.17	.36	No	No

Table 8. BLOCKFACE MEAN REFUSE RATINGS FOR CENSUS TRACTS — RATINGS BASED ON ALL BLOCKFACES IN A BLOCK VERSUS RATINGS BASED ON ONE RANDOMLY SELECTED BLOCKFACE PER BLOCK

Census Tracts	Mean Refuse Rating		Absolute Value of the Difference Between the Two Ratings	Significant Difference Established at	
	Based on All Blockfaces in the Sample Blocks	Based on One Randomly Selected Blockface in Each Sampled Block		90% Confidence Level	95% Confidence Level
1	3.01	2.75	.26	No	No
2	3.43	3.90	.47	No	No
3	1.43	1.53	.10	No	No
4	2.85	3.93	1.08	Yes	Yes
5	2.13	1.65	.48	Yes	Yes
6	3.23	3.91	.68	Yes	No
7	2.36	2.64	.28	No	No
8	1.47	1.33	.14	No	No
9	3.11	3.35	.24	No	No
10	3.91	3.94	.03	No	No

Table 9. RATER AGREEMENT FOR MEASUREMENTS OTHER THAN RATING SCALES

Recording Method	Variable	% of Observations Where Observers Agreed
Yes-No	Rat Indicators	96.3
	Odors	97.7
	Insects	99.8
	Refrigerator With Door	100.0
Number Observed	Bulk Items	96.9
	Clogged Drain Basins	97.7
	Abandoned Vehicles	99.7
	Dead Animals	99.8
None/Minor/Major	Fire Hazards	98.0

These results should be viewed with some caution, however. This is because many of the variables listed in Table 9 were among those found to exhibit little variation. That is, many of them did not frequently deviate from their lowest value. (See Table 5 on page 65.)

Observers Exhibit a Fair Degree of Consistency with Respect to the More Subjective-Type Measurements—Namely, the Glass, Garbage, and Refuse Ratings

It is to be expected that measurements based on the three rating scales, because of their subjective nature, are likely to exhibit a higher degree of inconsistency than would be the case with other types of measurements made during the field test.

The field test data indicated that, on the average, the observers differed by less than one point from one another on a seven-point scale, with 1 representing the most favorable conditions and 7 representing the most unfavorable conditions. The average difference across all six observational areas between pairs of observers for the garbage, glass, and refuse ratings was found to be .32, .65, and .82 points, respectively. This information is summarized in the table below, which also indicates the frequency with which raters agreed within one point or less of one another as well as the percent of observations where observers agreed within two points or less of one another.

Table 10. RATER AGREEMENT FOR RATING SCALE MEASUREMENTS

Type of Rating Scale	% of Observations Where Raters Agreed Within 1 Point or Less	% of Observations Raters Agreed Within 2 Points or Less	Mean Rating Difference
Garbage	89.1	98.4	.32
Glass	79.1	94.6	.65
Refuse	73.7	93.7	.82

The difference between observers in their assigned ratings was also analyzed to see whether the size of the rating discrepancy was related to the scale points. The initial hypothesis that observers would tend to show fairly close agreement for points 1 and 7 (the scale extremes) and much less agreement for the middle points of the scale did not prove out. Rather, inspection of a graphical plot of the rating discrepancy by scale point revealed that the amount of variation between raters showed a tendency to increase rapidly at the lower scale points and then stabilize. That is, the close agreement that was expected at the upper end of the scale did not materialize.

This relationship between the scale points and the mean differences was estimated statistically for each of the three rating scales, using the following function form:

$$Y = a - b/X$$

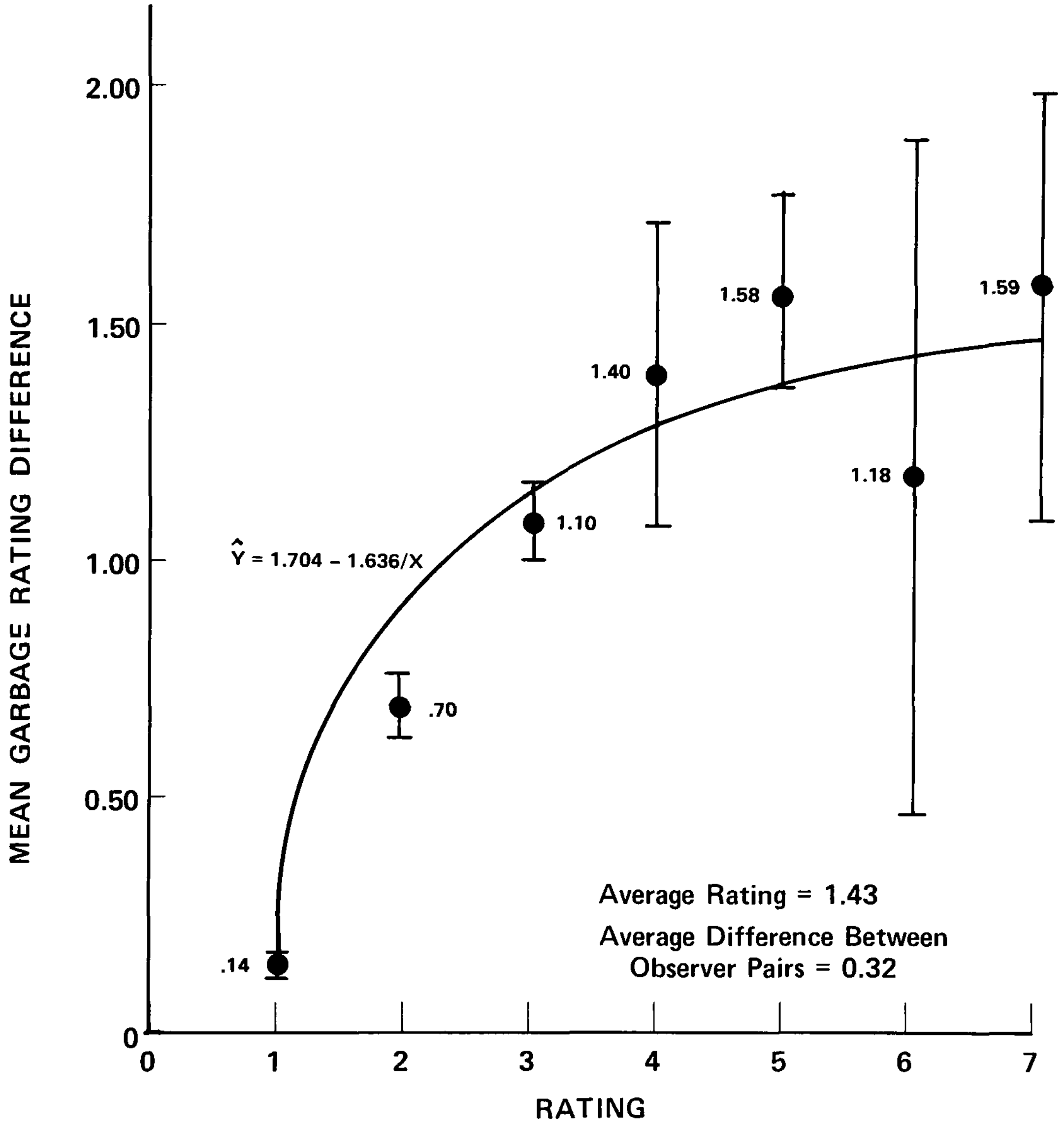
where X = the rating scale point
Y = the mean difference associated with the given scale point.

The following regression equations were developed, each of which was found to provide a fairly good fit to the data:

Type of Rating Scale	Estimating Equation	R^2
Garbage	$\hat{Y} = 1.704 - 1.636/X$	.895
Glass	$\hat{Y} = 1.465 - 1.226/X$	.884
Refuse	$\hat{Y} = 1.358 - 0.967/X$	.896

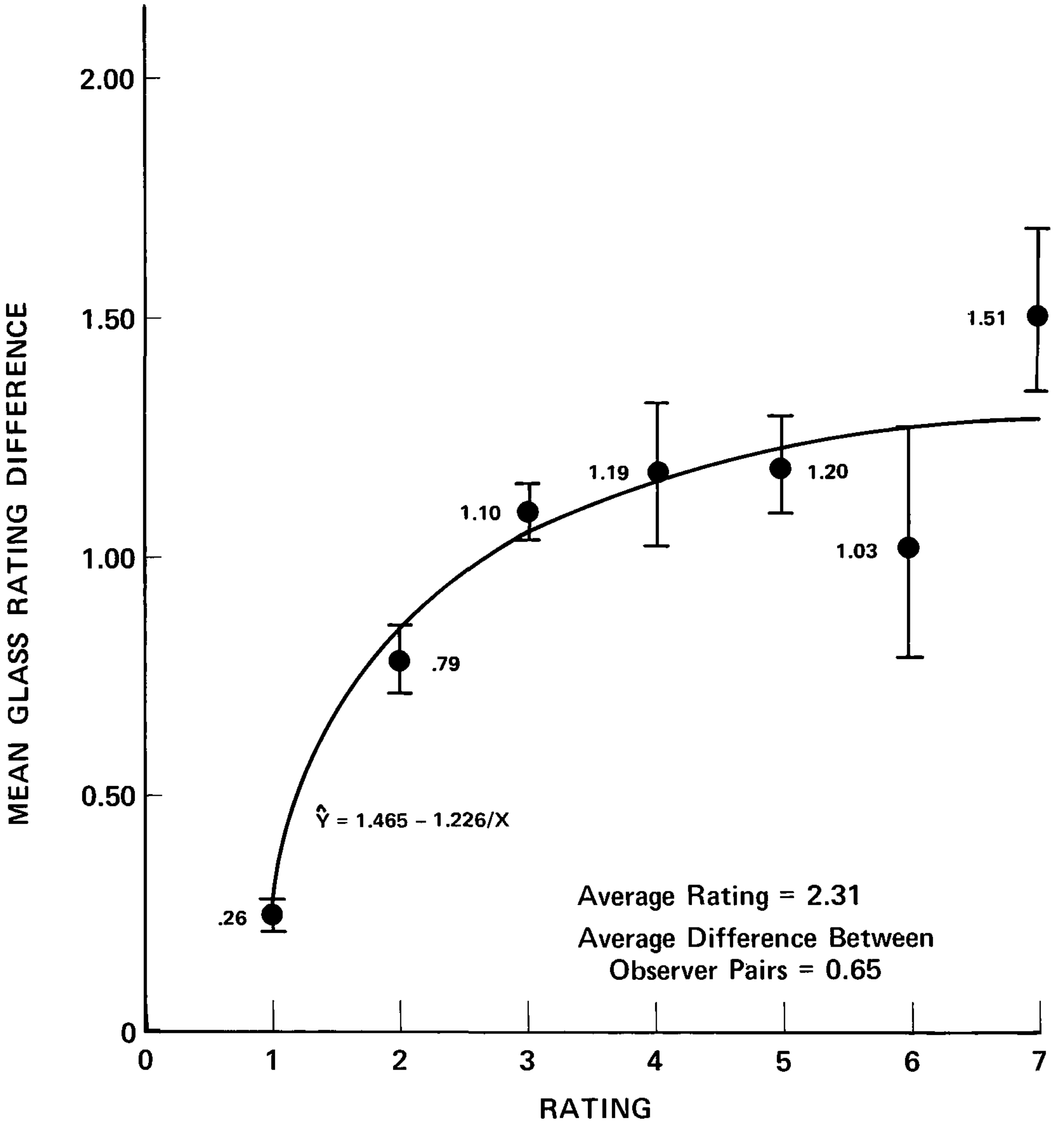
Figures 5 through 7, shown on the following three pages, present graphically the rater variation by scale point for the garbage, glass, and refuse rating scales, respectively. These figures show the actual average difference between observer pairs at each scale point and the 95 percent confidence interval that is associated with the respective mean differences. In cases where there were a sizeable number of observations (over a hundred or so) for a given scale point, the confidence bands are fairly tight. However, where the observations are few in number, the interval about the mean difference tends to be much larger. The figures also show the estimating relationships that were developed.

Appendix F, Tables F-15 through F-17 present estimates of the variance and standard deviation and the numeric values associated with the 95 percent confidence interval about the mean differences contained in Figures 5 through 7. A description of the methodology used in developing these findings is provided in Appendix E.



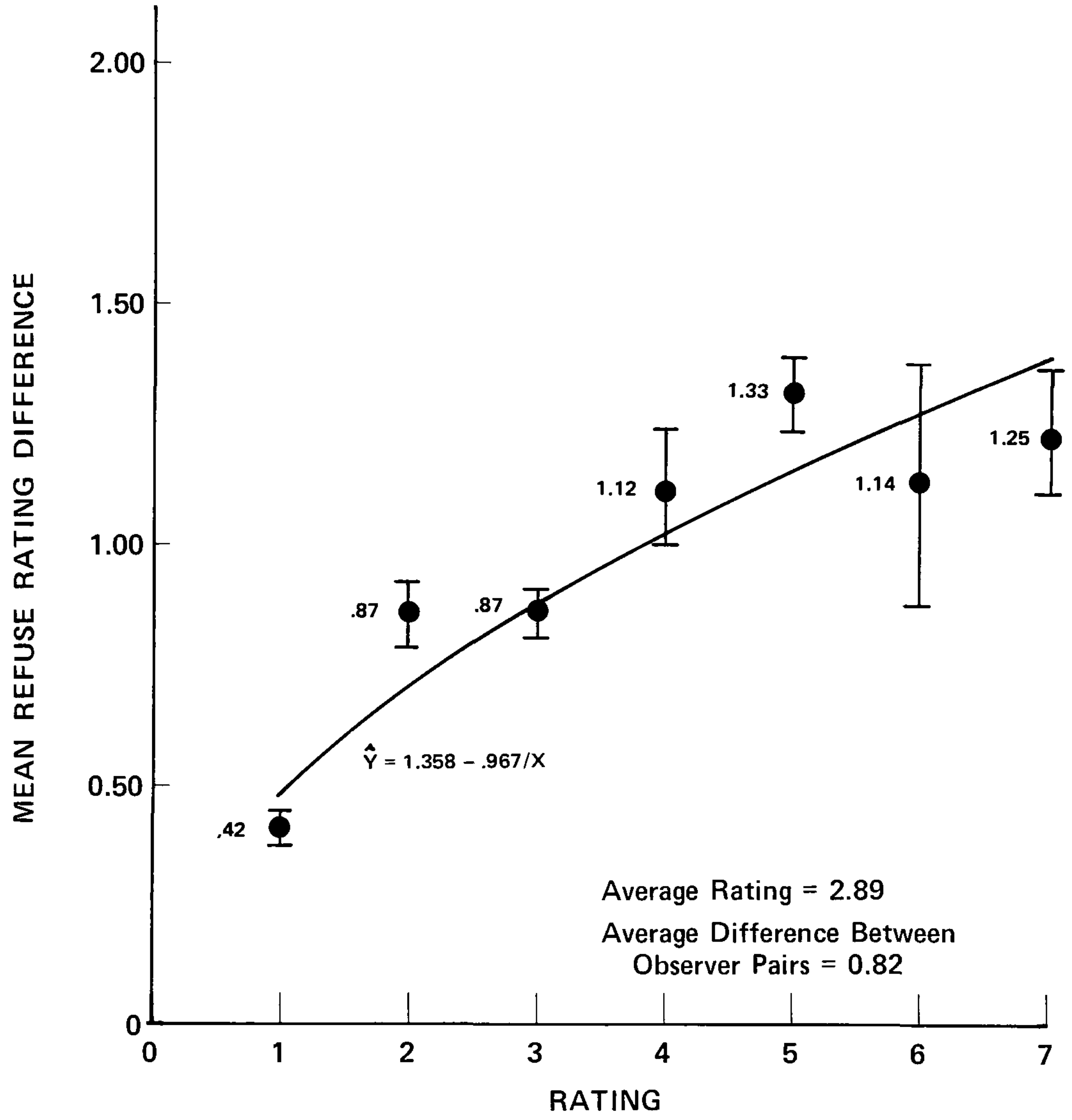
Total No. of Comparisons:	5052	329	598	30	154	11	66
% of Total:	80.9%	5.2%	9.6%	0.5%	2.5%	0.2%	1.1%

Figure 5: Mean Garbage Rating Difference by Scale Point (including 95% confidence interval about the mean differences)



Total No. of Comparisons:	3410	479	1224	158	481	65	419
% of Total:	54.7%	7.7%	19.6%	2.5%	7.7%	1.1%	6.7%

Figure 6: Mean Glass Rating Difference by Scale Point (including 95% confidence interval about the mean differences)



Average Rating = 2.89

Average Difference Between Observer Pairs = 0.82

Total No. of Comparisons:	2162	710	1684	273	731	71	605
% of Total:	34.7%	11.4%	27.0%	4.4%	11.7%	1.1%	9.7%

Figure 7: Mean Refuse Rating Difference by Scale Point (including 95% confidence interval about the mean differences)

Variation Among Observers Accounts for Approximately 15 to 20 Percent of the Total Variation When Measuring a Tract Mean

In making statistical inferences about a given geographical area, there are a number of sources of variability that can effect the estimate. Some of the major components that contribute to the variance of a census tract mean glass, garbage, or refuse rating are the following:

- Variation among blocks in the tract
- Variation among observers
- Random sources of variation.

When these variance components were estimated, by applying analysis of variance techniques to the data that were collected on the common tract, it was discovered that the major source of variation stemmed from the block to block differences. This accounted for approximately 75 percent of the total variation in the sample. The observers, on the other hand, contributed only 15 to 20 percent. Random effects made up the balance of the total variation. Both the block effect and the rater effect were significant at the .001 level. Table 11 on the following page summarizes the findings with respect to these three sources of variation in the average refuse and glass ratings for blockfaces and alleys in the common tract.

These findings provide insight into the relevance (or practical importance) of the fact that observers do not always agree on the rating to be assigned to a particular condition. In addition, the findings indicate that if objective measurements were to be utilized, the variance about the mean could be reduced, but the real key to reducing the variance about a tract mean that is estimated from sample data is the number of blocks in the sample.

The methodology used in developing these conclusions is described in Appendix E.

Table 11. THE VARIANCE COMPONENTS FOR A CENSUS TRACT MEAN VALUE FOR GLASS AND REFUSE RATINGS IN BLOCKFACES AND ALLEYS
(expressed in percentage terms)

Type of Rating Scale	% Contribution of Variance Components		
	Blocks	Raters	Random Effects
Glass Rating (Blockfaces)	76	18	6
Glass Rating (Alleys)	78	15	7
Refuse Rating (Blockfaces)	73	20	7
Refuse Rating (Alleys)	76	18	6

On the Whole, Observers Tend to Be Fairly Accurate in Their Assessment of Garbage, Glass, and Refuse Conditions; A Few, However, Were Always High or Always Low in Their Assigned Ratings

In general, the observers were able to estimate the average tract ratings for a census tract to within ± one-half of a scale point. This is illustrated more clearly in Table 12 on the following page. This table indicates the amount by which each observer differed from the overall tract value for selected rating scales. Data from the common tract were used in this regard. The distribution of the rater means about the overall tract mean is displayed graphically in Appendix F, Figures F-3 through F-8.

Table 12 suggests that a few of the observers tended to systematically assign substantially higher or lower ratings than the rest of the group. In particular, observer 8 appears to consistently assign ratings that exceed the mean by approximately one scale point, while observer 10 tends to always be on the low side of the overall tract average by about one scale point. This observation was borne out when the blockface mean glass and refuse ratings of the observers were compared against the overall mean ratings by block for the ten blocks that comprise the common tract.

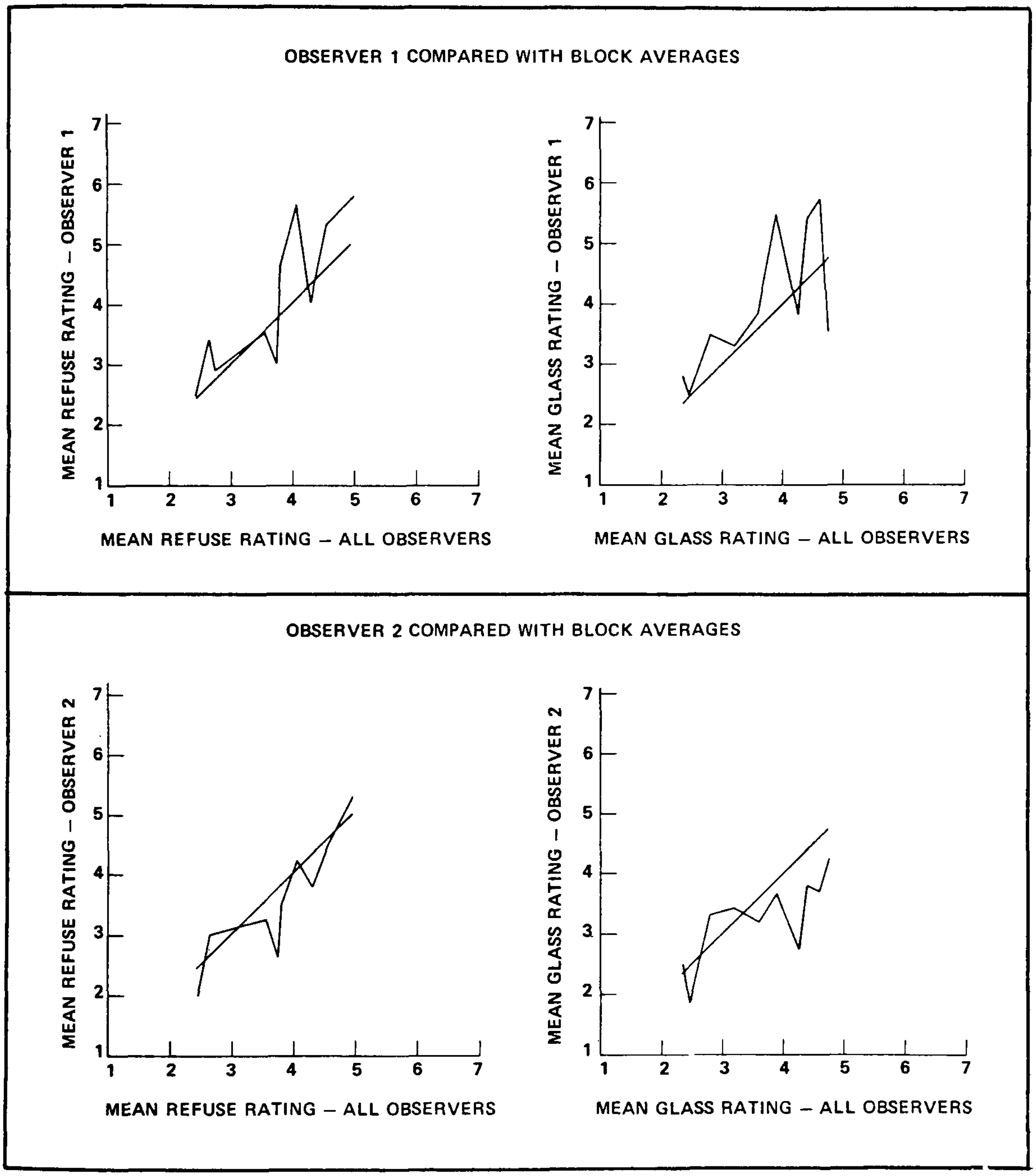
The results are shown in Figure 8 which is presented on five separate pages, starting on page 81. The graphs contained in this figure illustrate that:

- Some of the raters were inherently more variable than the rest; i.e., some tended to show larger fluctuations about the block means.
- Observer 8 systematically gave higher ratings to each block in the common tract.

Table 12. DIFFERENCES BETWEEN THE AVERAGE RATER VALUES AND THE OVERALL AVERAGE VALUES OF THE GARBAGE, GLASS, AND REFUSE RATINGS FOR BLOCKFACES AND ALLEYS IN THE COMMON TRACT

Observer	Difference Between Average Rater Values and the "True" Average Values[a]					
	Blockface Garbage Rating	Blockface Glass Rating	Blockface Refuse Rating	Alley Garbage Rating	Alley Glass Rating	Alley Refuse Rating
1	.11	.26	.38	.12	.13	-.01
2	.02	-.27	-.13	-.33	-.32	-.91
3	-.02	.10	-.34	.45	.11	-.04
4	.08	.27	.21	-.42	-.31	.03
5	.44	-.42	-.63	1.71	-.09	.18
6	-.30	-.05	-.39	1.02	.19	.45
7	.01	.90	.49	-.42	1.07	.81
8	.34	1.22	1.68	.41	1.33	1.22
9	-.37	-.33	.24	-1.08	.27	-.09
10	-.09	-.97	-.94	-.67	-1.95	-1.14
Overall Tract Mean	1.42	3.42	3.51	3.15	5.02	5.31

[a] The "true" average values reflect the overall tract means across all ten observers.



