TRANSPORTATION RESEARCH RECORD 746

Bus Transit Management and Performance

TRANSPORTATION RESEARCH BOARD

COMMISSION ON SOCIOTECHNICAL SYSTEMS
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1980

Transportation Research Record 746
Price $5.20
Edited for TRB by Susan Singer-Bart

mode
2 public transit

subject areas
11 administration
12 planning
13 forecasting
14 finance
16 user needs
52 human factors
53 vehicle characteristics
54 operations and traffic control
55 traffic flow, capacity, and measurements

Library of Congress Cataloging in Publication Data
National Research Council. Transportation Research Board.
Bus transit management and performance.

(Transportation research record; 746)
Thirteen reports prepared for the 59th annual meeting of the Transportation Research Board.
Includes bibliographical references.
1. Motor bus lines—United States—Addresses, essays, lectures. 2. Local transit—United States—Addresses, essays, lectures. I. Title. II. Series.
TE7.H5 no. 746 [HE5623] 380.5s [388.4'1322]
ISBN 0-309-03055-2 ISSN 0361-1981 80-19208

Sponsorship of the Papers in This Transportation Research Record

GROUP 1–TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
Leon M. Cole, Library of Congress, chairman

Public Transportation Section
George Edward Gray, California Department of Transportation, chairman

Committee on Bus Transit Systems
Herbert S. Levinson, Wilbur Smith and Associates, chairman
James C. Echols, Tidewater Transportation District Commission, vice chairman
Robert B. Deuser, Beiswenger Hoch and Associates, Inc., secretary
Richard J. Anderson, J. J. Bakker, Alvin G. Basmadjian, John W. Bates, Eugene T. Canty, Karen P. Damsgaard, Frank Dicesare, Curt M. Elmberg, Gordon J. Fielding, Edward R. Fleischman, Marvin C. Gersten, Jack R. Gilstrap, Andrew Hollander, Ronald I. Hollis, Linda Sue Hunkins, Carol A. Keck, Henry M. Mayer, Neil Craig Miller, Donald A. Morin, Martin C. Nizlek, Vasant H. Surti, F. V. Webster

GROUP 3–OPERATION AND MAINTENANCE OF TRANSPORTATION FACILITIES
Adolf D. May, University of California, Berkeley, chairman

Committee on Transit Service Characteristics
James E. Reading, Ohio Transit Authority, chairman
Peter Wood, Mitre Corporation, secretary
Bert Arrillaga, Arline L. Bronzaft, John T. Doolittle, Jr., Peter B. Everett, Gordon J. Fielding, Martin Flusberg, Harold H. Geissenheimer, Benita H. Gray, Judith Kaplan, Karla H. Karash, Faye L. Mench, David R. Miller, James H. Miller, Ray A. Mundy, Ronald C. Pfefer, Philip J. Ringo, Roger P. Snoble

W. Campbell Graeub, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each report. The organizational units and officers and members are as of December 31, 1979.

Contents

Authors of the Papers in This Record

Allen, William G., Jr., Barton-Aschman Associates, Inc., 1730 K Street NW, Suite 720, Washington, DC 20006
Blume, Steven W., Mass Transit Administration, 109 E. Redwood Street, Baltimore, MD 21202
Bodin, Lawrence D., College of Business and Management, University of Maryland, College Park, MD 27042
Boyle, Daniel K., Planning and Research Bureau, New York State Department of Transportation, 1220 Washington Avenue, State Campus, Albany, NY 12232
Brög, Werner, Socialdata GmbH, Institute for Empirical Social Research, Hans-Grässel-Weg 1, D-8000 Munich 70, West Germany
Casey, Robert, Transportation Systems Center, U.S. Department of Transportation, Kendall Square, Cambridge, MA 02142
Cherwony, Walter, Transportation Consulting Division, Booz, Allen, and Hamilton, Inc., 1346 Chestnut Street, Philadelphia, PA 19107
Dial, Robert B., Urban Mass Transportation Administration, Washington, DC 20590
Falcocchio, John C., Transportation Training and Research Center, Polytechnic Institute of New York, 333 Jay Street, Brooklyn, NY 11201
Fandialan, Angel, Bureau of Urban and Public Transportation, Michigan Department of Transportation, 425 West Ottawa Street, Lansing, MI 48909
Förg, Otto G., Socialdata GmbH, Institute for Empirical Social Research, Hans-Grässel-Weg 1, D-8000 Munich 70, West Germany
Golenberg, Marvin, SG Associates, Inc., 4200 Daniels Avenue, Annandale, VA 22003
Grimm, Lewis G., Barton-Aschman Associates, Inc., 1730 K Street NW, Suite 720, Washington, DC 20006
Holec, James M., Jr., Peat, Marwick, Mitchell and Company, 1990 K Street NW, Washington, DC 20006
McShane, William R., Transportation Training and Research Center, Polytechnic Institute of New York, 333 Jay Street, Brooklyn, NY 11201
Menaker, Paul, P. R. C. Harris Consultants, 15 Bank Street, Stamford, CT 06901; formerly with the Polytechnic Institute of New York
Mundle, Subhash R., Transportation Consulting Division, Booz, Allen, and Hamilton, Inc., 1346 Chestnut Street, Philadelphia, PA 19107
Nelson, Gary G., Consultant, 128 First Street, Troy, NY 12180
Papacostas, C. S., Department of Civil Engineering, University of Hawaii at Manoa, 2540 Dole Street, Honolulu, HI 96822
Prosperi, David C., Department of Urban and Regional Planning, University of Illinois at Urbana-Champaign, 1003 West Nevada, Urbana, IL 61801
Roess, Roger P., Transportation Training and Research Center, Polytechnic Institute of New York, 333 Jay Street, Brooklyn, NY 11201
Schnell, Gary S., Department of Civil Engineering, University of Hawaii at Manoa, 2540 Dole Street, Honolulu, HI 96822
Schwager, Dianne S., Peat, Marwick, Mitchell and Company, 1990 K Street NW, Washington, DC 20006
Spielberg, Frank, SG Associates, Inc., 4200 Daniels Avenue, Annandale, VA 22003
Turnquist, Mark A., School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14853