OBSERVER 1 COMPARED WITH BLOCK AVERAGES

OBSERVER 2 COMPARED WITH BLOCK AVERAGES

Figure 8: Blockface Glass and Refuse Ratings of Observers Compared With Overall Ratings for Blocks in Common Tract

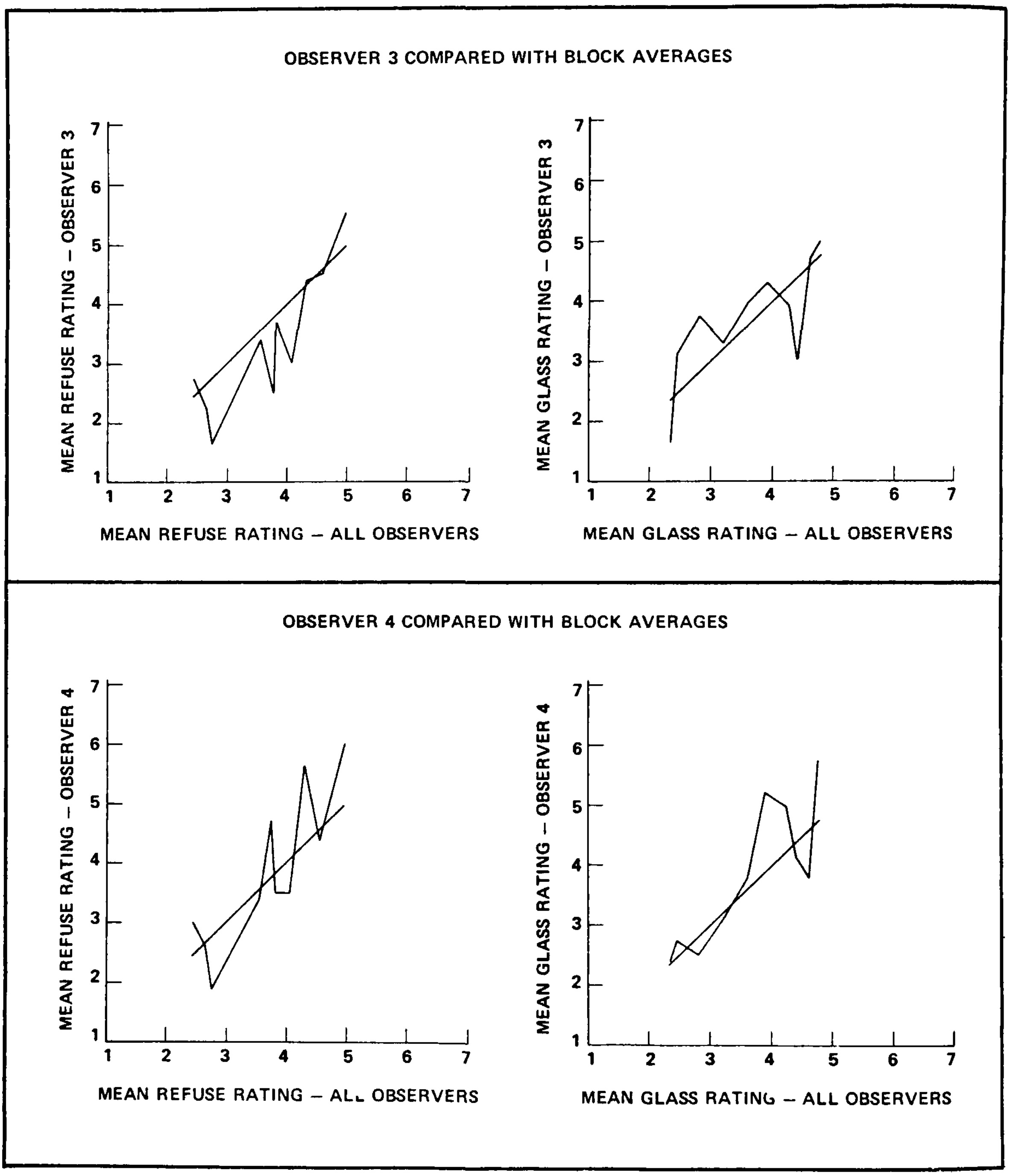


Figure 8 (continued)

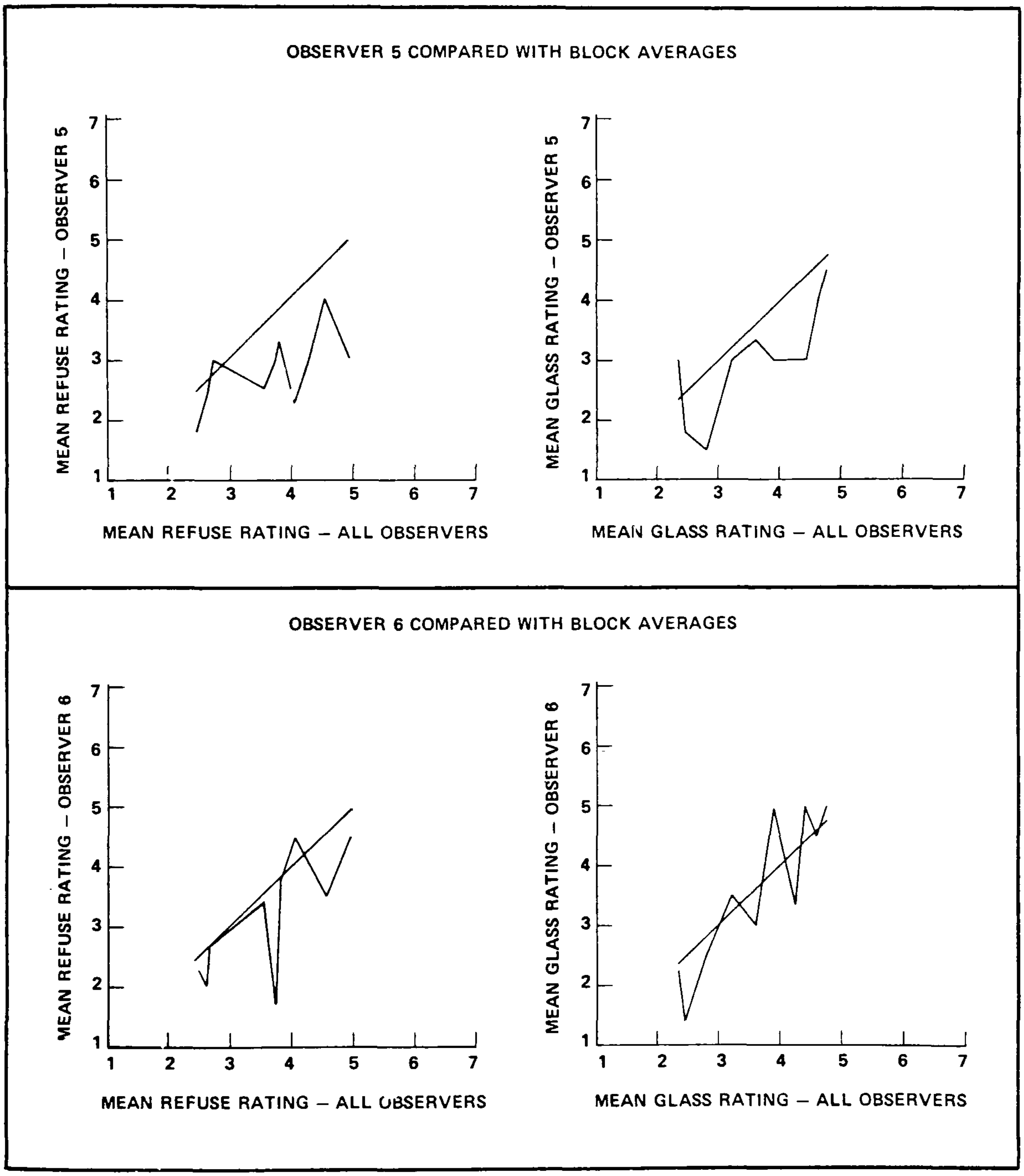


Figure 8 (continued)

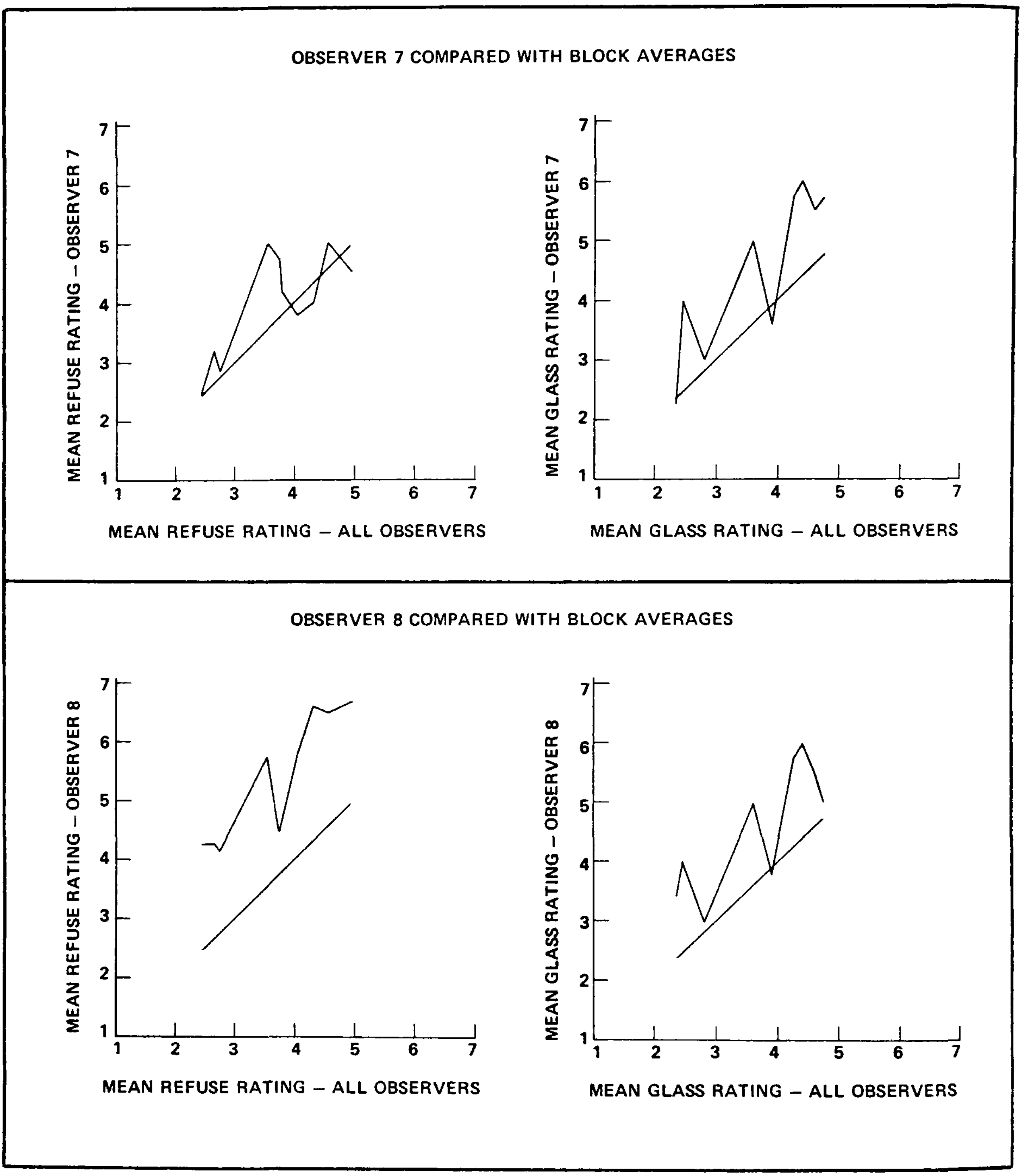


Figure 8 (continued)

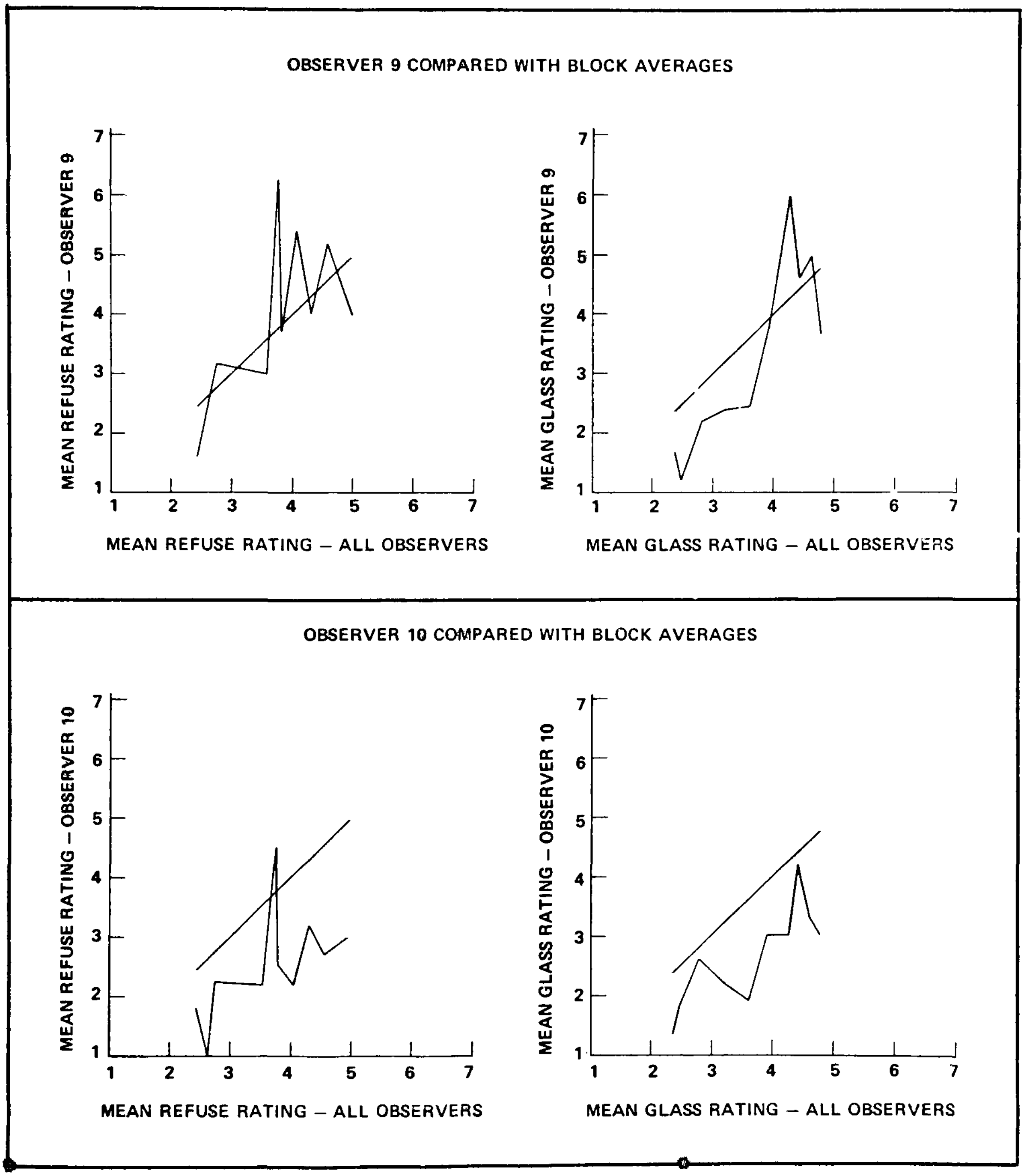


Figure 8 (continued)

- Observer 10 systematically assigned lower values to each block in the common tract.

Although Figure 8 indicates that observers 5 and 7 also tend to be consistently different, similar patterns for these two observers were not found to hold true for some of the other rating scales. Observers 8 and 10, on the other hand, were systematically different across all rating scales.

There Are a Number of Statistically Significant Correlations Among the Variables; However, the Explained Variation Tends to Be Low

Correlation analysis was used to test the degree to which individual pairs of variables were related with one another. These tests were conducted on the set of variables used to assess alley conditions as well as on the set of variables used to measure blockface conditions.

The results indicated that statistically significant correlations exist among pairs of variables. That is, there is a high probability that the real association between the two variables is not zero. However, as illustrated in Tables 13 and 14, shown on the following page, the correlation coefficients between pairs of variables are not high. Thus, although statistically meaningful relationships do exist, they do not explain a sufficient amount of the variation to be considered as good predictors of each other. This implies that if one is interested in knowing the amount of garbage, a separate measurement of garbage conditions will be required.

The correlations shown in Tables 13 and 14 are based on data from the common tract. This tract was used because the individual variables showed more sensitivity in the common tract than they did in any of the other tracts.

Table 13. CORRELATIONS AMONG THE VARIABLES USED TO MEASURE BLOCKFACE CONDITIONS

Variable Pairs	Correlation Coefficient (R)	Amount of Explained Variation (R^2)	Level of Significance
Refuse and Glass Ratings	.590	.348	.001
Refuse and Garbage Ratings	.207	.043	.001
Glass and Garbage Ratings	.164	.027	.001

Table 14. CORRELATIONS AMONG THE VARIABLES USED TO MEASURE ALLEY CONDITIONS

Variable Pairs	Correlation Coefficient (R)	Amount of Explained Variation (R^2)	Level of Significance
Refuse & Glass Ratings	.597	.356	.001
Refuse & Garbage Ratings	.473	.223	.001
Rat Indicators & Garbage Rating	.365	.133	.001
Rat Indicators & Refuse Rating	.346	.120	.001
Glass & Garbage Ratings	.308	.095	.001
Rat Indicators & Glass Rating	.287	.082	.001
Bulk Items & Refuse Rating	.261	.068	.001
Fire Hazards & Garbage Rating	.244	.060	.001
Fire Hazards & Refuse Rating	.202	.041	.001
Bulk Items & Glass Rating	.173	.030	.001

It is Possible to Develop Composite Measures of Effectiveness Using Multivariate Techniques

Overall measures of effectiveness were developed for blockfaces and alleys, based on data obtained from inspections made in the common tract. These particular observational areas were focussed upon because of a prior finding indicating that conditions in these two areas are related to conditions in three of the other observational areas in the sample. Indices for lot conditions were not developed because of the small number of observations obtained for lots. The methods used to develop the indices are described in more detail in Appendix E.

Blockface Indices

The functional relationships associated with the linear, conjunctive, and disjunctive models were estimated using regression techniques. The following least-squares equations were developed:

Type of Model	Estimating Equation	R^2
Linear	$\hat{Y}_{1B} = 8.236 - .448 X_{1B} - .289 X_{2B}$	.883
Conjunctive	$\ln \hat{Y}_{2B} = 0.900 + .345 \ln (8 - X_{1B}) + .214 \ln (8 - X_{2B})$	.879
Disjunctive	$\ln \hat{Y}_{3B} = 1.960 - .169 \ln (X_{1B} - .93) - .110 \ln (X_{2B} - .93)$	.865

where X_{1B} = value of the blockface refuse rating

X_{2B} = value of the blockface glass rating

$\hat{Y}_{jB}$ = estimated value for the blockface overall effectiveness rating.

In each case, only the coefficients associated with the refuse and glass ratings proved to be significant at statistically acceptable levels.

These equations were developed by averaging the values that individual observers assigned to blockfaces in the common tract. Similar functional relationships were also estimated for selected observers. However, in these cases the linear and conjunctive models were clearly favored over the disjunctive model. (See Appendix F, Tables F-18 through F-22.)

Of interest in comparing the three equations is that all assign approximately the same relative weights to the independent variables. This may be more clearly seen when the estimating parameters are normalized so that their sum adds up to one, as illustrated below.

Type of Model	Relative Weight for X_1 (Refuse Rating)	Relative Weight for X_2 (Glass Rating)
Linear	.608	.392
Conjunctive	.617	.383
Disjunctive	.606	.394

For comparative purposes, the three regression equations are illustrated graphically on the following two pages in Figures 9 and 10. Figure 9 presents the relationship between refuse rating values and estimated values of the overall effectiveness rating when the glass rating is held constant at various levels. Figure 10 illustrates the relationship between glass rating values and estimated values of the overall effectiveness rating when the refuse rating is held constant at various levels.

It should be pointed out that at certain values for the refuse and the glass ratings, the graphs indicate large discrepancies among the three curves, discrepancies much larger than would be expected based on the R^2 values shown above. This is explained by the fact that the data used to fit the three equations contained relatively few of the extreme values.[*] The largest discrepancies among the three equations

---

[*] This is due primarily to the fact that averages across observers were used. Any averaging technique tends to obscure the extreme values.

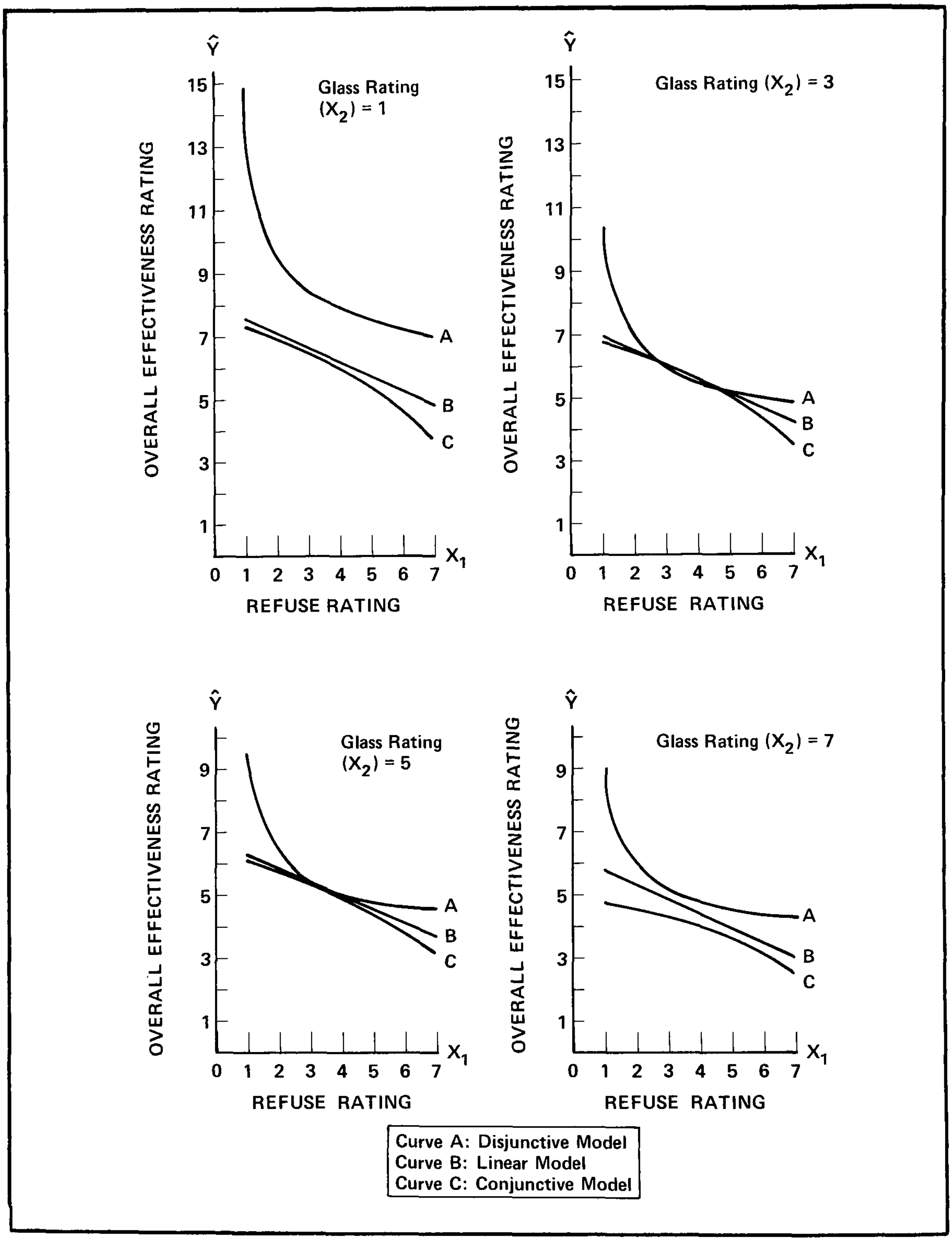


Figure 9: Graphical Presentation of the Regression Equations Developed for Blockface Data: Relationship Between Refuse Rating and Overall Effectiveness Rating When Glass Rating is Held Constant

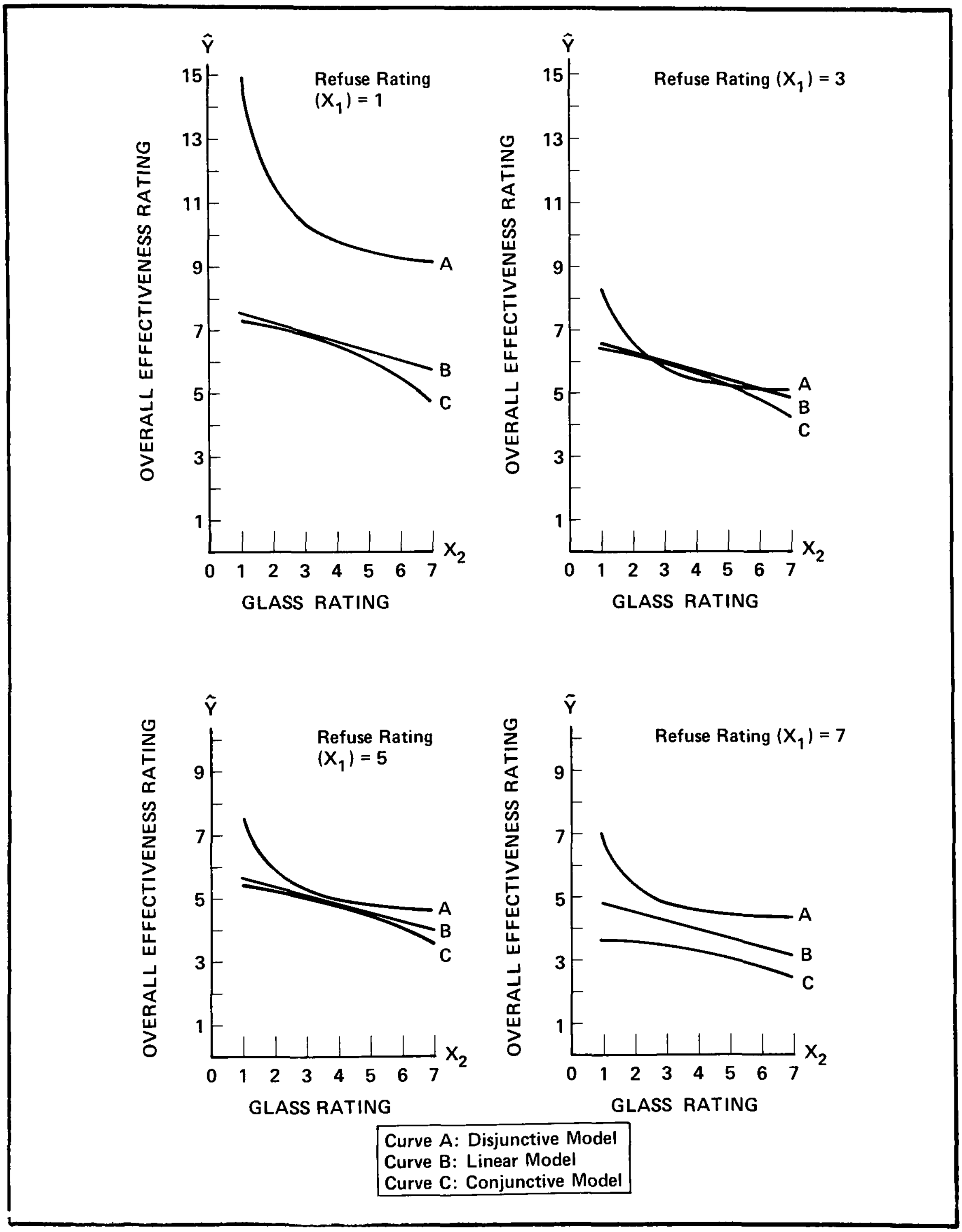


Figure 10: Graphical Presentation of the Regression Equations Developed for Blockface Data: Relationship Between Glass Rating and Overall Effectiveness Rating When Refuse Rating is Held Constant

occur when one or both of the independent variables is shown to have an extreme value. Thus, in the relevant range for which the data were fit, all three curves may be said to be fairly close to one another.

In the case of the three equations presented above, the coefficient of determination (R^2) provides little guidance for selecting a decision model that might underlie an overall effectiveness measure. However, the disjunctive model is probably a poor choice for two reasons. First, it takes on most of its driving force when the independent variables are at their extreme values. However, as mentioned above, the data used to fit the model contained relatively few extreme values, leading one to suspect that the R^2 value may be artificially inflated. Second, when the model was tested using individual data for selected observers, where extreme values were included, the fit to the data was not so good. Rather, observers tended to exhibit linear or conjunctive type combining tendencies.

The three equations were transformed to indices, having scales that range from 0 to 1, where 1 reflects the preferred value, by use of the following mathematical manipulation:

$$\left(\hat{Y}_j - Y_{j(min)}\right) \Big/ \left(Y_{j(max)} - Y_{j(min)}\right)$$

where $\hat{Y}_j$ = value predicted by the regression equation (j = 1, 2, 3).

The resultant index formulas for the three models are shown in Table 15 on the following page.

Thus, it is possible to construct indices for blockface conditions; however, it should be pointed out that the refuse rating by itself can serve as a good proxy for an overall measure. When considered by itself, the refuse rating explained 83 percent of the variation in the case of the linear and conjunctive models presented above and 80 percent of the variation in the disjunctive model. Relatively high correlations between the refuse rating and the overall rating were also found in the estimating relationships developed for individual observers.

Table 15. INDEX FORMULAS
FOR BLOCKFACE CONDITIONS

Type of Model	Index Formula
Linear	$I_{1B} = .226\ \hat{Y}_{1B} - .698$
Conjunctive	$I_{2B} = .208\ \hat{Y}_{2B} - .521$
Disjunctive	$I_{3B} = .094\ \hat{Y}_{3B} - .404$

Alley Indices

When the functional relationships associated with the linear, conjunctive, and disjunctive models were estimated using alley data, the linear and conjunctive models provided a substantially better fit than did the disjunctive model. The least-squares equations that resulted for the former two models are as follows:

Type of Model	Estimating Equation	R^2
Linear	$\hat{Y}_{1A} = 8.804 - .501 X_{1A} - .219 X_{2A} - .295 X_{3A}$	.977
Conjunctive	$\ln \hat{Y}_{2A} = .338 + .287 \ln (8 - X_{1A}) + .167 \ln (8 - X_{2A}) + .395 \ln (8 - X_{3A})$	.956

where X_{1A} = value of the alley refuse rating
X_{2A} = value of the alley glass rating
X_{3A} = value of the alley garbage rating
Y_{jA} = estimated value for the alley overall effectiveness rating.

In both cases, the coefficients associated with the refuse, glass, and garbage ratings were statistically significant at the .001 level.

Unlike the estimating equations for blockfaces, these equations do not assign the same relative weights to the independent variables. As shown below, when the estimating parameters are normalized so that their sum adds up to one, the linear model places heaviest weight on the refuse rating, while the conjunctive model places heaviest weight on the garbage rating:

Type of Model	Relative Weight for X_1 (Refuse Rating)	Relative Weight for X_2 (Glass Rating)	Relative Weight for X_3 (Garbage Rating)
Linear	.493	.216	.291
Conjunctive	.338	.197	.465

Index formulas, scaled from 0 to 1, for these two models are provided in Table 16 below. Here also, however,

Table 16. INDEX FORMULAS
FOR ALLEY CONDITIONS

Type of Model	Index Formula
Linear	$I_{1A} = .164 \hat{Y}_{1A} - .279$
Conjunctive	$I_{2A} = .169 \hat{Y}_{2A} - .236$

the refuse rating by itself can serve as a good proxy for an overall measure. When considered by itself, the alley refuse rating accounted for 91 percent of the variation in the case of the linear model and 86 percent of the variation in the case of the conjunctive model.

FINDINGS RELATED TO THE SAMPLE DESIGN

In addition to the general findings described above, there were several findings that related to the sample design. They may be summarized as follows:

- Large differences in the mean tract ratings were found to exist between two groupings of the strata.
- The sampling plan was adequate to detect changes of one-half a point or less in the blockface ratings; larger samples would be required to detect an equivalent change in the alley ratings.
- The mean tract ratings did not appear to be biased by the day of the week when the inspection was made.

These findings are discussed in the paragraphs that follow.

Large Discrepancies in the Mean Tract Ratings Were Found to Exist Between Two Groupings of the Strata

As described in Chapter V of this report, the sampling plan was based on the concept of strata; i.e., mutually exclusive sets of census tracts. The five strata used in the project were: Dirty, Model Cities, Income Level 1, Income Level 2, and Income Level 3.

When the mean garbage, glass, and refuse values for each stratum were compared with those of the other strata, two distinct groups appeared to emerge, consisting of the following strata:

Group 1	Group 2
Dirty	Income Level 2
Model Cities	Income Level 3
Income Level 1	

This two-fold split is illustrated in Table 17, shown below, which presents the mean blockface and alley ratings for each stratum.

Table 17. AVERAGE REFUSE, GLASS, AND GARBAGE RATINGS FOR BLOCKFACES AND ALLEYS BY STRATUM

Stratum	Blockface Mean Ratings			Alley Mean Ratings		
	Refuse Rating	Glass Rating	Garbage Rating	Refuse Rating	Glass Rating	Garbage Rating
Dirty	3.55	3.42	1.47	4.90	5.17	2.88
Model Cities	2.95	2.43	1.32	4.23	4.17	1.91
Income Level 1	3.27	2.59	1.45	4.36	4.74	2.23
Income Level 2	1.95	1.53	1.05	3.28	2.53	1.44
Income Level 3	1.78	1.39	1.04	2.58	1.93	1.26

Significance tests on the difference between the mean values confirmed that, in general, when a mean rating for a stratum from Group 1 was compared with a mean rating for a stratum from Group 2, the two ratings were found to be different at statistically significant levels. On the other hand, when the difference between the mean ratings for two strata within the same group was tested, no statistically significant difference between the two means could be established.*

---

* The t-test was used to determine whether or not the mean values between stratum were significantly different.

In interpreting the results, it should be pointed out that the Model Cities and Dirty Strata consisted almost entirely of census tracts where the average annual family income (in 1970) was less than $9,000. This corresponds to the income range that was used to define the Income Level 1 Stratum. Thus, Group 1 can be said to consist of census tracts where the average annual family income is less than $9,000, while Group 2 contains those census tracts where the average annual family income is $9,000 or more. Insofar as the mean ratings for these two groups were found to be substantially different, with Group 1 having much higher ratings than Group 2, the results tend to support a hypothesis that many have postulated; namely, the amount of glass, garbage, and refuse found in a given area is related to the income level of the area.

The Sampling Plan Was Adequate to Detect Changes of One-Half a Point or Less in the Blockface Ratings

In general, the higher the tract mean rating, the larger the standard error. In spite of this, the sampling plan proved adequate to detect a change of one-half a point or less in the average tract values for blockface ratings.

As illustrated in Table 18 on the following page, the highest average refuse rating for blockfaces in the ten census tracts sampled was 3.91. Given its standard error of the mean of .242, a change of ± .49 points or more is discernible. Where the mean refuse ratings were lower, the standard errors are smaller and, therefore smaller changes in the mean ratings are detectable.

Similar results were obtained when the average garbage and glass ratings for blockfaces were analyzed; i.e., the sampling plan enabled one to detect changes of one-half a point or less in the average values for a census tract. The refuse rating has been used (in Table 18) to illustrate the point because it generally had larger standard errors and slightly higher mean values than the other two blockface rating scales.

Since higher mean ratings and larger variances tend to be associated with census tracts where the average annual family income is less than $9,000, this finding implies that more intensive sampling is required in these areas to detect a given amount of change in the blockface ratings than is needed in census tracts where the income level is $9,000 or more.

Table 18. AMOUNT OF CHANGE IN THE CENSUS TRACT MEAN BLOCKFACE REFUSE RATINGS DETECTABLE AT THE 95 PERCENT CONFIDENCE LEVEL (based on sample design for field test)

Census Tract Mean Refuse Ratings	Standard Error of the Mean ($\sigma_{\overline{X}}$)	Change in Rating Detectable at 95% Level of Confidence
3.91	.242	.49
3.43	.202	.41
3.23	.233	.47
3.11	.168	.34
3.01	.186	.38
2.85	.187	.38
2.36	.141	.28
2.13	.127	.26
1.47	.097	.20
1.43	.085	.17

For alley ratings, on the other hand, the sampling plan proved inadequate in all but a few cases to detect a change of one-half a point or less in the mean tract ratings. Table 19 on the following page summarizes the mean alley refuse ratings for the ten census tracts sampled, the standard error associated with each mean, and the amount of change that would be detectable. The results imply that a larger number of alleys would need to be sampled if one wants to be able to detect small changes in the alley ratings.

Table 19. AMOUNT OF CHANGE IN THE CENSUS TRACT MEAN ALLEY REFUSE RATINGS DETECTABLE AT THE 95 PERCENT CONFIDENCE LEVEL (based on sample design for field test)

Census Tract Mean Refuse Ratings	Standard Error of the Mean ($\sigma_{\overline{X}}$)	Change in Rating Detectable at 95% Level of Confidence
5.38	.306	.64
4.46	.336	.70
4.40	.359	.75
4.30	.267	.56
4.27	.360	.75
4.15	.384	.80
3.61	.321	.67
3.30	.337	.70
1.78	.152	.32
1.50	.139	.29

The Mean Tract Ratings Did Not Appear to Be Biased by the Day of the Week When the Inspection Was Made

Analysis of variance tests were conducted using the data collected by several of the observers to see if there were any statistically significant day-to-day changes in the ratings in the common tract. As mentioned previously, the common tract was the area that exhibited the worst conditions overall. The three observers for whom the analysis was carried out were among the three most accurate and most conscientious raters.

The results tend to suggest that in areas where the mean ratings are high, inspections can be made on any day of the week and still be representative. In only a few of the cases tested was there a significant difference or bias in the ratings due to the day of the week.

VII. RECOMMENDATIONS

The recommendations stemming from this project are of two types:

- General recommendations on how to develop a measurement system that is specific to a given community.
- Detailed recommendations on how to implement an ongoing measurement system using the findings from the field test data.

Each is briefly described below.

RECOMMENDATIONS RELATED TO DEVELOPING A MEASUREMENT SYSTEM

The following recommendations are offered for communities who want to use the general procedures that were developed during this project as an aid in designing a solid waste effectiveness measurement system that is unique to their own community:

(1) Review the list of measures and measurement techniques (provided in Table 3 on pages 30 through 36) and select those measures most useful.

(2) Use the basic survey design developed for this project to obtain preliminary data on those measures that require direct observation of existing conditions. Income ranges appear to be the most viable candidate for establishing sampling strata.

(3) Utilize several observers in each tract and have them make the same measurements.

(4) Apply techniques similar to those used in this project to determine the appropriate sample size for an ongoing measurement system and the relevant variables for which measurements should be made.

RECOMMENDATIONS RELATED TO IMPLEMENTING AN ONGOING MEASUREMENT SYSTEM

The recommendations presented below are offered to communities who want to implement an effectiveness measurement system that is based upon the results of the field demonstration conducted in the City of Baltimore.

(1) Collect data on blockfaces, alleys, and lots only.

(2) Sample approximately ten blocks in each census tract where the overall conditions are bad to detect a change of one-half a point or less in the blockface garbage, glass, and refuse ratings; inspect fewer blocks in areas where the overall conditions are good. Use the income level of the census tract as an initial means for classifying conditions.

(3) Inspect only one randomly selected blockface in each block; inspect all alleys and lots in the block.

(4) Utilize several observers and have them inspect different blocks in the same census tract in order to reduce the variation associated with inconsistency among observers.

(5) Periodically compare the observations of raters within the same tract to see if any of the observers are consistently high or consistently low.

(6) Make measurements only of the amount of refuse found in these areas. Use this as an indicator of overall conditions.

<u>or</u>

Make measurements of the amount of refuse, glass, and garbage found in the areas, the presence of rat signs (alleys only) and the number of bulk items (alleys only). Report the measurements separately and/or as a composite measure.

(7) Report the location of fire hazards, bulk items, abandoned vehicles, clogged basins, and other items of interest so that corrective action can be taken.

CITED REFERENCES

1. Blair, L.H., and A.I. Schwartz. How Clean Is Our City? Washington, The Urban Institute, 1972.

2. "Community Litter Survey Technique." New York, Keep America Beautiful, Inc., May 5, 1973. (Working Paper.)

3. Ralph Stone and Company, Inc. The Use of Bags for Solid Waste Storage and Collection. U.S. Environmental Protection Agency, 1972. [Distributed by National Technical Information Service, Springfield, Va., as PB-212 590.] p. 21.

4. Fund for the City of New York. Unpublished data, 1973.

5. D. C. Department of Environmental Services. Unpublished data.

6. "Sanitation Management Information System Concept." CONSAD Corporation, January 1972. (Unpublished Report.)

7. "Effectiveness Report on the Use of Plastic Bags as Trash Receptacles in the City of Baltimore." Baltimore, Department of Public Works, September 24, 1973. (Working Paper.)

8. Einhorn, H. J. "The Use of Nonlinear, Noncompensatory Models in Decision Making." Psychological Bulletin, 73(3): 221-230, 1970.

9. Klee, A. J. "The Role of Decision Models in the Evaluation of Competing Environmental Health Alternatives." Management Science, 18(2): B-52 to B-67, October 1971.

10. Helmer, O. "Analysis of the Future: The Delphi Method." Defense Doc. Center Document AD 649640. The Rand Corporation, 1967. pp 1-11.

11. Baumol, W. J. Economic Theory and Operations Analysis. 2d ed. New Jersey, Prentice-Hall, Inc., 1965. pp 355-385.

12. Arrow, K. J. Social Choice and Individual Values. Cowles Commission Monograph No. 12. New York, John Wiley and Sons, Inc., 1951.

13. Klee, A. J. "The Utilization of Expert Opinion in Decision-Making." AIChE Journal, 18 (6): 1107-1115, November 1972.

14. Webb, K., and H. P. Hatry. Obtaining Citizen Feedback: An Application of Citizen Surveys to Local Governments. Washington, The Urban Institute, 1973.

15. U.S. Bureau of the Census. 1970 Census of Housing: Block Statistics. Final Report HC (3)-106, Baltimore, Maryland Urbanized Area. Washington, U.S. Government Printing Office, 1971.

APPENDIX A

FURTHER DISCUSSION OF THE LINEAR, CONJUNCTIVE, AND DISJUNCTIVE DECISION MODELS

This appendix presents mathematical representations for the linear, conjunctive, and disjunctive decision models that were used in the project. In addition, geometric illustrations are provided for the latter two models.

THE LINEAR MODEL

The linear model produces an index that is an additive sum of the component variables. For example, if health, safety, and appearance are the principal indicators in assessing solid waste system effectiveness, the values of the variables associated with these indicators are weighted to reflect their relative importance and then added together to obtain an overall score. This score is the index value for the given values of the variables.

The linear model may be represented by the following general form:

$$E = \sum_{i=1}^{n} w_i x_i$$

where E represents an overall (or global) assessment of conditions. The w_i are the weights that are given to the individual component variables, x_i.

THE CONJUNCTIVE MODEL

The conjunctive model says that the effectiveness of a solid waste system depends on whether each of the component variables used to measure effectiveness surpasses a threshold value set for it. If $E = f(x_1, x_2, \ldots, x_n)$ represents the vector of effectiveness variables, and $T = g(t_1, t_2, \ldots, t_n)$ represents the vector of thresholds, then only if x_i is greater than t_i for all i will the system be effective. Since a certain minimum value is required for all the variables, this implies that a high score on one variable cannot compensate for a low score on another variable, as is the case with the linear model. Thus, the conjunctive model adheres to a multiple cutoff procedure rather than a linear compensatory procedure.

In the strict sense in which the model is defined, the function is a discontinuous one that can take on the values of "1" or "0" only. The value of the function is "1" whenever all the x_i are greater than the corresponding t_i; otherwise the value is "0." A geometric representation of this model is shown in Figure A-1(a) on the following page. There are a number of mathematical forms that are continuous in nature that can be used to approximate the conjunctive model. In this project, the following parametric function was used:

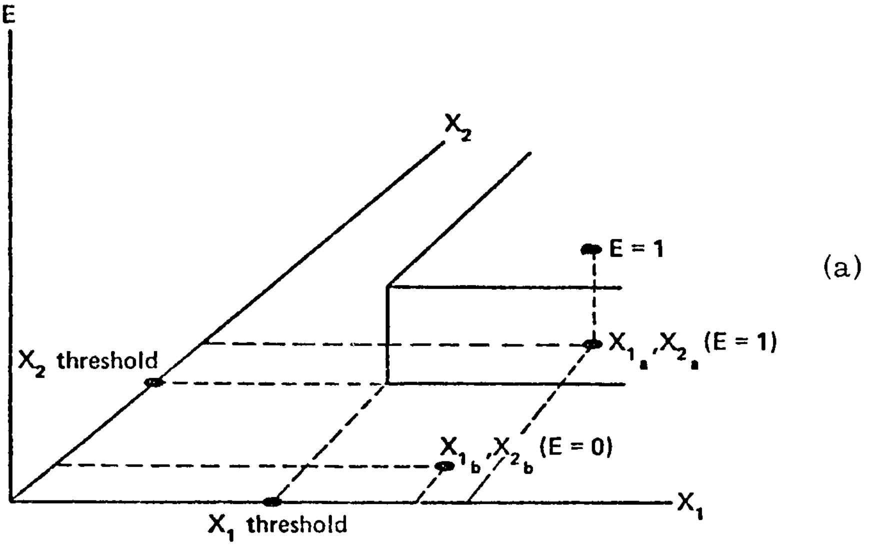
$$E = \prod_{i=1}^{n} x_i^{w_i}$$

where the variables are defined as in the linear model.

A geometric representation of this function is provided in Figure A-1(b). The nature of this function is such that a low value for any one of the component variables will produce a low overall value. Moreover, the low value cannot be compensated for by high scores on the other variables, since it is the product of the variables that is the important factor.

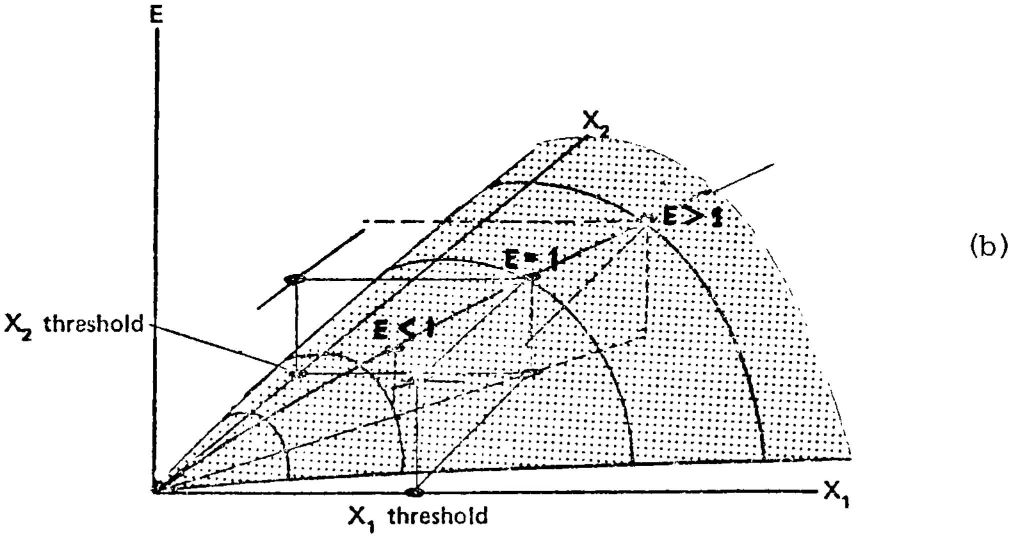
To facilitate the use of ordinary least square methods in fitting the data to the model, a linear transformation of the above function can be performed by taking the logarithms of both sides of the equation. The model thus becomes:

$$\ln E = \sum_{i=1}^{n} w_i \ln(x_i)$$



Conjunctive E = 1 if $X_1 > X_1$ threshold and $X_2 > X_2$ threshold

E = 0 otherwise



Continuous, differentiable approx. to a conjunctive model

Figure A-1: Geometric Representations of the Conjunctive Model

Because the raw data to be used in fitting the least squares function were scaled such that the smaller the value the better the conditions, the variables were converted to a reverse scale. This had the effect of producing a function that retained the properties of the one being discussed but was its mirror image.

THE DISJUNCTIVE MODEL

This disjunctive model says that the effectiveness of a solid waste system depends on whether at least one of the component variables used to measure effectiveness exceeds the individual threshold set for it. If $E = f(x_1, x_2, \ldots, x_n)$ represents the vector of effectiveness variables, and $S = h(s_1, s_2, \ldots, s_n)$ represents the vector of thresholds, then if x_i exceeds s_i for at least one i, the system is effective.* Since only one component variable needs to exceed its standard, the disjunctive model is called a maximum evaluation function, as compared with the conjunctive model which is a minimum evaluation function.

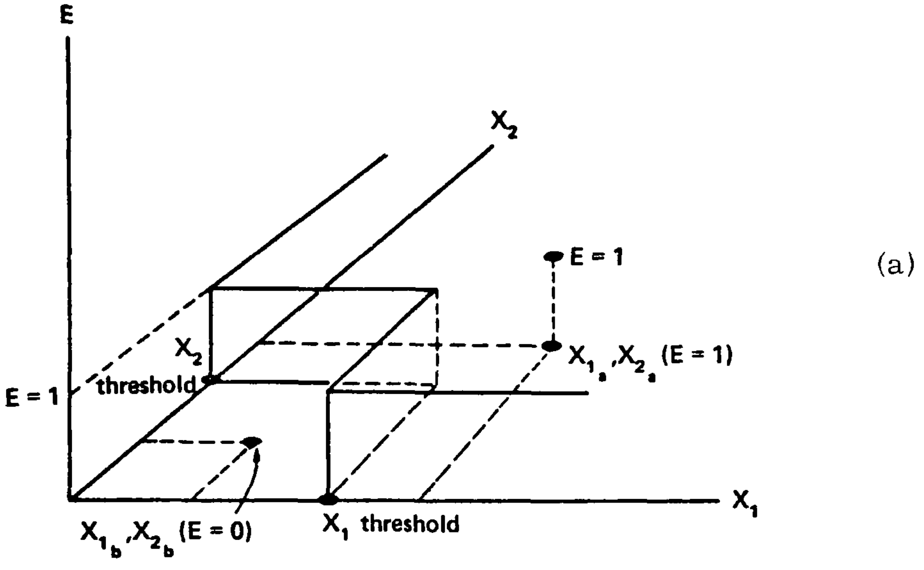
In the strict sense in which the model is defined, the function is a discontinuous one that can take on the values of "1" or "0" only. The value of the function is "1" whenever any one of the x_i exceed the corresponding s_i; otherwise the function has a value of "0." A geometric representation of this model is presented in Figure A-2(a), shown on the following page.

As in the case of the conjunctive model, there are a number of mathematical forms that can be used to approximate the disjunctive model and give it a continuous shape. A hyperbolic function, having the following form, was used in this project:

$$E = \prod_{i=1}^{n} (\rho_i - x_i)^{-w_i}$$

---

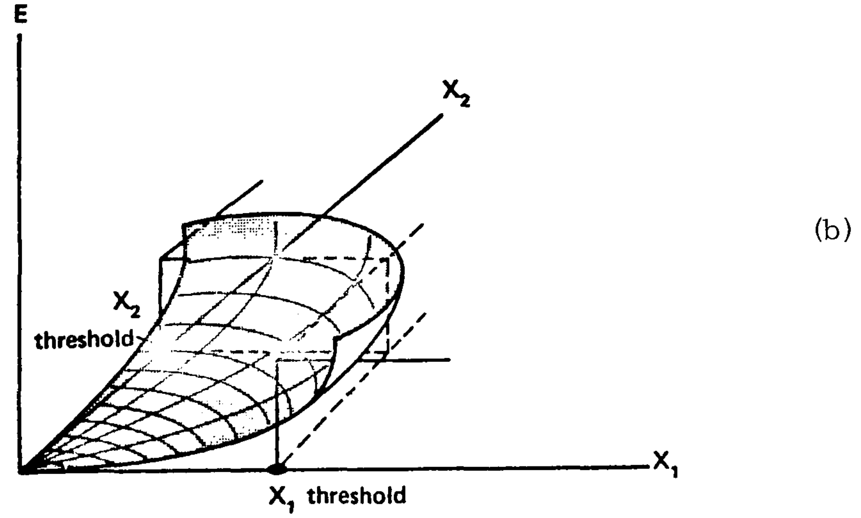
* s_i is used here as a threshold symbol instead of t_i to indicate that the threshold values for the disjunctive model may be different from those for the conjunctive model.



Disjunctive

E = 0 if $X_1 \leqslant X_1$ threshold and $X_2 \leqslant X_2$ threshold

E = 1 if $X_1 > X_1$ threshold or $X_2 > X_2$ threshold



Continuous, differentiable approx. to a disjunctive model within a region of desired application.

Figure A-2: Geometric Representations of the Disjunctive Model where the ρ_i are the asymptotic values of the function. They are arbitrarily set for each x_i, such that they exceed the maximum value that the x_i can attain.

The function is illustrated geometrically in Figure A-2(b). The nature of this function is such that a high value for any one of the component variables will produce a high value for the overall measure.

Performing a linear transformation, the function becomes:

$$\ln E = -\sum_{i=1}^{n} w_i \ln (\rho_i - x_i)$$

As in the case of the conjunctive model, the raw data were converted to a reverse scale, and the estimated function was a mirror image of the one being discussed.

APPENDIX B

AN ASSESSMENT OF THE USEFULNESS OF THE CANDIDATE EFFECTIVENESS MEASURES

This appendix presents the results of a survey in which a group of officials in the pilot test city were asked to assess the usefulness of the effectiveness measures, illustrated in Table 3 on pages 30 through 36. The group of evaluators consisted of representatives from the sanitation, health, and planning departments of that city. The evaluators separately assessed each variable in terms of how useful it would be to them in their decision-making needs. They were asked to assign one of the following three descriptors to each variable:

- important
- of limited value
- not important.

The variables were then scored as follows, based on the descriptors assigned:

2 = important
1 = of limited value
0 = not important.

Mean scores were computed by major measurement category (health, appearance, safety, etc.) and by solid waste activity (storage, collection, etc.). These are shown for all evaluators combined in Table B-1 on the following page.

The results indicate that in terms of solid waste activities, the most useful variables are those related to:

- Storage
- Collection
- Cleaning.

Table B-1. VARIABLE SCORES FOR THE CANDIDATE EFFECTIVENESS MEASURES BY SOLID WASTE ACTIVITY AND BY MEASUREMENT CATEGORY (Averages for all Evaluators Combined)

Measurement Category	Solid Waste Activity				All Activities Combined[a]
	Storage	Collection	Cleaning	Local Transportation	
Public Health	1.6	1.7	1.8		1.7
Public Safety	1.5	1.6	1.7		1.6
Appearance of Community	1.8	2.0	1.6	1.2	1.7
Odor	0.8	0.7			0.7
Satisfaction of Storage and Collection Needs	1.6	1.6			1.6
Compliance with Standards	1.6	1.4			1.5
Inconvenience to Public		1.1		0.8	1.0
All Categories Combined[a]	1.6	1.4	1.6	0.9	1.4

[a] Weighted averages.

In terms of measurement categories, the most useful variables are those related to:

- Public Health
- Public Safety
- Appearance of Community
- Satisfaction of Community's Storage and Collection Needs
- Compliance with Standards.

Table B-2 on pages 116 through 120 presents the score for each effectiveness measure by measurement category. It reflects the average score across all evaluators.

Table B-2. VARIABLE SCORES FOR EACH OF THE CANDIDATE EFFECTIVENESS MEASURES BY MEASUREMENT CATEGORY

(Averages for all Evaluators Combined)

Measurement Category	Measures of Effectiveness	Variable Score
PUBLIC HEALTH	— Average rat count for storage areas — Percent of storage areas where rat count exceeds given threshold	1.4
	— Average fly count for storage areas — Percent of storage areas where fly count exceeds given threshold	1.2
	— Percent of storage areas found to contain health hazards	1.8
	— Average health hazard rating for storage areas — Percent of storage areas where health hazard rating exceeds a given threshold	2.0
	— Percent of blocks where more than "X" percent of the storage areas are found to contain health hazards	1.6
	— Number of containers per block set out the night before collection which pose potential health hazards	1.8
	— Percent of inspections found to contain health hazards on the public way	1.8
	— Percent of inspections where health hazards are found on lots or public areas	1.8
PUBLIC SAFETY	— Number of fires caused partially by improper storage of combustible solid waste per 1,000 persons	1.2

Table B-2 (continued)

Measurement Category	Measures of Effectiveness	Variable Score
PUBLIC SAFETY (Continued)	— Percent of storage areas found to contain safety hazards	1.8
	— Average safety hazard rating for storage areas — Percent of storage areas where safety hazard rating exceeds a given threshold	1.6
	— Percent of blocks where more than "X" percent of the storage areas is found to contain safety hazards	1.6
	— Percent of inspections found to contain safety hazards on the public way	1.8
	— Percent of inspections where safety hazards are found on lots or public areas	1.6
APPEARANCE OF COMMUNITY	— Average appearance rating for storage areas — Percent of storage areas where appearance rating exceeds a given threshold	1.8
	— Percent of storage areas found to contain abandoned or discarded bulk items	1.8
	— Percent of blockfaces containing spilled or scattered refuse subsequent to collection (where curbside collection is performed) + Percent of alleys containing spilled or scattered refuse subsequent to collection (where alley collection is performed)	2.0
	— Percent of inspections where abandoned or discarded bulk items are observed	2.0
	— Percent of inspections where abandoned automobiles or trucks are observed	2.0

Table B-2 (continued)

Measurement Category	Measures of Effectiveness	Variable Score
APPEARANCE OF COMMUNITY (Continued)	— Average litter rating for streets and alleys — Percent of streets and alleys where litter rating exceeds a given threshold	1.8
	— Number of unsolicited complaints from citizens about the appearance of their community per 1,000 persons	1.2
	— Index of citizen satisfaction with the appearance of their community	1.2
	— Average litter rating for vacant lots and public areas	1.6
	— Number of drain basins which are clogged per basin inspected	2.0
	— Percent of collection fleet likely to cause spillage while in transport	1.2
ODOR	— Percent of storage areas found to contain offensive odors	0.8
	— Percent of blocks where more than "X" percent of the storage areas is found to contain offensive odors	0.8
	— Number of unsolicited complaints from citizens about the existence of offensive odors per 1,000 persons	0.6
SATISFACTION OF STORAGE & COLLECTION NEEDS	— Amount to which the combined storage-collection capacity falls short of "true" storage-collection needs per block (in pounds of refuse)	1.4
	— Percent of blocks where the combined storage-collection capacity falls short of the "true" storage-collection needs	1.4

Table B-2 (continued)

Measurement Category	Measures of Effectiveness	Variable Score
SATISFACTION OF STORAGE & COLLECTION NEEDS (Continued)	— Percent of storage areas found to contain an inadequate number of containers	1.8
	— Percent of blocks where more than "X" percent of the storage areas contains an inadequate number of containers	1.8
COMPLIANCE WITH STANDARDS	— Percent of storage areas having containers which do not comply with the regulations	1.6
	— Percent of blockfaces where mixed refuse remains uncollected for one or more days (where curbside collection is performed) + Percent of alleys where mixed refuse remains uncollected for one or more days (where alley collection is performed)	1.8
	— Average delay time in meeting regular pickup schedules for alleys/blockfaces	1.4
	— Number of unsolicited complaints from citizens about delays in pickup of mixed refuse per 1,000 persons	1.6
	— Percent of instances in which there was a delay of one day or more in the pickup of bulky items	1.6
	— Average delay time in the pickup of bulky items	1.0
	— Number of unsolicited complaints from citizens about delays in pickup of special or bulk items per 1,000 persons	1.2
INCONVENIENCE/ DISCOMFORT TO PUBLIC	— Average amount of time spent per household per month preparing refuse for collection	0.6
	— Percent of collection areas where noise from collection exceeds a given threshold	1.0

Table B-2 (continued)

Measurement Category	Measures of Effectiveness	Variable Score
INCONVENIENCE/DISCOMFORT TO PUBLIC (Continued)	— Number of collection miles taking place during early morning hours as a percent of total collection miles	1.4
	— Number of collection stops taking place during early morning hours as a percent of total collection stops	1.2
	— Number of unsolicited complaints from citizens about noise caused by refuse collection activities per 1,000 persons	1.0
	— Number of miles of major and secondary arterial roads where refuse collection is performed during peak hours — Amount of peak hour time during which refuse collection is taking place along major and secondary arterial roads	1.0
	— Number of reported instances of property damage caused by collection equipment or collection personnel per 1,000 persons	1.2
	— Total dollar value of property losses caused by collection equipment or collection personnel	1.2
	— Amount of peak hour time during which collection fleet is enroute to (or from) central deposit source	1.2
	— Percent of vehicles where air pollution rating exceeds a given threshold	0.4

APPENDIX C

DESCRIPTION OF THE SURVEY DESIGN AND ITS IMPLEMENTATION DURING THE PILOT TEST

This appendix describes the field survey design and how it was used during the pilot test phase of the project. The information provided in this appendix may be useful to communities in designing their own measurement systems to assess conditions (such as health hazards, safety hazards, appearance, and so forth) in various areas such as census tracts, sanitation districts, or entire cities.

The survey design is based on a concept as to the manner by which the data at particular observational points are blended together to produce estimates for designated areas. This concept enables one to determine the relative importance and impact of various sources of variability on the conditions that are being measured.

Some of the sources of variability whose relative importance should be determined are:

- Among days of the week
- Among routine observers
- Among observation points in the same block
- Among blocks in tracts
- Among tracts in sanitation districts
- Among sanitation districts.

In addition to providing estimates of variation from different sources, the survey design has a structure that permits unbiased estimates of the measured conditions to be made for alternative areas. Although the estimates as derived from this survey may not be sufficiently precise to be of practical value in themselves, they are sufficient to evaluate the soundness of the estimation methodology and to determine the size of possible future surveys so as to achieve a specified level of precision.

GENERAL DESCRIPTION OF THE SURVEY STRUCTURE

The survey structure is based on the concept of sampling from among and within strata. Strata are defined as mutually exclusive sets of census tracts; that is, there is no overlap in the area represented by a given stratum. The primary reason for formulating strata is to be able to group the sampling units in a manner that minimizes the sampling variation among units within a stratum. The reduction of variance within stratum tends to increase the precision of the estimates.

There are a number of ways to define strata. They could correspond to different geographical areas, as, for example, sanitation districts. Alternatively, they could be identified with income levels.

Once the strata are defined, two census tracts from each stratum are randomly selected, and ten blocks are selected at random from each tract. Thus, altogether, ten census tracts and a total of 100 blocks are covered. The ten blocks of a census tract are then assigned at random to form five block pairs: 1, 2, 3, 4, and 5. The numbering of the block pairs is independent from tract to tract; that is, there is no relationship between block pair j in census tract i and block pair j in some other census tract. Since there are 10 census tracts and 5 block pairs to each census tract, there are 50 block pairs.

Ten observers perform the data gathering. They are formed into five teams of two observers each. The teams are identified as teams A, B, C, D, and E. With respect to the basic survey design, a pair of observers inspects a total of ten block pairs. The 50 block pairs are thus divided into five block pair groups, and each pair of observers is assigned to a different block pair group.

This basic design is augmented in one tract by having all observers inspect all blocks in that tract. This tract is defined as the "intensive study" tract. Thus, each observer observes 28 blocks, composed of two blocks each in nine tracts and ten blocks in the special intensive study tract.

Because of the weekly mixed refuse collection, estimates of conditions at a given observation point vary from one day of the week to another. Consequently, any snapshot for an area should be a "composite blur" over the conditions of an entire week. For this reason, the five block pairs of a tract, in addition to being associated with different observer pairs, are also associated with different days of the week. Thus, at the census tract level, effects of days of the week will be completely confounded with the observer pair. When data are combined over tracts, it is possible, using analysis of variance techniques, to separate the effects of pairs of observers and days of the week.

Table C-1 on the following page shows the basic survey design before randomization. Each letter entry (M, T, W, Th, F) of the table (indicating the day of the week) refers to the inspection of a specified block pair by a specified observer pair in a designated tract. The subscripts of the letters refer to weeks 1 and 2 of the inspection period. Thus, each inspection team inspects all tracts and makes inspections on all five days of the week in a balanced manner. Each tract is inspected on each of the five days of the week by different observer pairs. Each day of the week is, therefore, balanced over tracts and observer pairs. In addition to the inspections indicated in this table, all observers inspect all blocks in a selected tract termed the "common" tract. These visits are all made at approximately the same time by all teams.

Before application of the design set out in Table C-1, various random selections and randomizations must occur. These are as follows:

(1) List census tracts in each of five strata and pick two census tracts at random from each stratum.

(2) Pick 10 blocks at random from each census tract.

Table C-1. BASIC SURVEY DESIGN[a]

			Observer Pair				
Stratum	Census Tract	A	B	C	D	E	Number of Block Pairs Per Tract
1	1	M_1^1	T_1	W_1	Th_1	F_1	5
	2	M_2^6	T_2	W_2	Th_2	F_2	5
2	1	T_1	W_1^3	Th_1	F_1	M_1	5
	2	T_2	W_2^8	Th_2	F_2	M_2	5
3	1	W_1	Th_1	F_1^5	M_1	T_1	5
	2	W_2	Th_2	F_2^{10}	M_2	T_2	5
4	1	Th_1	F_1	M_1	T_1^2	W_1	5
	2	Th_2	F_2	M_2	T_2^7	W_2	5
5	1	F_1	M_1	T_1	W_1	Th_1^4	5
	2	F_2	M_2	T_2	W_2	Th_2^9	5
Five Strata	Ten Census Tracts	Number of Block Pairs Per Observer Pair					
		10	10	10	10	10	

[a] The superscripts on the day codes indicate the order of visits to be made by the person who performed the training. His first inspection is indicated by the superscript 1, the second by the superscript 2, and so on. The observational pattern is balanced over strata, observer pairs, and days of the weeks. The subscripts refer to weeks 1 and 2 of the inspection period.

(3) Randomly form the 10 blocks within a census tract into five block pairs. Number these from 1 to 5. Each block pair should be identified by the stratum (i), the tract within the stratum (j), and the block pair (k). The latter are to be numbered from one to five. The blocks within a pair should be identified by the subscript (m). The values 1 and 2 taken by the subscript m should be assigned at random to the two blocks of a pair.

(4) Randomly associate the letters A, B, C, D, and E to the five block pairs in each tract. There should be a new randomization for each tract.

(5) Randomly form five pairs of observers from 10 routine observers. Randomly associate the letters A, B, C, D, and E to the five pairs of observers. This is done only once.

(6) Randomly permute the columns of the main body of Table C-1 (that portion that contains the letters indicating days). After that is done, also randomly permute pairs of rows; that is, the two rows associated with a stratum should be treated as one row in the permutation.

SURVEY PLAN AS UTILIZED IN FIELD TEST

The initial survey plan for the field test conducted in the City of Baltimore utilized the city's five sanitation districts as the strata. However, when members of the project team surveyed the census tracts selected from each stratum, they found that there was not a sufficient amount of variation in the conditions among census tracts to test the measurement apparatus. For this reason, an alternative scheme that would ensure an adequate amount of variation was used for defining the strata. The revised strata were defined as follows:

- Dirty Stratum—Those census tracts the sanitation and health department personnel defined as particularly dirty.

- Model Cities Stratum—Those census tracts contained within the Model Cities areas of the city.

- Income Stratum No. 1—Those census tracts where the average family income in 1970 was less than $9,000.

- Income Stratum No. 2—Those census tracts where the average family income in 1970 was between $9,000 and $11,999.

- Income Stratum No. 3—Those census tracts where the average family income in 1970 was $12,000 or more.

To eliminate any overlap among strata, census tracts belonging to the Dirty Stratum were classified first, followed by tracts belonging to the Model Cities Stratum. The group of census tracts that remained were then classified into one of the three income-related strata. The total number of census tracts in each stratum were as follows:

Strata	Total Number of Census Tracts in Strata
Dirty	15
Model Cities	16
Income Level 1	59
Income Level 2	83
Income Level 3	28

Within each stratum, two census tracts were selected at random, and from each census tract ten blocks were randomly selected. The geographical distribution of the census tracts selected for the field survey within the City of Baltimore is shown in Figure C-1 on the following page. A breakdown of the survey census tracts by sanitation district and by income grouping is provided in Table C-2 on page 128. A similar breakdown for the entire City of Baltimore is presented in Table C-3 on the same page for comparative purposes.

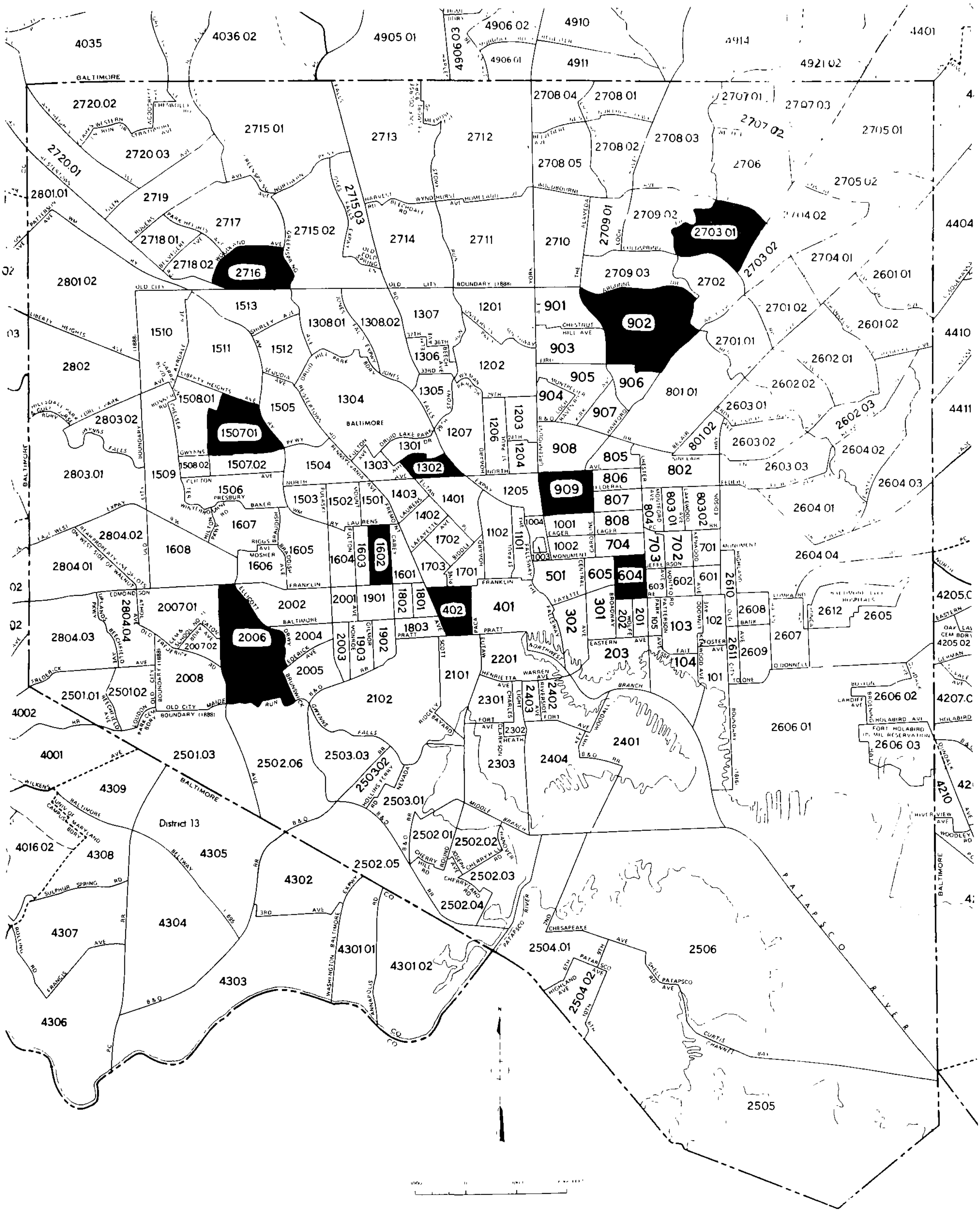


Figure C-1: Geographical Distribution of Baltimore City Census Tracts Included in the Field Test

Table C-2. DISTRIBUTION OF FIELD TEST CENSUS TRACTS BY INCOME GROUPING AND SANITATION DISTRICT

Sanitation District	Income Groupings [a]				Total
	$3,000- $5,999	$6,000- $8,999	$9,000- $11,999	$12,000 and over	
Northeast			1	1	2
Northwest		2		1	3
Eastern	2				2
Western	1	1	1		3
Central					0
Total	3	3	2	2	10

[a] Based on the average family income for 1970. There were no census tracts where the average family income in 1970 was less than $3,000.

Table C-3. DISTRIBUTION OF BALTIMORE CITY CENSUS TRACTS BY INCOME GROUPING AND SANITATION DISTRICT

Sanitation District	Income Groupings [a]				Total
	$3,000- $5,999	$6,000- $8,999	$9,000- $11,999	$12,000- and over	
Northeast	0	5	25	9	39
Northwest	0	15	18	14	47
Eastern	8	24	18	2	52
Western	11	23	26	2	62
Central	0	0	0	1	1
Total	19	67	87	28	201

[a] Based on the average family income for 1970. There were no census tracts where the average family income in 1970 was less than $3,000.

APPENDIX D

THE DATA COLLECTION FORMS AND PROCEDURES

This appendix contains replicas of the data collection forms that were developed to facilitate the collection of data during the field test. It includes the following five forms:

- Pre-Survey Form—Provides identifying information on the block to be surveyed.
- Blockfaces, Alleys, and Private Ways Form—Provides information on the conditions observed along blockfaces, alleys, and front yards and sidewalks.
- Storage and Backyard Area Form—Provides information on the conditions observed in storage and backyard areas.
- Vacant Lots, Public Parks, and Parking Lots Form—Provides information on the conditions observed on lots and parks.
- Summary Form—Provides information about collection and cleaning schedules and the length of time it took to collect the field data.

The information contained/requested on each of the five forms is briefly explained in this appendix. Detailed recording procedures for the forms were provided in the instruction booklet that was developed for the field test.

THE PRE-SURVEY FORM

The Pre-Survey Form provides information useful in identifying and locating the census block to be inspected. A replica of this form is provided in Figure D-1 on the following page. A

PRE-SURVEY FORM

Block Map:

Tract Number ______

Block Number ______

Sanitation District ______

Days of:

- regular collection ______
- bulk collection ______
- street cleaning ______
- alley cleaning ______

Number of:

- blockfaces ______
- alleys ______
- open spaces ______
- households ______

Procedures for Selection of Storage and Backyard Areas:

Special Notes:

Figure D-1: Replica of the Pre-Survey Form completed copy of this form was provided for each block that was surveyed during the field test. It contained:

- A hand-drawn map of the block to be surveyed.
- A set of code numbers to be used to identify the census block under inspection.
- The various collection and cleaning schedules that apply to the block.
- Some descriptive information about the block.
- Procedures to use when selecting storage and backyard areas for inspection.

The Block Map, which was hand-drawn on the left-hand side of the form, contained numeric codes between 1 and 19 to identify each blockface. These were to be used when recording information about blockfaces and private ways on the data collection forms. The Block Map also contained numeric identifiers for alleys. These ranged from 21 to 29 and were to be used when recording information about alleys on the data collection forms.

BLOCKFACES, ALLEYS, AND PRIVATE WAYS FORM

The Blockfaces, Alleys, and Private Ways Data Form is intended to be used for inspections made in the following areas:

Blockface: The area from the center of the street up to and including the curb and gutter, extending from any corner of a block to the adjacent corner.

Alley: A passageway, usually 5 to 10 feet wide, extending into or through the interior of a block.

Private Way (Sidewalk and Front Yard): The area from the front of the house to the curb, extending from any corner of a block to the adjacent corner.

A replica of this data form is provided in Figure D-2 on the following page.

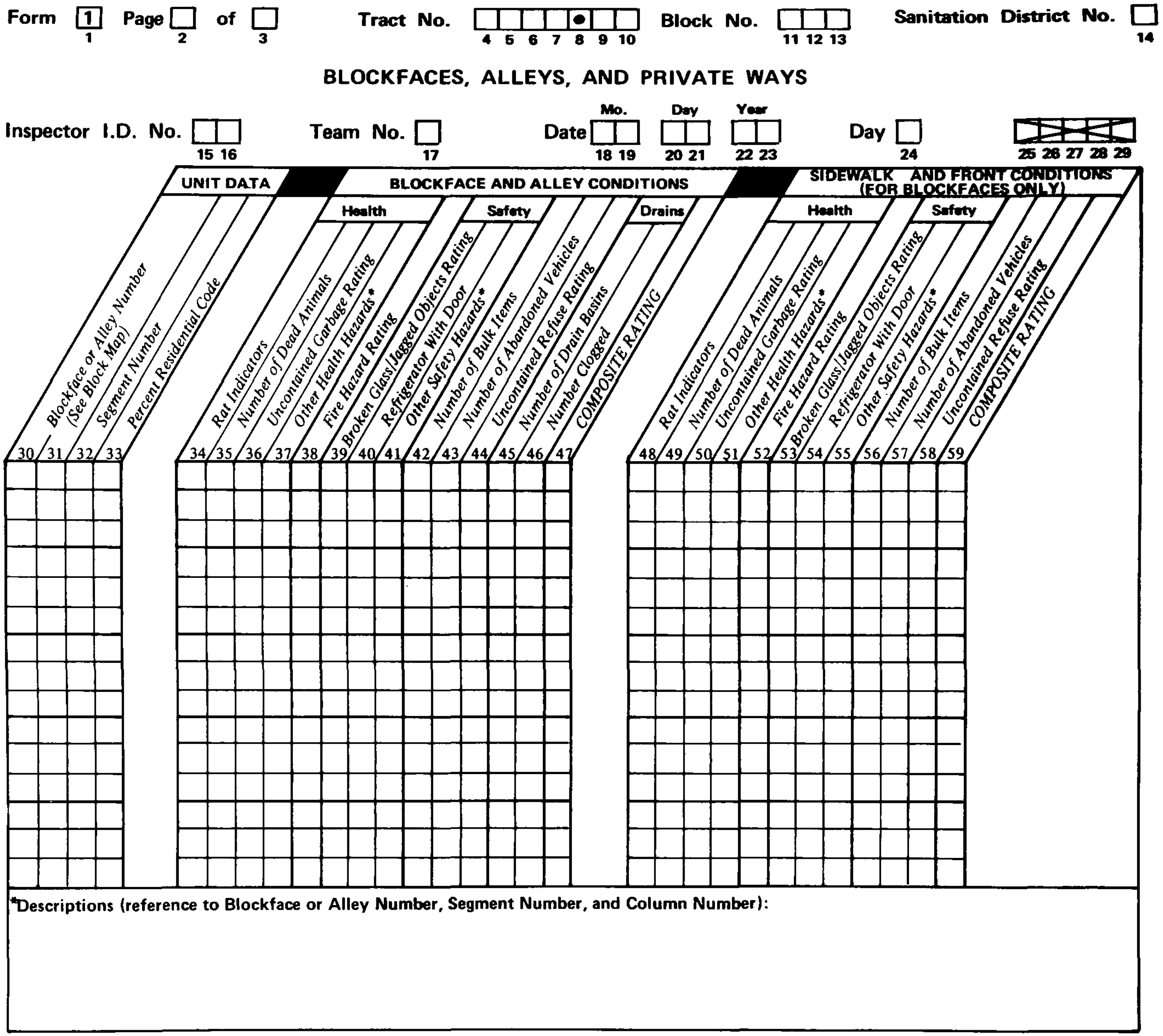
The above three areas were inspected in linear segments of approximately 100 to 200 feet each. Data for each blockface and private way segment were recorded on a separate line on the form. Similarly, data for each alley segment were recorded on a separate line. All entries on each line used were to be filled in, unless otherwise specified. Where one form was not sufficient to record the information on blockfaces, alleys, and private ways, additional data forms were used. A separate form was used for each block that was surveyed.

The form is divided into four major sections:

- Locator Data
- Unit Data
- Blockface and Alley Conditions
- Sidewalk and Front Yard Conditions.

Locator Data is contained at the top of the form in the boxes corresponding to items 1-29. It provides information by which to identify the census block being inspected, the date of the inspection, and the inspection team. Unit data, contained in columns 30-33 of the form, provides information by which to identify each sample area inspected. Blockface and Alley Conditions, contained in columns 34-47, provides information on the type of conditions observed along each blockface or alley segment inspected. Sidewalk and Front Yard Conditions, contained in columns 48-59, provides information on the types of conditions observed along the private way corresponding to the blockface segment being inspected.

A summary description of the entries on this form (excluding those connected with the Locator Data at the top of the form) is provided on pages 134 through 136.



Form ☐1 Page ☐ of ☐ Tract No. ☐☐☐☐●☐☐ Block No. ☐☐☐ Sanitation District No. ☐
1 2 3 4 5 6 7 8 9 10 11 12 13 14

BLOCKFACES, ALLEYS, AND PRIVATE WAYS

Inspector I.D. No. ☐☐ Team No. ☐ Date Mo. ☐☐ Day ☐☐ Year ☐☐ Day ☐ ☒☒☒☒☒
15 16 17 18 19 20 21 22 23 24 25 26 27 28 29

UNIT DATA			
Blockface or Alley Number (See Block Map)		Segment Number	Percent Residential Code
30	31	32	33

BLOCKFACE AND ALLEY CONDITIONS													
Health					Safety						Drains		
Rat Indicators	Number of Dead Animals	Uncontained Garbage Rating	Other Health Hazards*	Fire Hazard Rating	Broken Glass/Jagged Objects Rating	Refrigerator With Door	Other Safety Hazards*	Number of Bulk Items	Number of Abandoned Vehicles	Uncontained Refuse Rating	Number of Drain Basins	Number Clogged	COMPOSITE RATING
34	35	36	37	38	39	40	41	42	43	44	45	46	47

SIDEWALK AND FRONT CONDITIONS (FOR BLOCKFACES ONLY)											
Health					Safety						
Rat Indicators	Number of Dead Animals	Uncontained Garbage Rating	Other Health Hazards*	Fire Hazard Rating	Broken Glass/Jagged Objects Rating	Refrigerator With Door	Other Safety Hazards*	Number of Bulk Items	Number of Abandoned Vehicles	Uncontained Refuse Rating	COMPOSITE RATING
48	49	50	51	52	53	54	55	56	57	58	59

*Descriptions (reference to Blockface or Alley Number, Segment Number, and Column Number):

Figure D-2: Replica of the Blockfaces, Alleys, and Private Ways Data Form

Blockface or Alley Number — The identifying number of the blockface or alley segment being inspected, as indicated on the Block Map.

Segment Number — A sequential number assigned by the observers to each segment.

Percent Residential Code (blockfaces only) — A code number to describe the percent of residential structures along the blockface segment being inspected; it was recorded as follows:

1 = 0 percent residential
2 = 1 - 24 percent residential
3 = 25 - 49 percent residential
4 = 50 - 74 percent residential
5 = 75 - 99 percent residential
6 = 100 percent residential
9 = Not Applicable — an alley is being inspected.

Rat Indicators — Used to indicate the presence of rat signs; i. e., sighting of live or dead rats, rat gnawings, rat burrows, rat feces, and rat tracks.

Number of Dead Animals — The number of dead or decaying animals observed.

Uncontained Garbage Rating — Recorded on a scale of 1 to 7, with the scale points reflecting varying degrees of uncontained garbage as shown below:

1 = No uncontained garbage is observed
3 = Minor amounts of uncontained garbage are observed
5 = Moderate amounts of uncontained garbage are observed
7 = Substantial amounts of uncontained garbage are observed *or* garbage accumulation shows signs of rat or insect attraction.

Garbage is defined as waste resulting from the preparation, cooking, serving, or eating of food.

Other Health Hazards — Used to indicate the presence of conditions, not on the data form, that the observer considers to be potential contributors to disease or illness.

Fire Hazard Rating — A code number describing the presence of fire hazards. It was recorded as follows:

1 = No fire hazards observed
2 = Minor fire hazard exists
3 = Major fire hazard exists.

A major fire hazard is defined as an accumulation of solid waste materials sufficient to cause or contribute to a fire of such magnitude that property damage or personal injury is likely to occur. A minor fire hazard is defined as an accumulation of solid waste materials sufficient to cause or contribute to a fire, but unlikely to cause personal injury or property damage.

Broken Glass/Jagged Objects Rating — Recorded on a scale of 1 to 7, with the scale points reflecting varying degrees of glass, jagged objects, etc., as shown below:

1 = No broken glass, jagged objects, etc., are observed
3 = Minor quantities of broken glass, jagged objects, etc., are observed
5 = Moderate amounts of broken glass, jagged objects, etc., are observed
7 = Substantial amounts of broken glass, jagged objects, etc., are observed.

The items to be included are: broken glass, pieces of barbed wire, and other sharp or jagged objects.

Refrigerator With Door — Used to indicate the presence of a refrigerator with a door intact.

Other Safety Hazards — Used to indicate the presence of conditions, not on the data form, that the observer considers to be potential hazards to public safety.

Number of Bulk Items — The number of bulk items observed. Bulk items are defined as items that cannot fit into a storage container. These may include discarded furniture items or appliances, shipping cases, carpeting, automobile tires, and so forth.

Number of Abandoned Vehicles — The number of abandoned automobiles and abandoned trucks observed. Abandoned vehicles are vehicles which appear to be in an apparently inoperative condition. They are generally characterized by a lack of licenses or inspection stickers, or by expired licenses or stickers.

Uncontained Refuse Rating — Recorded on a scale of 1 to 7, with the scale points reflecting varying degrees of uncontained refuse as shown below:

1 = No uncontained refuse is lying on the ground
3 = A minor amount of uncontained refuse is observed
5 = Moderate amounts of uncontained refuse are observed
7 = Substantial amounts of uncontained refuse are observed.

The rating should reflect conditions, exclusive of what is included in making ratings of uncontained garbage, fire hazards, and broken glass/jagged objects.

Number of Drain Basins (blockfaces and alleys only) — The total number of storm drain basins observed.

Number of Clogged Drain Basins (blockfaces and alleys only) — The number of drain basins that appear to be clogged by debris.

Composite Rating — Recorded on a scale of 1 to 8, with 1 indicating the most favorable overall conditions and 8 indicating the worst overall conditions. This reflects the observer's subjective assessment of the overall conditions observed.

STORAGE AND BACKYARD AREA FORM

The Storage and Backyard Area Data Form is intended to be used for inspection of areas where solid waste materials are stored and for inspection of private areas (backyards) surrounding the storage locations. These sample areas are defined as follows:

> Solid Waste Storage Area: An area external to the structure which normally serves as a location for the containment of discarded solid waste materials; i. e., the place where waste storage containers are located.
>
> Private Area (Backyard): The area belonging to the back of the structure served by the storage area. It is bounded by the property lines of the adjacent structures and, in many cases, by an alley. The alley itself, however, is not part of the inspection area.

A replica of this data form is provided in Figure D-3 on the following page.

Only a sample of storage areas and backyards was to be inspected within the survey block. Procedures for selection of these areas were provided on the Pre-Survey Form for the block. Observers were asked to record data based only on what they observed. That is, they were not required to open the containers and determine their contents. In instances where there was no alley access to the storage/backyard area, the observers were to ask permission of the resident in order to gain access to these areas. Data for each storage area/backyard combination was to be recorded on a separate line on the form.

The format of the Storage and Backyard Area Data Form is similar to that of the Blockfaces, Alleys, and Private Ways Data Form. It is divided into the same four major sections and required many of the same types of measurements to be made.

A summary description of the entries that are unique to this form is provided on pages 139 through 141.

Form [2] Page ☐ of ☐ Tract No. ☐☐☐☐●☐☐ Block No. ☐☐☐ Sanitation District No. ☐
1 2 3 4 5 6 7 8 9 10 11 12 13 14

STORAGE AND BACKYARD AREA

Inspector I.D. No. ☐☐ Team No. ☐ Date Mo. ☐☐ Day ☐☐ Year ☐☐ Day ☐ ☒☒☒☒☒
15 16 17 18 19 20 21 22 23 24 25 26 27 28 29

UNIT DATA								STORAGE AREA CONDITIONS																							BACKYARD CONDITIONS											COMPOSITE RATING
								Size of Cans and Bins					Health							Safety					Condition of Cans or Bins						Health					Safety						
Sample Area Number		Blockface No. (See Block Map)		Structure Code	Collection Responsibility Code	No. Regular Collections/Week	Inspection Conditions Code	No. 10 Gallon Cans	No. 20 Gallon Cans	No. Other Cans*	No. Small Bins	No. Large Bins	Rat Indicators	Insect Indicators	No. of Dead Animals	Improperly Containerized Garbage	Uncontained Garbage Rating	Other Health Hazards*	Fire Hazard Rating	Brok. Glass/Jag. Obj. Rating	Refrigerator With Door	Other Safety Hazards*	No. of Bulk Items	No. of Unapproved Containers	Total No. of Non-complying Cans or Bins	No. of Cans or Bins in Poor Condition	No. of Cans With Size Violations	No. of Cans or Bins Without Tight Lid	Odors	Uncontained Refuse Rating	Rat Indicators	Number of Dead Animals	Uncontained Garbage Rating	Other Health Hazards*	Fire Hazard Rating	Brok. Glass/Jag. Obj. Rating	Refrigerator With Door	Other Safety Hazards*	No. of Bulk Items	No. of Abandoned Vehicles	Uncontained Refuse Rating	
30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
0	1																																									
0	2																																									
0	3																																									
0	4																																									
0	5																																									
0	6																																									
0	7																																									
0	8																																									
0	9																																									
1	0																																									
1	1																																									
1	2																																									
1	3																																									
1	4																																									
1	5																																									

*Descriptions (reference to Sample Area Number and Column Number):

Figure D-3: Replica of the Storage and Backyard Area Data Form

Sample Area Number — A pre-recorded number that was used as a reference number for each storage area/backyard being inspected.

Blockface Number — The identifying number of the blockface upon which the structure associated with the sample area was located.

Structure Code — A code number describing the type of structure associated with the sample area being inspected; it was recorded as follows:

1 = Residential
2 = Apartment Complex
3 = Restaurant, Fast Food Establishment
4 = Combination Residential and Business
5 = Business Only
6 = Public Building
7 = Industrial
8 = Other (Specify: ____________________)
9 = None.

Collection Responsibility Code — A code number describing the group responsible for mixed refuse collection; it was recorded as follows:

1 = Collection performed by city sanitation department
2 = Collection performed by private contractor
3 = Unknown (do not know who performs collection).

Number of Regular Collections Per Week — The number of times per week that mixed refuse is collected, as indicated on the Pre-Survey Form.

Inspection Conditions Code — A code number used to describe whether there were problems that prevented inspection of all or part of the designated sample area and, if so, the nature of the problem. It was recorded as follows:

1 = No problems encountered which hindered the inspection of either the storage area or the backyard
2 = No storage area — structure abandoned/uninhabited
3 = No storage area — structure inhabited
4 = No storage area — unable to determine if structure is inhabited
5 = Containers are at end of alley or along curb for pickup
6 = High fence — inspection could not be made
7 = No alley access to storage/backyard area — resident not at home, or resident would not allow access
8 = Other (Specify: ______________________)
9 = No clearly defined backyard area.

Size of Cans and Bins (storage areas only) — Used to record the number of storage containers by size and type of container.

Insect Indicators (storage areas only) — Used to indicate the presence of insect signs; i.e., sighting of swarming or crawling insects, sighting of insect larvae, and so forth.

Improperly Containerized Garbage (storage areas only) — A code number to describe the manner in which garbage is containerized; it was recorded as follows:

1 = All garbage is properly stored in covered cans or bins
2 = Garbage is lying open in at least one can or bin and/or garbage is contained in paper sacks, plastic bags, or other similar containers
3 = There are uncovered cans or bins and/or there are plastic bags, paper sacks, etc., but it is not possible to tell whether or not they contain garbage.

A code 2 condition was to be reported in preference to any other code that may fit the storage area under observation.

Number of Unapproved Containers (storage areas only) — The total number of plastic bags, paper sacks, wooden or cardboard boxes, and so forth, which are used as containers for solid waste materials.

Number of Non-Complying Cans or Bins (storage areas only) — The total number of cans or bins found in the storage area that do not comply with established city regulations. For the field test city, these regulations required that cans and bins:

- be constructed of metal
- have tight fitting lids (if filled or partially filled)
- be free of holes
- be not larger than 20 gallons.

Number of Cans or Bins in Poor Condition (storage areas only) — The number of cans or bins of a combustible nature and/or having holes observed in the storage area. A combustible can is one that is not constructed of metal.

Number of Cans With Size Violations (storage areas only) — The number of cans larger than the allowable size observed in the storage area. For the field test city, any can larger than 20 gallons was considered a size violation.

Number of Cans or Bins Without Tight Lid (storage areas only) — The number of cans or bins that lack tight lids. Included would be cans or bins with no lids, cans or bins with bent or obviously loose lids, and overpacked cans or bins for which the lid will not fit tightly.

Odors (storage areas only) — Used to indicate the presence of odors due to poor storage conditions.

VACANT LOTS, PUBLIC PARKS, AND PARKING LOTS FORM

The Vacant Lots, Public Parks, and Parking Lots Data Form is intended for use in the areas defined as follows:

Vacant Lots: Open spaces that serve no business or residential purpose.

Public Parks: Playgrounds, recreational, or scenic areas serving the public.

Parking Lots: Paved open areas, usually having lined spaces and identifying signs, that are used as places for parking vehicles.

A replica of this form is provided in Figure D-4 on the following page.

All vacant lots, public parks, and parking lots within the boundary of the designated block were to be inspected with the use of this form. Data for each vacant lot, public park, or parking lot inspected was to be recorded on a separate line on the form.

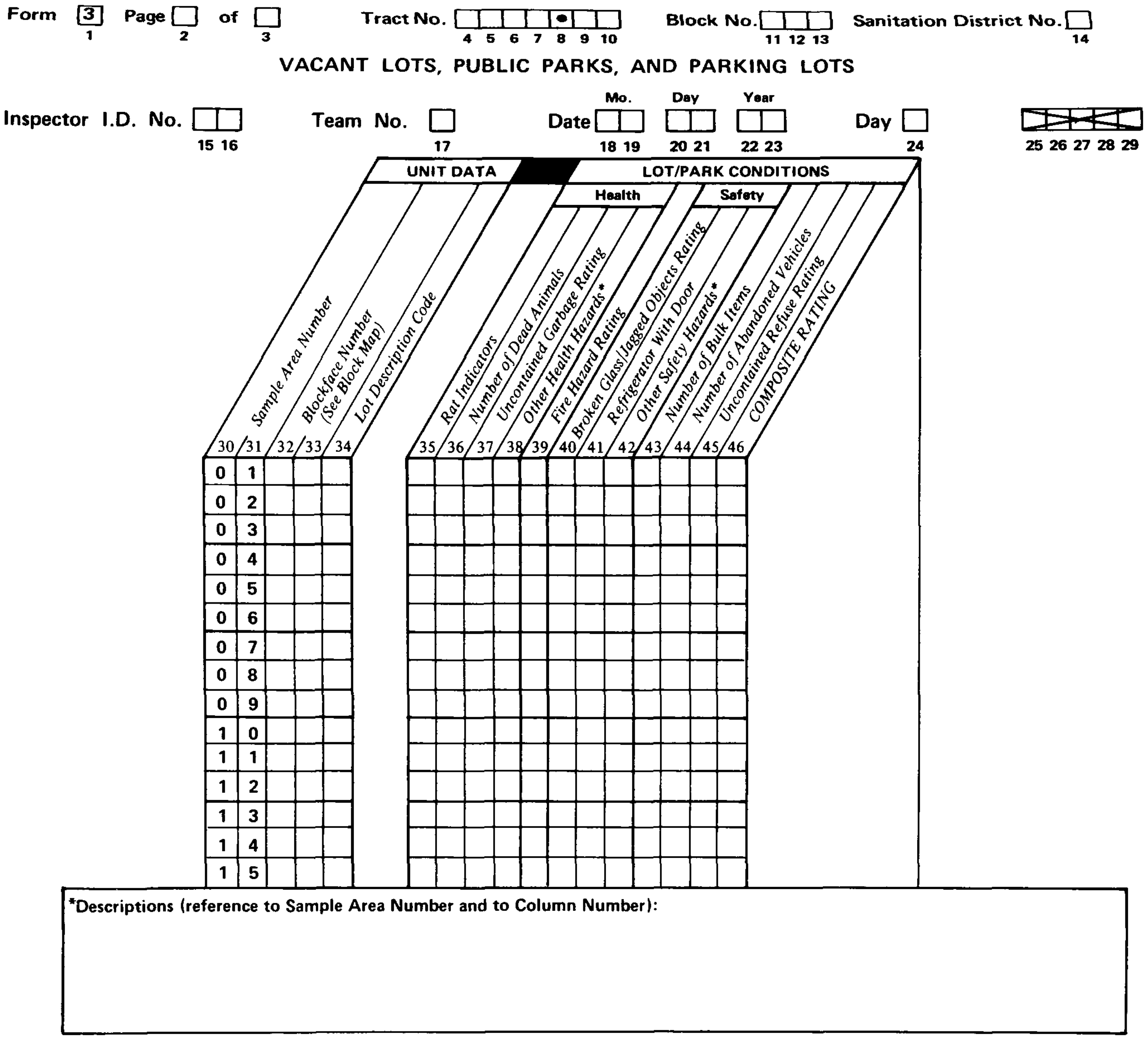
The format of this form is also similar to that of the two previous forms. A summary description of the entries that are unique to this form is provided below.

Sample Area Number — A pre-recorded number that was used as a reference number for each separate vacant lot, public park, or parking lot being inspected.

Blockface Number — The identifying number of the blockface upon which the lot or park is located.

Lot Description Code — A code number used to identify the inspection area; it was recorded as follows:

1 = Vacant Lot
2 = Public Park
3 = Parking Lot.



Form ☒3 (1) Page ☐ (2) of ☐ (3)   Tract No. ☐☐☐☐●☐☐ (4 5 6 7 8 9 10)   Block No. ☐☐☐ (11 12 13)   Sanitation District No. ☐ (14)

VACANT LOTS, PUBLIC PARKS, AND PARKING LOTS

Inspector I.D. No. ☐☐ (15 16)   Team No. ☐ (17)   Date Mo. ☐☐ (18 19) Day ☐☐ (20 21) Year ☐☐ (22 23)   Day ☐ (24)   ☒☒☒☒☒ (25 26 27 28 29)

UNIT DATA					LOT/PARK CONDITIONS											
						Health				Safety						
Sample Area Number 30	31	Blockface Number (See Block Map) 32	33	Lot Description Code 34	Rat Indicators 35	Number of Dead Animals 36	Uncontained Garbage Rating 37	Other Health Hazards* 38	Fire Hazard Rating 39	Broken Glass/Jagged Objects Rating 40	Refrigerator With Door 41	Other Safety Hazards* 42	Number of Bulk Items 43	Number of Abandoned Vehicles 44	Uncontained Refuse Rating 45	COMPOSITE RATING 46
0	1															
0	2															
0	3															
0	4															
0	5															
0	6															
0	7															
0	8															
0	9															
1	0															
1	1															
1	2															
1	3															
1	4															
1	5															

*Descriptions (reference to Sample Area Number and to Column Number):

Figure D-4: Replica of the Vacant Lots, Public Parks, and Parking Lots Data Form

THE SUMMARY FORM

The Summary Form is used to provide information on collection and cleaning activities as they relate to the day that the inspection was made and on block conditions relative to collection and cleaning activities. It also provides an estimate of the length of time that it took to collect the requested information about the block. Observers were to complete one copy of this form for each block they inspected. A replica of this data form is provided in Figure D-5 on the following page.

The Summary Form is organized as follows:

- Identifying Information — Contained in the upper left-hand corner of the form.
- Time Since Collection and Cleaning — Contained in the upper right-hand corner of the form.
- Start and Finish Time for Data Collection — Contained in the upper middle section of the form.
- Composite Block Rating — Contained in the upper middle section of the form.
- Selected Observations on Block Conditions — Questions 1 through 7 on the form.

A summary description of information requested on this rm is provided below:

Identifying Information — Used to identify the observer making the inspection and the sample block being surveyed.

Time Since Collection and Cleaning — Used to indicate the elapsed time since regular collection, bulk pick-up, and street and alley cleaning. This information was recorded by comparing the day that the inspection was made with the last scheduled day for collection and cleaning, respectively, as shown on the Pre-Survey Form. Zero (0) was to be recorded if the observer was there on a collection/cleaning day subsequent to the time the collection or cleaning took place. If collection or cleaning was ongoing at the time the observer arrived, he (or she) was to wait and inspect the area subsequent to the completion of the activity.

SUMMARY FORM

Name ____________________
Tract No. ____________________
Block No. ____________________
District No. ____________________

No. of Days Since Regular Collection ________
No. of Days Since Bulk Collection ________
No. of Days Since Street Cleaning ________
No. of Days Since Alley Cleaning ________

Start Time: ________ am/pm
Finish Time: ________ am/pm
Composite Block Rating ________ (1-8)

1. Was this the regular collection day for mixed refuse?
   ____ Yes ____ No

2. If yes to Q. 1, were you there
   ____ before collection?
   ____ after collection?

3. If yes to Q. 1, identify the place of collection by blockface (e.g., along alley, at end of alley, at curbside, from backyard):

   Blockfaces

   1. ____________
   2. ____________
   3. ____________
   4. ____________
   5. ____________
   6. ____________
   7. ____________
   8. ____________
   9. ____________
   10. ____________
   11. ____________
   12. ____________

4. Was there any uncollected refuse along any blockface or alley?
   ____ Yes ____ No

5. If yes to Q. 4, identify the type of uncollected refuse (e.g., mixed refuse, bulk items, and so forth) and the blockface/alley number.

Blockface/Alley No.	Type of Refuse
____________	____________
____________	____________
____________	____________
____________	____________
____________	____________
____________	____________
____________	____________
____________	____________
____________	____________

6. Were any of the streets or alleys being cleaned while you were there?
   ____ Yes ____ No

7. If yes to Q. 6, provide the appropriate blockface and/or alley numbers on the line below.
   ____

Figure D-5: Replica of the Summary Form

Start and Finish Time for Data Collection — Used to indicate the length of time that it took to collect data for the survey block.

Composite Block Rating — Recorded on a scale of 1 through 8, with 1 indicating the most favorable overall conditions and 8 indicating the worst overall conditions. This reflects the observer's subjective assessment of overall block conditions.

Selected Observations on Block Conditions — Used to obtain further information about block conditions relative to collection and cleaning activities.

APPENDIX E

DESCRIPTION OF THE ANALYSIS PLAN

This appendix provides a detailed description of how some of the major findings presented in Chapter VI of the report were developed. Specifically, it describes the techniques used to:

- Analyze the consistency (or reliability) with which the measurements were made, particularly the rating scale measurements because of the higher degree of subjectivity associated with these types of measurements.
- Estimate the variance components for a census tract mean.
- Assess the accuracy of the measurements.
- Determine correlations among the six observational areas and among the variables.
- Develop composite measures of effectiveness.

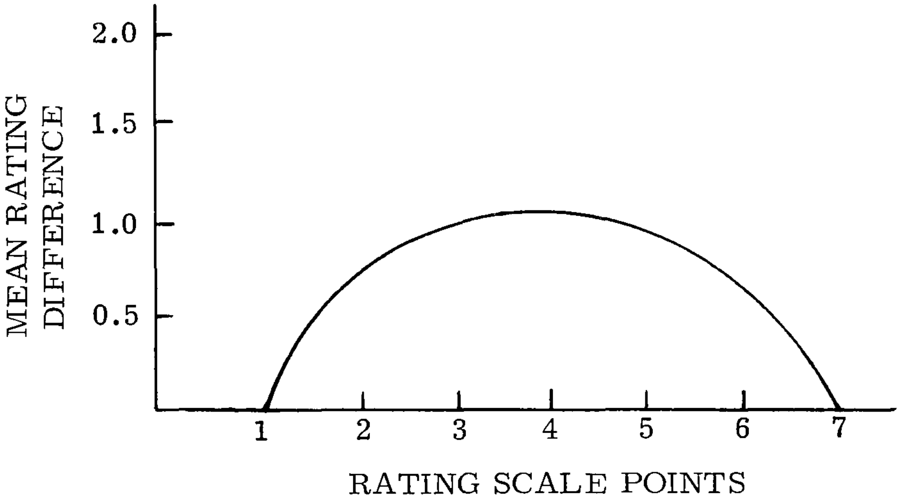
EVALUATING THE CONSISTENCY AMONG RATERS

Reliability or consistency refers to the ability of two or more observers to independently assign the same value when measuring the same phenomena. In assessing the reliability of the data collected during the field test, separate techniques were developed for each of the three generic types of measures—rating scales, counts, and yes-no type measures. These are presented in the following paragraphs.

Consistency Among Raters—Rating Scales

As described in Chapter V, the observers worked in pairs, with each person separately recording his (or her) measurements on the data collection forms. To evaluate the reliability of the rating scale measurements, the mean discrepancy between all observer pairs at each of the seven scale points was analyzed. The analysis was performed separately for each type of rating scale—garbage, glass, and refuse—in each of the six observational areas where data were collected (blockfaces, alleys, storage areas, etc.). The results were subsequently combined across the six observational areas by type of rating to develop the findings presented in Chapter VI.

In performing the analysis, we hypothesized that a relationship of the following type would exist between the size of the mean rating difference and the rating scale points.



That is, we expected to find fairly close agreement between observers for conditions 1 and 7 (the scale extremes); i.e., mean rating differences approximately equal to zero. Toward the midpoint of the scale, however, we expected to find much less agreement; i.e., mean differences much greater than zero. We felt that observers were likely to have no trouble recognizing extremely good and extremely poor conditions, whereas conditions in between the two extremes were likely to pose the most problems in terms of rater agreement.

To test the hypothesis, we let:

q_i = value assigned by Rater Q to the i^{th} observation point (where Rater Q was one member of a rater pair)

p_i = value assigned by Rater P to the i^{th} observation point (where Rater P was the other member of the rater pair)

d_i = value of the absolute difference between a rater pair at the i^{th} observation point; i.e., $/q_i - p_i/$

For each observational area, we first looked at the discrepancies between rater pairs (d_i) and assumed that Rater Q was correct in his assignment of rating values, q_i. By cross-tabulating the discrepancies between rater pairs with the values that Rater Q assigned, we obtained a matrix that contained:

- Frequencies for each cell
- Frequencies and percents for each row and column
- Values for the mean difference between rater pairs at each of the 7 rating values, along with the standard deviations associated with each mean difference.

The matrix looked as follows:

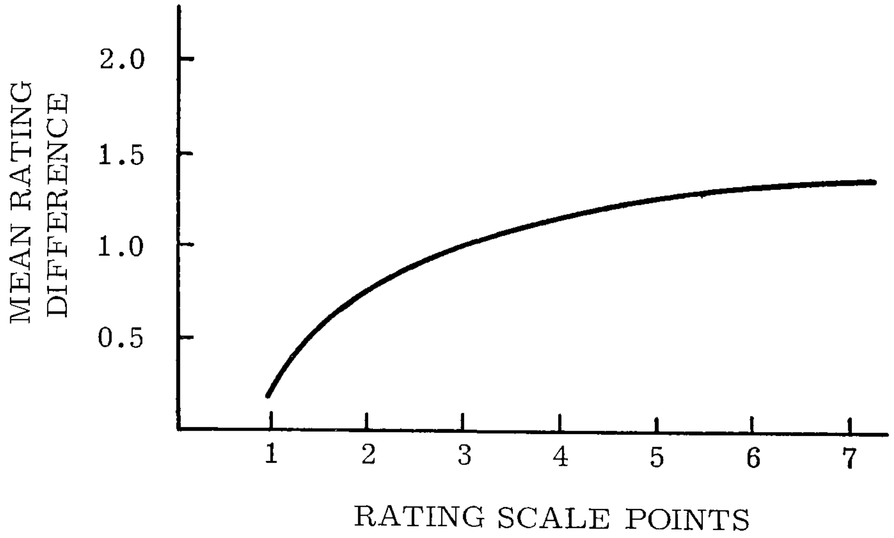
Value Assigned by Rater Q (q_i)	Absolute Value of the Discrepancy Between Rater Q and Rater P (d_i)							Total Frequency Percent of Total	Mean Value Standard Deviation
	0	1	2	3	4	5	6		
1									
2									
3									
4									
5									
6									
7									
Freq. % of Total									

Our particular interest from an analytical point of view was the information contained in the last column; namely, the mean difference by scale point and its standard deviation.

Now, since we had no a priori reason to assume that Rater Q was the correct rater, we proceeded to look at the discrepancies between rater pairs, on the assumption that Rater P was correct in his assignment of rating values, p_i. When the discrepancies in the observations of the two raters were cross-tabulated with the values assigned by Rater P, we obtained a matrix identical in form to the one presented above, having Rater P's assigned values.

The mean differences by scale point that resulted from the two cross-tabulations were combined and a weighted average was developed. This average reflected the mean difference across all observer pairs for each scale point of a given rating scale. This procedure was repeated for each of the six observational areas. The resultant data were pooled to develop new weighted averages for the mean rating differences across all observational areas. The standard error of the mean and the 95 percent confidence interval about the mean rating differences were then computed for each scale point.

When a plot of the mean rating differences versus the scale points was made, it was found to exhibit a curvilinear relationship somewhat different than originally hypothesized. The plotted data exhibited the following shape for each of the three rating scales.



The data thus revealed fairly close agreement between raters at the lowest scale point. The amount of variation between raters tended to increase rapidly between scale points 1 and 3 and then tended to stabilize at the higher scale points. Thus, the close agreement that was expected at the upper end of the scale did not materialize.

The relationship between the mean rating differences and the rating scale points was then estimated statistically using least-squares techniques to fit a curve of the form:

$$Y = a - b/X$$

where X = the rating scale point
Y = the mean difference associated with the given scale point.

The resultant regression equations provided a good fit to the data. All had regression coefficients that were significant at the .001 level.

The procedures described above may be summarized in mathematical notation for each observational area as follows:

(1) Let: d_{i1j} = absolute difference between Raters 1 and 2 of each rater pair at the i^{th} observation point, assuming Rater 1 is correct in his assignment of rating j

d_{i2j} = absolute difference between Raters 1 and 2 of each rater pair at the i^{th} observation point, assuming Rater 2 is correct in his assignment of rating j

(2) Then, the mean difference for the j^{th} rating point on the scale, assuming Rater 1 of each rater pair is correct, (m_{1j}) is:

$$m_{1j} = \frac{\sum d_{i1j}}{n_{1j}}$$

and, the mean discrepancy for the j^{th} rating point on the scale, assuming Rater 2 of each rater pair is correct, (m_{2j}) is:

$$m_{2j} = \frac{\sum d_{i2j}}{n_{2j}}$$

(3) The mean discrepancy across all pairs of raters for the j^{th} rating point (MD_{tj}) is:

$$MD_{tj} = \frac{\sum d_{i1j} + \sum d_{i2j}}{n_{1j} + n_{2j}}$$

(4) And the variance of the mean discrepancy at the j^{th} point $\left(\sigma^2_{MD_{tj}}\right)$ is:

$$\sigma^2_{MD_{tj}} = \frac{n_{1j}\ SD^2_{1j} + n_{2j}\ SD^2_{2j}}{n_{1j} + n_{2j} - 2}$$

(5) And the 95 percent confidence interval for the mean discrepancy at the j^{th} point is:

$$MD_{tj} \pm 1.96 \frac{\sigma_{MD_{tj}}}{\sqrt{n_{1j} + n_{2j}}}$$

Consistency Among Raters—Counts

To assess the reliability of measurements that required the observers to count the number of similar type items that were present (e.g., bulk items, abandoned vehicles), the following procedure was developed: we would determine for each count value the percent of instances in which there was complete agreement between pairs of raters, the percent of instances in which they disagreed by 2 counts, and so forth.

The technique proposed to handle this issue was similar to that proposed to handle the question of consistency when rating scales were used. The only difference was that instead of looking at the mean difference for each count value, we would be looking at percent of observations that differ by 0, 1, 2, etc. at each count value. Percentages were to be used rather than mean differences because, whereas it makes sense to speak of a garbage rating of 3.5, it makes no sense to speak of 3.5 bulk items.

However, when the data on measurements that required counts were reviewed, it became apparent that the agreement was so close that further analysis of the reliability of these measurements was unnecessary. That is, raters were found to be in agreement 97 percent of the time or better when it came to assessing the number of similar type items that were present.

Consistency Among Raters—Yes-No Type Measurements

To determine the reliability of measurements that required an assessment of the presence or absence of a given condition (e.g., rat indicators, odors), chi-square tests were performed. In all cases the results indicated a close degree association between the responses of paired observers at the .001 significance level. The percent of agreement was also quite high—96 percent of the time or better raters were in complete agreement.

ESTIMATING VARIANCE COMPONENTS FOR A CENSUS TRACT

In making statistical inferences about a geographical area (e.g., the mean garbage rating for a census block, census tract, sanitation district, etc.), it must be recognized that there are a number of sources of variability that can affect the estimate, rater inconsistency being only one of these. These other sources of variability include: variation among points in a block, variation among blocks in a tract, variation among tracts within a stratum, etc.

To the extent that these other sources of variability are large relative to the variation caused by rater inconsistency, the disagreement among raters takes on less significance. For this reason, the analysis phase of the project included an attempt to estimate the relative contribution of the following variance components to the total variance of a tract mean:

- Variation among blocks in the tract
- Variation among observers
- Random sources of variation.

The data collected for the common tract, where all observers went each day, were used to perform the analysis.

The technique employed to estimate the above variance components may be summarized by the statements provided below.

(1) A model for the observation taken by the j^{th} observer in the i^{th} block was postulated:

$$y_{ij} = u + b_i + r_j + e_{ij}$$

where y_{ij} = average block value recorded by the j^{th} rater for the i^{th} block

u = true mean score for the tract

b_i = block effect

r_j = rater effect

e_{ij} = random effect

(2) The distribution was assumed to be such that the expected values and the variances for b_i, r_j, and e_{ij} would be as follows:

$$0, \sigma_b^2 \quad \text{for the } b_i$$

$0, \sigma_r^2$ for the r_j

$0, \sigma_e^2$ for the e_{ij}

(3) Thus, an estimate of the mean score for the tract would be provided by:

$$y_{\cdots} = \frac{1}{rb} + \sum_i \sum_j y_{ij}$$

where $y_{\cdots}$ = average tract value

y_{ij} = average block value recorded by the j^{th} observer for the i^{th} block

r = number of raters

b = number of blocks

(4) And the variance of $y_{\cdots}$ would be:

$$\sigma^2_{y\cdots} = \frac{\sigma_b^2}{b} + \frac{\sigma_r^2}{r} + \frac{\sigma_e^2}{rb}$$

(5) By performing an analysis of variance, the expected value of the mean squares (EMS) associated with the sources of variation indicated in (4) was obtained. The relationship between the EMS and the above variance components is:

- For blocks: $EMS_b = \hat{\sigma}_e^2 + r\,\hat{\sigma}_b^2$
- For raters: $EMS_r = \hat{\sigma}_e^2 + b\,\hat{\sigma}_r^2$
- For the error term: $EMS_e = \hat{\sigma}_e^2$

(6) The equations listed in (5) were solved in a simultaneous fashion and estimates of σ_b^2, σ_r^2, and σ_e^2 were developed.

(7) The values of the variances obtained in (6) were then substituted into (4) to estimate the total variance of y... and its components: $\frac{\sigma_b^2}{b}$, $\frac{\sigma_r^2}{r}$, and $\frac{\sigma_e^2}{rb}$.

ASSESSING THE ACCURACY OF OBSERVERS

Accuracy refers to the degree to which the measured value approaches the true value. In the case of the rating scales, the true value of conditions at a given site would be the average value across an infinite population of observers looking at the same site. As an approximation of the true value, the mean rating across the 10 observers can be used.

Thus, in order to get an idea of how accurate each observer was, we used the data from the common tract where an average value across 10 observers could be made. Specifically, we compared the tract rating for a given observer with the overall tract rating that resulted when an average value was estimated across all observers. The 95 percent confidence interval about each observer's mean value was also computed. In addition, we compared the average block ratings for each observer with the overall block ratings.

DETERMINING THE CORRELATIONS

Correlation analysis was used to test the degree of association between:

- Similar variables measured in each of the six observational areas for which data were collected.
- Different variables measured within the same observational area.

This was done in order to determine whether it was possible to reduce the number of observational areas for which measurements should be made and/or the number of different variables.

In both instances, the correlation matrices were developed on the basis of the Pearson product-moment correlation coefficient. This is a measure of association that can be used when the variables represent at least an ordered continuum of some kind (low to high, agree to disagree, etc.). It was selected because it is more powerful than other tests of association; that is, the probability of making a correct inference based on the value of the correlation coefficient is higher for this type of test than for others.

Correlations Among Variables Across Observational Areas

In measuring the correlations among the six observational areas—blockfaces, private ways, alleys, storage areas, backyards, and lots—the average garbage, glass, and refuse ratings were computed for each of the 100 blocks that were surveyed. From these averages, simple correlation coefficients were developed to reflect the strength of the relationship between the garbage ratings in blockfaces and the garbage ratings in each of the other areas. Simple correlation coefficients were also developed to assess the strength of the relationship between glass and refuse ratings in blockfaces and those in other observational areas. In a similar manner, correlations were developed between ratings in alleys and those in the other observational areas, and so forth.

Correlations Among Variables in the Same Observational Area

Simple correlation coefficients among the variables were developed for blockface and alley conditions. The analysis was performed only for these two observational areas because when the results of the correlation analysis across observational areas were reviewed, conditions in these two areas were found to be related at statistically significant levels to conditions in the other observational areas included in the field test.

DEVELOPING THE COMPOSITE MEASURES

As described in Chapter IV, it was decided to utilize the following techniques to develop indices of solid waste effectiveness:

(1) Have the observers assign a composite (or global) rating to each area they surveyed at the same time they made measurements of the individual variables. In assigning the composite rating, the observers were to use an 8-point scale, with 1 representing the most favorable overall conditions and 8 representing the most unfavorable overall conditions.

(2) Use regression techniques to estimate the functional forms corresponding to the three types of decision models. In the regressions, the composite rating was to be used as the dependent variable; the other variables were to be used as the independent variables.

(3) Compare the resultant regressions to determine the most appropriate decision model and use this to formulate an index of solid waste effectiveness.

In implementing this approach, the following general forms were used to specify the three decision models.

Type of Model	Functional Form
Linear	$Y_1 = a_1 + \sum_{i=1}^{n} b_{i1} Z_i$
Conjunctive	$\ln Y_2 = a_2 + \sum_{i=1}^{n} b_{i2} \ln(Z_i)$
Disjunctive	$\ln Y_3 = a_3 + \sum_{i=1}^{n} b_{i3} \ln(\phi_i - Z_i)$

where Y = dependent variable
Z_i = independent variables
a = regression constant
b_i = regression coefficients
ϕ_i = asymtotic values associated with each independent variable for the disjunctive model.

In developing the estimating equations associated with the above functional forms, we let:

$Y = 9 - E$ (where E was the value of the composite rating, along a scale of 1-8)

$Z_i = 8 - X_i$ (where X_i was the value of the measured variables, each scaled so that the lowest value was 1 and the highest value was 7)

$\phi_i = 7.07$

That is, we used a reverse scaling of the variables.

The estimating equations for the three types of decision models that resulted from application of least-squares techniques thus became:

Type of Model	Estimating Equation
Linear	$\hat{Y}_1 = a_1 + \sum_{i=1}^{n} b_{i1} (8 - X_i)$
Conjunctive	$\ln\hat{Y}_2 = a_2 + \sum_{i=1}^{n} b_{i2} \ln (8 - X_i)$
Disjunctive	$\ln\hat{Y}_3 = a_3 - \sum_{i=1}^{n} b_{i3} \ln \left[7.07 - (8 - X_i)\right]$
	$= a_3 - \sum_{i=1}^{n} b_{i3} \ln (X_i - .93)$

In all of these formulations, the higher the value of $\hat{Y}_j$, the better the overall conditions. In the case of the conjunctive model, one can say that all of the measured variables (X_i) must be no worse than a given level. In the case of the disjunctive model, one can say at least one of the measured variables must be no worse than a given level.

Stepwise linear regression techniques were used in estimating the equations. Variables significant at the .05 level or greater were retained. The equations were estimated both for selected observers and for the group as a whole. The latter estimations were developed by averaging the values that individual observers assigned to the areas they inspected in the common tract.

The resultant equations were then transformed into indices that could take on values between 0 and 1. This was done by first finding the maximum value for $\hat{Y}_j$; i.e., the value that would be obtained when the X_i values are at their minimum. This was assigned the value of $Y_{j(max)}$. Next, the value for $Y_{j(min)}$ was found; i.e., the value that would be obtained when the X_i values are at their maximum. We then set $Y_{j(min)}$ = zero and $Y_{j(max)}$ = 1. The intermediate scale values were determined by application of the following formula:

$$\left(\hat{Y}_j - Y_{j(min)}\right) \Big/ \left(Y_{j(max)} - Y_{j(min)}\right)$$

where $\hat{Y}_j$ = value predicted by the regression equation (j = 1, 2, 3).

APPENDIX F

TABLES, CHARTS, AND GRAPHS THAT SUPPORT FINDINGS

This appendix contains a number of detailed charts, tables, and graphs that support the findings presented in Chapter VI. The tabular displays are presented first, followed by figures that contain the charts and graphs.

Table F-1. FREQUENCY DISTRIBUTION FOR BULK ITEMS BY AREA OF OBSERVATION—ALL TRACTS IN SAMPLE

Area of Observation	% Where None Observed	% Where 1 Observed	% Where 2-5 Observed	% Where More Than 5 Observed	Total Observations
Blockfaces	99.6	0.3	0.1	0.0	747
Private Ways	95.7	2.3	1.6	0.4	735
Alleys	83.0	9.3	6.1	1.6	313
Storage Areas	92.3	3.2	4.0	0.5	572
Backyards	80.6	8.4	7.2	3.8	692
Lots	69.0	6.9	13.8	10.3	58

Table F-2. FREQUENCY DISTRIBUTION FOR BULK ITEMS BY AREA OF OBSERVATION—COMMON TRACT ONLY

Area of Observation	% Where None Observed	% Where 1 Observed	% Where 2-5 Observed	% Where More Than 5 Observed	Total Observations
Blockfaces	99.4	0.6	0.0	0.0	348
Private Ways	93.6	4.7	1.7	0.0	343
Alleys	72.5	12.1	15.4	0.0	182

Table F-3. FREQUENCY DISTRIBUTION FOR DEAD ANIMALS BY AREA OF OBSERVATION—ALL TRACTS IN SAMPLE

Area of Observation	% Where None Observed	% Where 1 Observed	% Where 2-5 Observed	% Where More Than 5 Observed	Total Observations
Blockfaces	99.3	0.7	0.0	0.0	747
Private Ways	99.3	0.7	0.0	0.0	735
Alleys	97.8	1.6	0.6	0.0	313
Storage Areas	99.8	0.2	0.0	0.0	574
Backyards	99.7	0.3	0.0	0.0	694
Lots	98.3	1.7	0.0	0.0	58

Table F-4. FREQUENCY DISTRIBUTION FOR DEAD ANIMALS BY AREA OF OBSERVATION—COMMON TRACT ONLY

Area of Observation	% Where None Observed	% Where 1 Observed	% Where 2-5 Observed	% Where More Than 5 Observed	Total Observations
Blockfaces	99.1	0.9	0.0	0.0	348
Private Ways	98.5	1.5	0.0	0.0	343
Alleys	92.9	5.5	1.6	0.0	183

Table F-5. FREQUENCY DISTRIBUTION FOR ABANDONED VEHICLES BY AREA OF OBSERVATION—ALL TRACTS IN SAMPLE

Area of Observation	% Where None Observed	% Where 1 Observed	% Where 2-5 Observed	% Where More Than 5 Observed	Total Observations
Blockfaces	100.0	0.0	0.0	0.0	747
Private Ways	100.0	0.0	0.0	0.0	735
Alleys	97.7	1.0	1.3	0.0	313
Storage Areas	NA	NA	NA	NA	NA
Backyards	98.3	1.7	0.0	0.0	694
Lots	86.2	10.3	3.5	0.0	58

Table F-6. FREQUENCY DISTRIBUTION FOR ABANDONED VEHICLES BY AREA OF OBSERVATION—COMMON TRACT ONLY

Area of Observation	% Where None Observed	% Where 1 Observed	% Where 2-5 Observed	% Where More Than 5 Observed	Total Observations
Blockfaces	100.0	0.0	0.0	0.0	348
Private Ways	100.0	0.0	0.0	0.0	343
Alleys	98.9	1.1	0.0	0.0	183

Table F-7. FREQUENCY DISTRIBUTION FOR CLOGGED DRAIN BASINS BY AREA OF OBSERVATION—ALL TRACTS IN SAMPLE

Area of Observation	% Where None Observed	% Where 1 Observed	% Where 2-5 Observed	% Where More Than 5 Observed	Total Observations
Blockfaces	93.3	6.6	0.1	0.0	746
Alleys	98.7	1.3	0.0	0.0	313

Table F-8. FREQUENCY DISTRIBUTION FOR CLOGGED DRAIN BASINS BY AREA OF OBSERVATION—COMMON TRACT ONLY

Area of Observation	% Where None Observed	% Where 1 Observed	% Where 2-5 Observed	% Where More Than 5 Observed	Total Observations
Blockfaces	90.2	9.8	0.0	0.0	348
Alleys	100.0	0.0	0.0	0.0	183

Table F-9. FREQUENCY DISTRIBUTION FOR FIRE HAZARDS BY AREA OF OBSERVATION—ALL TRACTS IN SAMPLE

Area of Observation	% Where No Fire Hazard Observed	% Where Minor Fire Hazard Observed	% Where Major Fire Hazard Observed	Total Observations
Blockfaces	100.0	0.0	0.0	747
Private Ways	99.6	0.3	0.1	735
Alleys	94.6	4.5	1.0	313
Storage Areas	98.3	1.0	0.7	574
Backyards	94.4	4.0	1.6	694
Lots	93.1	6.9	0.0	58

Table F-10. FREQUENCY DISTRIBUTION FOR FIRE HAZARDS BY AREA OF OBSERVATION—COMMON TRACT ONLY

Area of Observation	% Where No Fire Hazard Observed	% Where Minor Fire Hazard Observed	% Where Major Fire Hazard Observed	Total Observations
Blockfaces	100.0	0.0	0.0	348
Private Ways	100.0	0.0	0.0	343
Alleys	88.6	10.9	0.5	183

Table F-11. FREQUENCY DISTRIBUTION FOR RAT INDICATORS BY AREA OF OBSERVATION—ALL TRACTS IN SAMPLE

Area of Observation	% Where No Rat Indicators Present	% Where Rat Indicators Present	Total Observations
Blockfaces	98.9	1.1	747
Private Ways	93.2	6.8	735
Alleys	67.4	32.6	313
Storage Areas	91.3	8.7	574
Backyards	81.4	18.6	694
Lots	72.4	27.6	58

Table F-12. FREQUENCY DISTRIBUTION FOR RAT INDICATORS BY AREA OF OBSERVATION—COMMON TRACT ONLY

Area of Observation	% Where No Rat Indicators Present	% Where Rat Indicators Present	Total Observations
Blockface	97.7	2.3	348
Private Ways	80.5	19.5	343
Alleys	29.5	70.5	183

Table F-13. FREQUENCY DISTRIBUTION FOR INSECT INDICATORS BY AREA OF OBSERVATION—ALL TRACTS IN SAMPLE

Area of Observation	% Where No Insect Indicators Present	% Where Insect Indicators Present	Total Observations
Storage Areas	99.8	0.2	574

Table F-14. FREQUENCY DISTRIBUTION FOR ODORS BY AREA OF OBSERVATION—ALL TRACTS IN SAMPLE

Area of Observation	% Where No Odors Present	% Where Odors Present	Total Observations
Storage Areas	96.0	4.0	574

Table F-15. GARBAGE RATING SUMMARY STATISTICS FOR CONSISTENCY AMONG RATERS

Scale Point	Mean Rating Difference	Variance σ^2	Standard Deviation σ	95% Confidence Interval About the Mean Difference	Total Number of Comparisons
1	.1360	.2676	.5173	±.0143	5052
2	.7022	.5025	.7089	±.0766	329
3	1.1020	1.0298	1.0148	±.0813	598
4	1.4000	.8000	.8944	±.3333	30
5	1.5780	1.8403	1.3566	±.2142	154
6	1.1819	1.1637	1.0788	±.7151	11
7	1.5909	2.9840	1.7274	±.4254	66

Table F-16. GLASS RATING SUMMARY STATISTICS FOR CONSISTENCY AMONG RATERS

Scale Point	Mean Rating Difference	Variance σ^2	Standard Deviation σ	95% Confidence Interval About the Mean Difference	Total Number of Comparisons
1	.2599	.5439	.7375	±.0247	3410
2	.7871	.6702	.8187	±.0733	749
3	1.0956	1.4112	1.1879	±.0664	1224
4	1.1899	.7791	.8827	±.1376	158
5	1.1975	1.3714	1.1711	±.1047	481
6	1.0308	.9679	.9838	±.2442	65
7	1.5108	3.1741	1.7816	±.1705	419

Table F-17. REFUSE RATING SUMMARY STATISTICS FOR CONSISTENCY AMONG RATERS

Scale Point	Mean Rating Difference	Variance σ^2	Standard Deviation σ	95% Confidence Interval About the Mean Difference	Total Number of Comparisons
1	.4163	.7957	.8920	±.0376	2162
2	.8705	.9282	.9634	±.0710	710
3	.8694	1.1119	1.0545	±.0504	1684
4	1.1173	.8319	.9121	±.1082	273
5	1.3270	1.2424	1.1146	±.0808	731
6	1.1409	1.1228	1.0596	±.2489	71
7	1.2463	2.5734	1.6042	±.1278	605

Table F-18. BLOCKFACE ESTIMATING EQUATIONS FOR LINEAR, CONJUNCTIVE, AND DISJUNCTIVE MODELS FOR OBSERVER 1

Type of Model	Estimating Equations[a] (all coefficients significant at .001 level)	R^2
Linear	$\hat{E} = -.48 + .51 X_1 + .45 X_2 + .25 X_3$	.911
Conjunctive	$\ln \hat{E} = -.14 + .56 \ln(X_1) + .44 \ln(X_2) + .19 \ln(X_3)$	.845
Disjunctive	$\ln \hat{E} = 2.93 - .19 \ln(7.07 - X_1) - .24 \ln(7.07 - X_2) - .84 \ln(7.07 - X_3)$	.594

[a] $\hat{E}$ = blockface composite rating; X_1 = blockface refuse rating; X_2 = blockface glass rating; X_3 = blockface garbage rating.

Table F-19. BLOCKFACE ESTIMATING EQUATIONS FOR LINEAR, CONJUNCTIVE, AND DISJUNCTIVE MODELS FOR OBSERVER 2

Type of Model	Estimating Equations[a] (all coefficients significant at .001 level)	R^2
Linear	$\hat{E} = .72 + .45 X_1 + .25 X_2$	.733
Conjunctive	$\ln \hat{E} = .27 + .38 \ln(X_1) + .34 \ln(X_2)$	.706
Disjunctive	$\ln \hat{E} = 2.17 - .58 \ln(7.07 - X_1) - .30 \ln(7.07 - X_2)$	.631

[a] $\hat{E}$ = blockface composite rating; X_1 = blockface refuse rating; X_2 = blockface glass rating.

Table F-20. BLOCKFACE ESTIMATING EQUATIONS FOR LINEAR, CONJUNCTIVE, AND DISJUNCTIVE MODELS FOR OBSERVER 6

Type of Model	Estimating Equations[a] (all coefficients significant at .01 level)	R^2
Linear	$\hat{E} = 1.57 + .28X_1 + .26X_2$	.508
Conjunctive	$\ln \hat{E} = .59 + .23 \ln (X_1) + .29 \ln (X_2)$	.541
Disjunctive	$\ln \hat{E} = 1.31 - .19 \ln (7.07-X_2)$	.214

[a] $\hat{E}$ = blockface composite rating; X_1 = blockface refuse rating; X_2 = blockface glass rating.

Table F-21. BLOCKFACE ESTIMATING EQUATIONS FOR LINEAR, CONJUNCTIVE, AND DISJUNCTIVE MODELS FOR OBSERVER 7

Type of Model	Estimating Equations[a] (all coefficients significant at .01 level)	R^2
Linear	$\hat{E} = 1.67 + .19X_1 + .25X_2 + .52X_3$	.595
Conjunctive	$\ln \hat{E} = .79 + .23 \ln (X_1) + .25 \ln (X_2)$	.529
Disjunctive	$\ln \hat{E} = 3.85 - .10 \ln (7.07-X_2) - 1.39 \ln (7.07-X_3)$	.449

[a] $\hat{E}$ = blockface composite rating; X_1 = blockface refuse rating; X_2 = blockface glass rating; X_3 = blockface garbage rating.

Table F-22. BLOCKFACE ESTIMATING EQUATIONS FOR LINEAR, CONJUNCTIVE, AND DISJUNCTIVE MODELS FOR OBSERVER 10

Type of Model	Estimating Equations[a] (all coefficients significant at .001 level)	R^2
Linear	$\hat{E} = .31 + .39\,X_1 + .22\,X_2$	.692
Conjunctive	$\ln\hat{E} = -.02 + .42\ln(X_1) + .28\ln(X_2)$	.654
Disjunctive	$\ln\hat{E} = 2.13 - .32\ln(7.07-X_1) - .77\ln(7.07-X_2)$	.539

[a] $\hat{E}$ = blockface composite rating; X_1 = blockface refuse rating; X_2 = blockface glass rating.

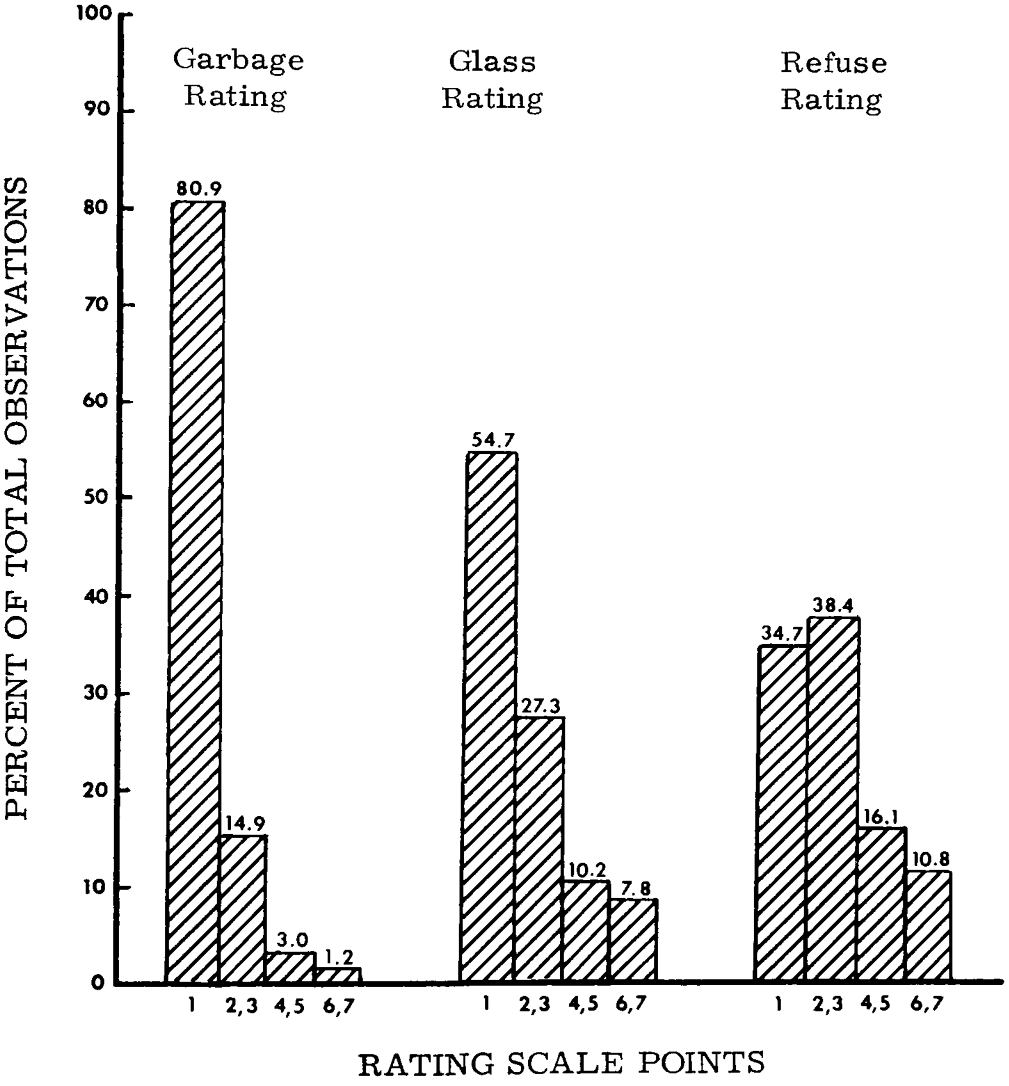


Figure F-1: Frequency Distribution for Garbage, Glass, and Refuse Rating Scales — All Tracts in Sample

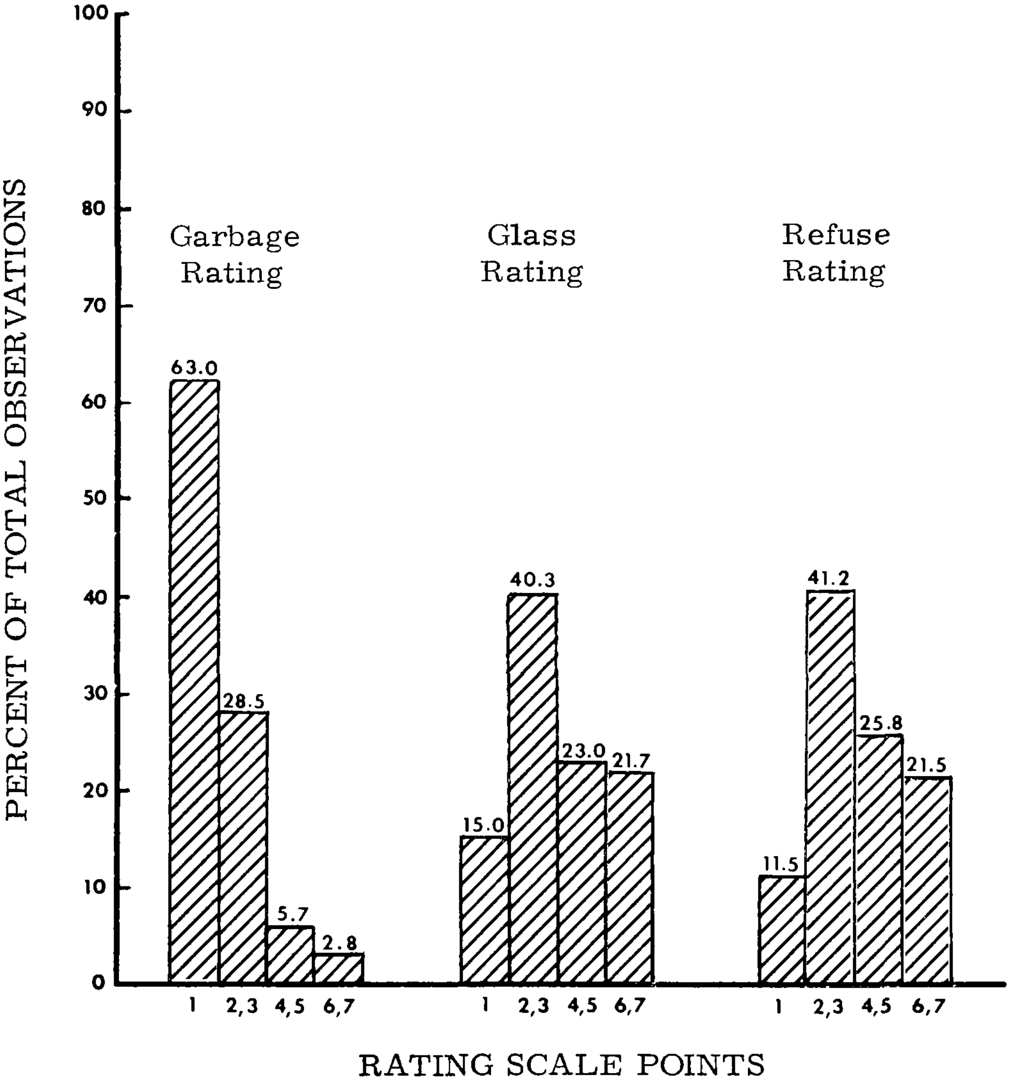


Figure F-2: Frequency Distribution for Garbage, Glass, and Refuse Rating Scales — Common Tract Only

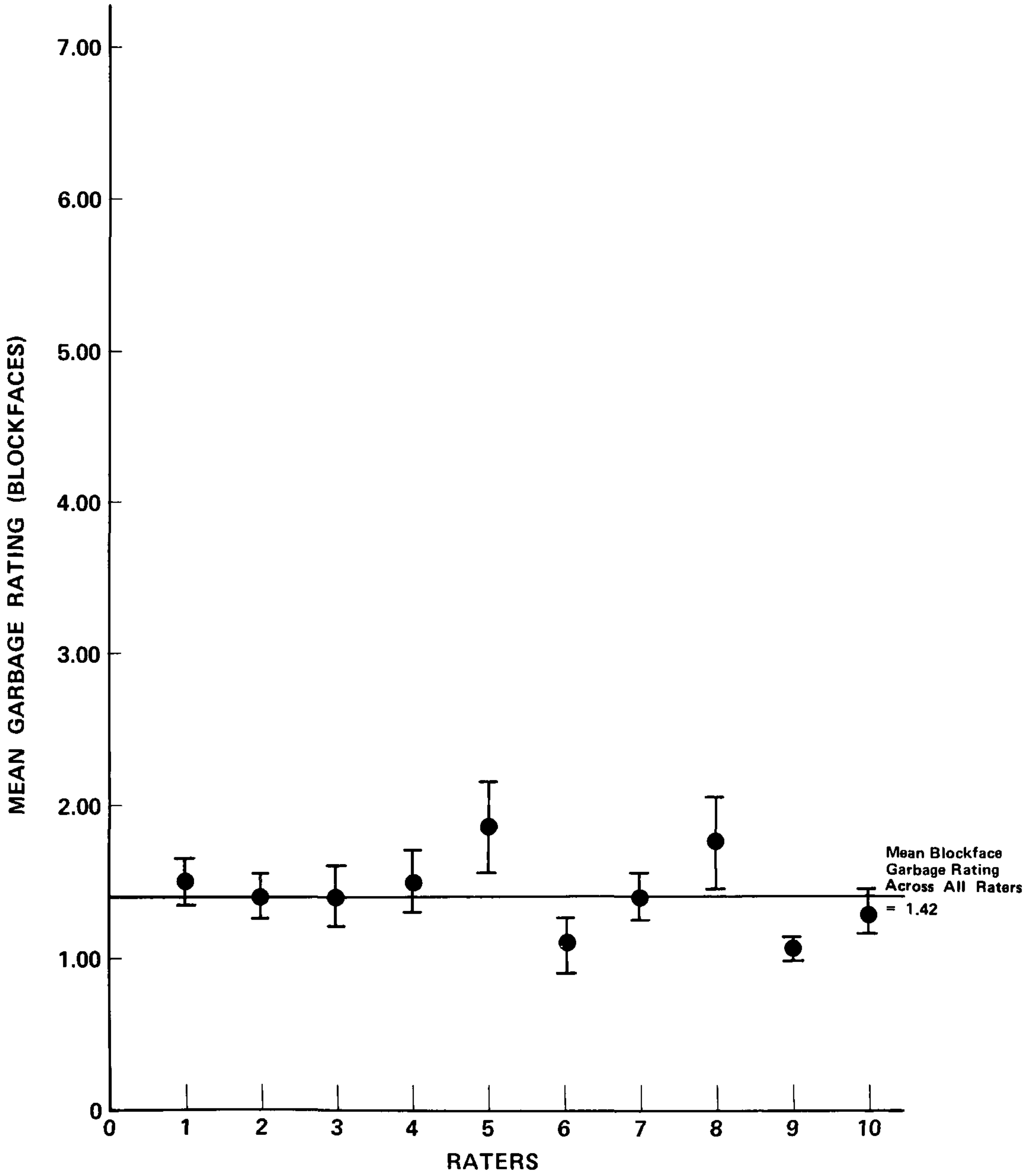


Figure F-3: Average Garbage Rating for Blockfaces in the Common Tract: Overall and by Rater (including 95% confidence interval about the rater means)

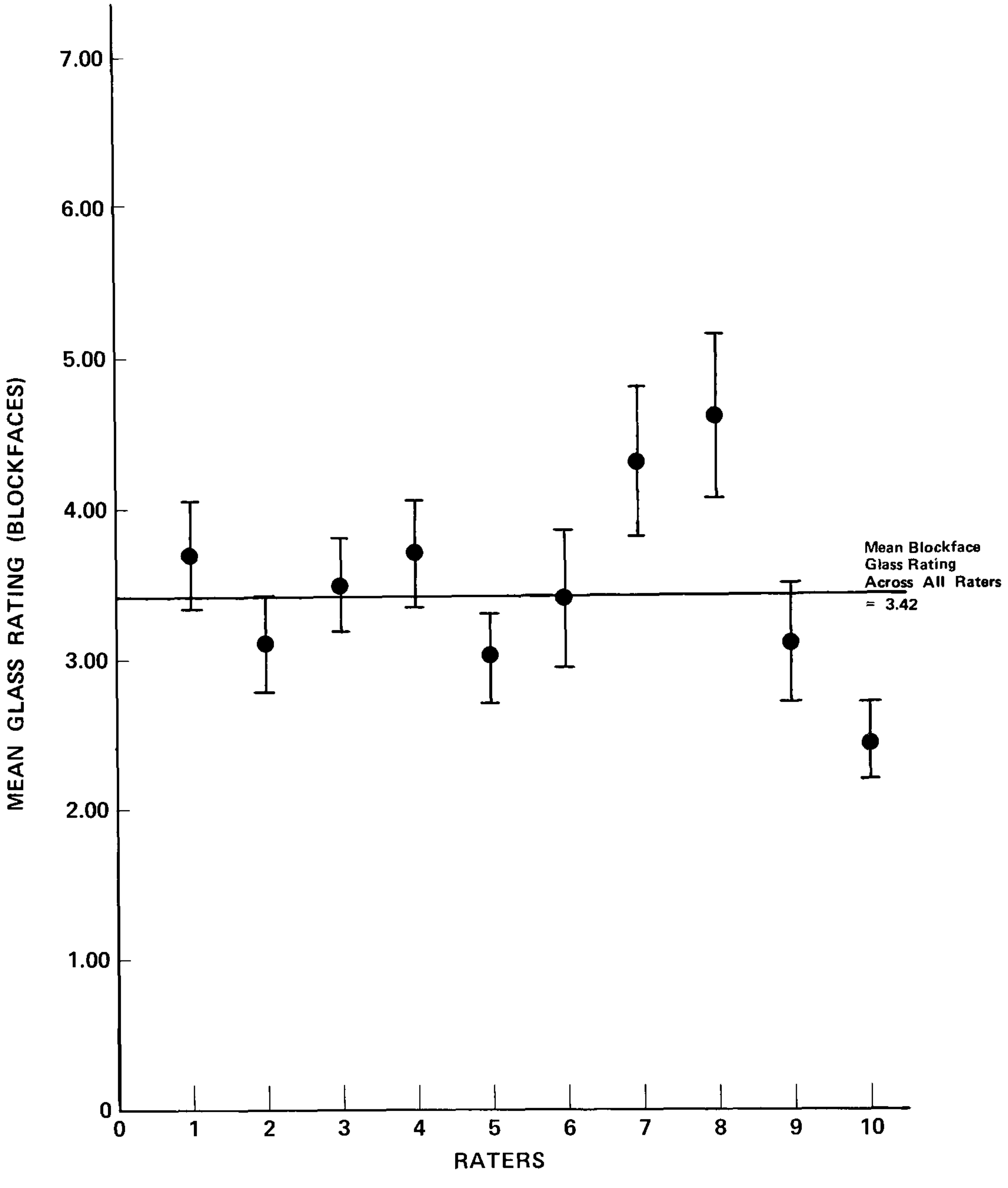


Figure F-4: Average Glass Rating for Blockfaces in the Common Tract: Overall and by Rater (including 95% confidence interval about the rater means)

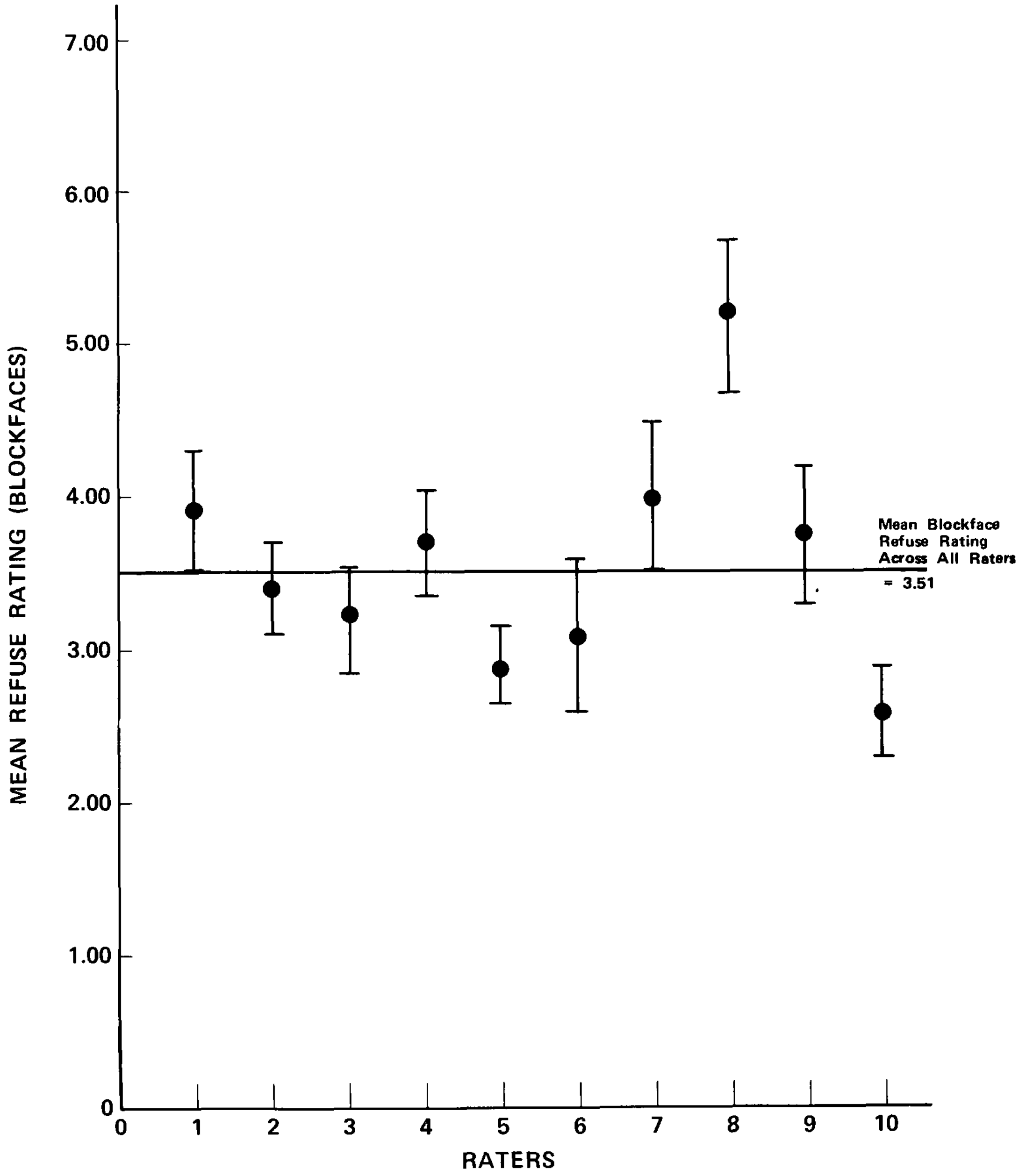


Figure F-5: Average Refuse Rating for Blockfaces in the Common Tract: Overall and by Rater (including 95% confidence interval about the rater means)

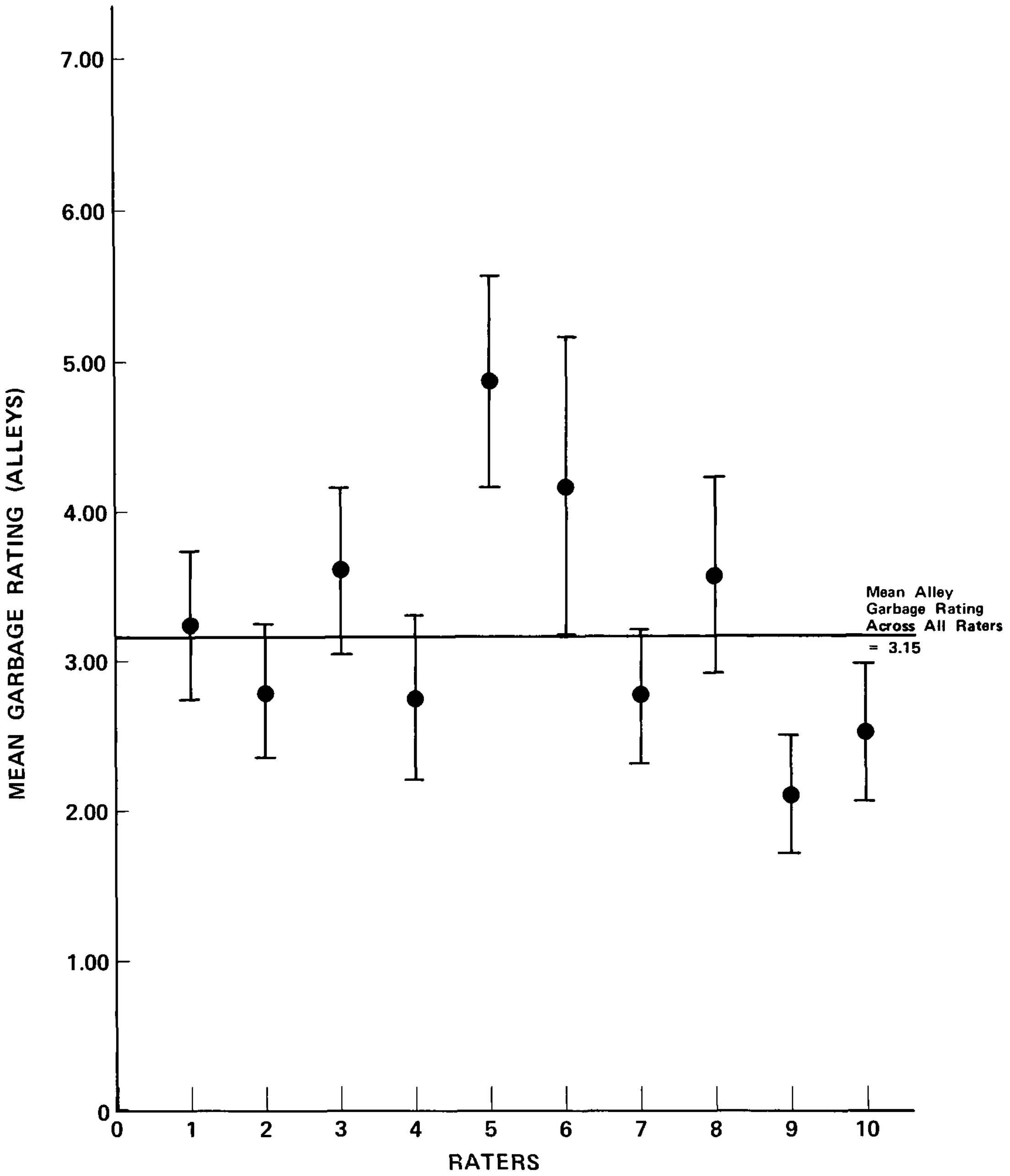


Figure F-6: Average Garbage Rating for Alleys in the Common Tract: Overall and by Rater (including 95% confidence interval about the rater means)

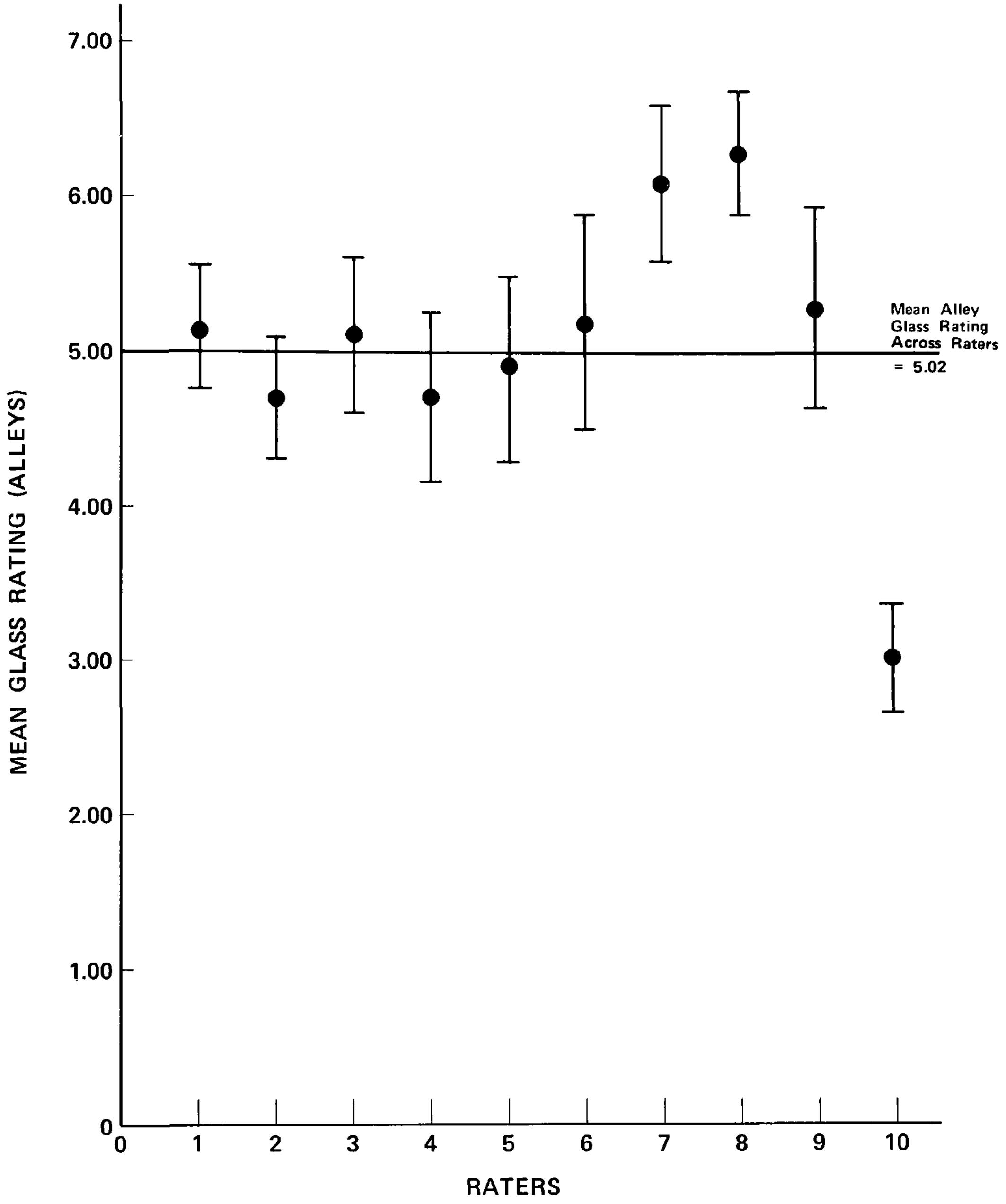


Figure F-7: Average Glass Rating for Alleys in the Common Tract: Overall and by Rater (including 95% confidence interval about the rater means)

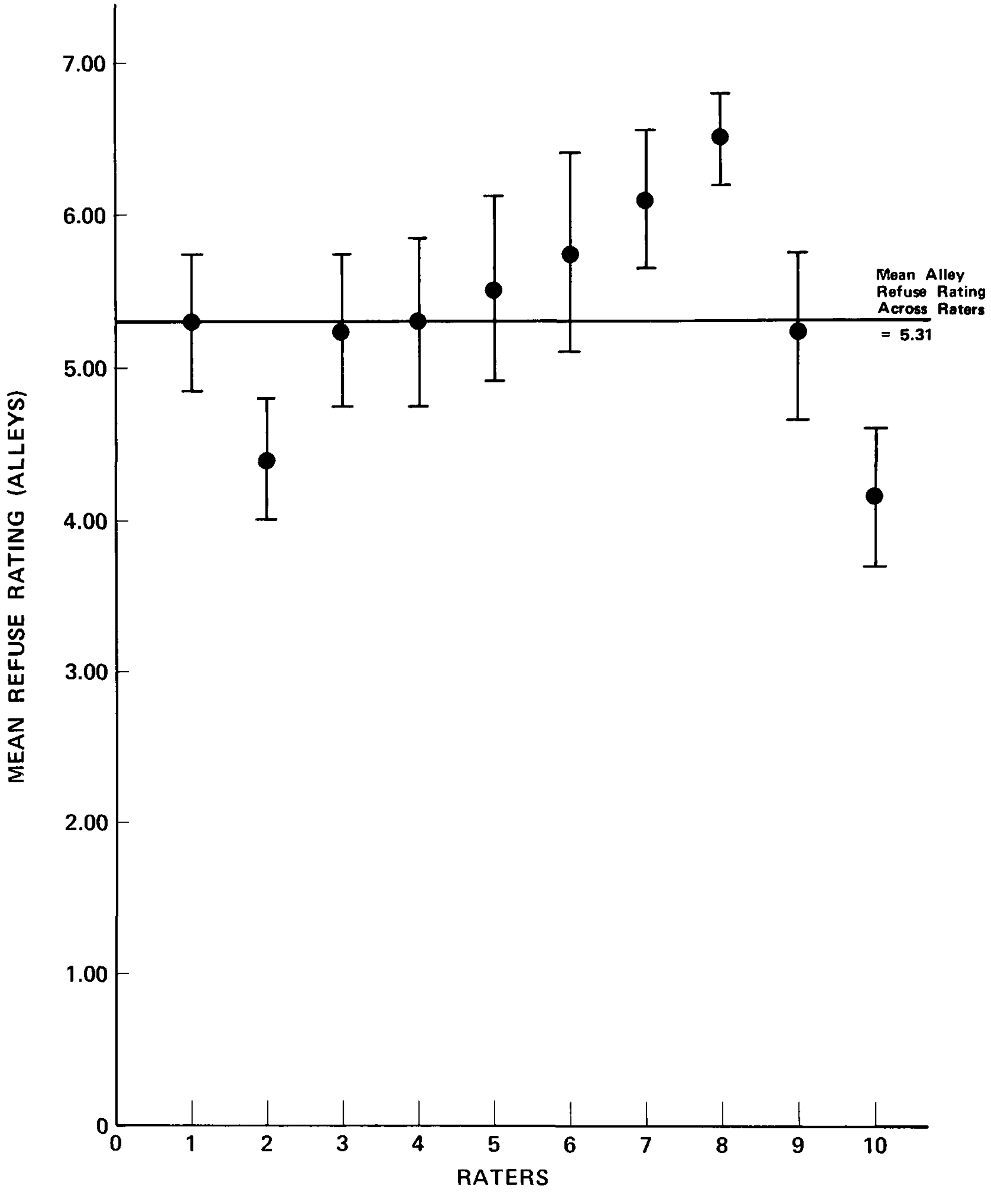


Figure F-8: Average Refuse Rating for Alleys in the Common Tract: Overall and by Rater (including 95% confidence interval about the rater means)

TECHNICAL REPORT DATA

(Please read Instructions on the reverse before completing)

1. REPORT NO.
EPA-670/2-74-082

2.

3. RECIPIENT'S ACCESSION NO.

4. TITLE AND SUBTITLE
MEASURES OF EFFECTIVENESS FOR REFUSE STORAGE, COLLECTION, AND TRANSPORTATION PRACTICES

5. REPORT DATE
November 1974; Issuing Date

6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)
Messer Associates, Inc.

8. PERFORMING ORGANIZATION REPORT NO.

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Messer Associates, Inc.
8555 16th Street, Suite 706
Silver Spring, Maryland 20910

10. PROGRAM ELEMENT NO.
1DB063/ROAP 02AAE/Task 06

11. CONTRACT/~~GRANT~~ NO.
68-03-0260

12. SPONSORING AGENCY NAME AND ADDRESS
National Environmental Research Center
Office of Research and Development
U.S. Environmental Protection Agency
Cincinnati, Ohio 45268

13. TYPE OF REPORT AND PERIOD COVERED
Final

14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT Perhaps between 75 to 80 percent of a solid waste system cost is due to storage, collection, and transportation, the remainder being attributable to disposal. Given an adequate accounting system, the monetary costs of a solid waste management system are much easier to compute than are the benefits produced and the nonmonetary cost incurred. Thus, although a community may have an accurate estimate of what it is spending on its system, it often is uncertain as to whether or not it is receiving reasonable value in the benefits returned; i.e., it has little or no idea of its "cost effectiveness." This report presents the results of a project that focused on the systematic development of a set of measures and measurement tools that could be used to assess the effectiveness of solid waste storage, collection, and transportation practices. The project included a pilot test of the measurement methodology in an urban community. The measurement system presented in this report is intended to support municipal decision-makers who have responsibility for such services as mixed refuse collection, street and alley cleaning, sanitary code enforcement, sanitation education, and other related activities. It provides a model or prototype that municipal representatives can use to design effectiveness measures that are specific to their own solid waste management needs and activities. The report includes a comprehensive list of candidate effectiveness measures along with the measurement techniques and sampling procedures needed to collect data to formulate the candidate measures.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
*Measurement Refuse Storage Collection Transportation	*Solid waste management *Effectiveness measurements	13B

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS *(This Report)*	21. NO. OF PAGES
RELEASE TO PUBLIC	UNCLASSIFIED	201
	20. SECURITY CLASS *(This page)*	22. PRICE
	UNCLASSIFIED	

EPA Form 2220-1 (9-73)

☆ U.S. GOVERNMENT PRINTING OFFICE: 1974-657-586/5311 Region No. 5-11