Transit Ridership in an Intense Transit Environment: Some Observations

William P. McShane, Paul Menaker, Roger P. Roess, and John C. Falcocchio

Five transit services in an intense transit environment (the city of New York) were surveyed: four bus routes and one rail rapid transit route. In addition, surveys of express bus and automobile ridership on a section of the Long Island Expressway were considered to provide some further mode comparisons. The prime trip purposes were work and school: Work trips accounted for about two out of four trips; except for the premium services, school trips accounted for one out of four trips. Occupation and income generally reflected the source populations. The gender split varied from service to service; buses had the most females (60-80 percent) and automobiles the least (15 percent). Relative to the automobile, riders stated the prime reasons for transit as "automobile not available" or "parking problems." Express bus services drew heavily from public transit; the preferences for it were expressed primarily as comfort and convenience in terms that rank it as a mimic of the automobile—climate control, no transfers, and proximity to trip ends. A picture emerges of a hierarchy preference of modes: (a) automobile, (b) something that mimics automobile, and (c) conventional transit. A case study to replicate the modal gender differences required that two bias coefficients be introduced into a logit model that describes the situation: a distinct preference for bus as a transit mode and a disutility for the automobile that is equivalent to an incremental cost of $2/trip.

This paper presents the results of a set of surveys of transit riders conducted in a relatively well-served, intense transit environment (1). Rider surveys were generally conducted by using mail-back forms that were distributed on board five local transit services: four fixed-route bus routes and one rail rapid transit line. Results from other studies conducted at the Polytechnic Institute are integrated to provide a systematic view of the range of transportation alternatives available to the individual in the environment studied. These other studies include a survey of New York City express bus services and a study of Long Island Expressway (LIE) users.

The ridership studies were complemented by an extensive origin-destination study on the rail rapid transit service. These results are also reported here.

The intent of the study was to relate the ridership observed to both the source population and the ridership of other services and to deduce differences that might be specific to the mode or useful in the planning of transit services.

SERVICES SURVEYED

The five services surveyed are located in the city of New York, an environment that has a substantial transit infrastructure. There are, nonetheless, variations in the amount and type of service available within the city as well as variations in the density and character of the areas themselves.

The five services surveyed are shown in Figure 1 on a map of the city. They are as follows:

1. Two local bus routes in Queens—Queens is one of the five boroughs of New York City; a substantial portion of the residential population commutes to Manhattan as well as to the several central business districts (CBDs), industrial, and commercial areas within Queens.
2. Two bus routes in Brooklyn—Brooklyn is another of the five boroughs of New York City (each is also a county of the state of New York). It has generally higher densities and a higher concentration of low-income areas than does Queens. It too has commercial and industrial areas, and a CBD that, if considered independently, would be the third largest in the nation.
3. One rail rapid transit line on Staten Island—Staten Island is another of the five boroughs, but it was only connected in 1964 to the others directly by the construction of the Verrazano Narrows Bridge. Previously, the only connection was by ferry. Staten Island is almost suburban in character and is currently experiencing significant growth.

Because of the diversity of economic activity, and the multiplicity of CBDs and other concentrations, it is both feasible and practical to view the city and its surrounding areas as an environment that has many transportation alternatives, including one or more feasible transit alternatives in most areas. It is this routine availability of some transit that is of interest.

Two other studies in which some of us were involved were considered to be especially relevant to the present purposes: one of express bus users and one of LIE users. The first was undertaken in 1973 for the New York City Transportation Administration (2). The second was undertaken as part of a study of improvement alternatives of the western section of the LIE, which is located in Queens (3). For convenience, the study section is shown dashed in Figure 1.

These two studies are of particular interest because they represent key alternatives to the local transit services surveyed (i.e., express bus and automobile). Further, these studies involve services that share origins and destinations with the services surveyed here.

SURVEY EXECUTION

Table 1 summarizes the basic facts of each survey: date of execution, direction, forms distributed, forms returned, survey method, crew size, and any relevant additional comments. Note the following:

1. A total of 47 247 forms were distributed, and 17 123 returns were processed;
2. The Staten Island Rapid Transit Railway Company (SIRT) survey was a major effort in logistics; 125 people were retained, trained, scheduled, and deployed for a massive one-day effort; and
3. A substantial diversity of services and areas are represented in the seven services listed.

TRIP PURPOSE AND RIDER DEMOGRAPHICS

The basic ridership of the services studied may be characterized in terms of occupation, purposes, income, gender, and age. The occupation results were somewhat ambiguous, probably due to the way in which people classified themselves. Nonetheless, some interesting patterns were noted and are discussed below.

Trip Purpose

The prime trip purposes are work and school. These encompass between 62 and 93 percent of all trips. All surveys include both peak and off-peak service, although the off-peak service on the LIE was limited.

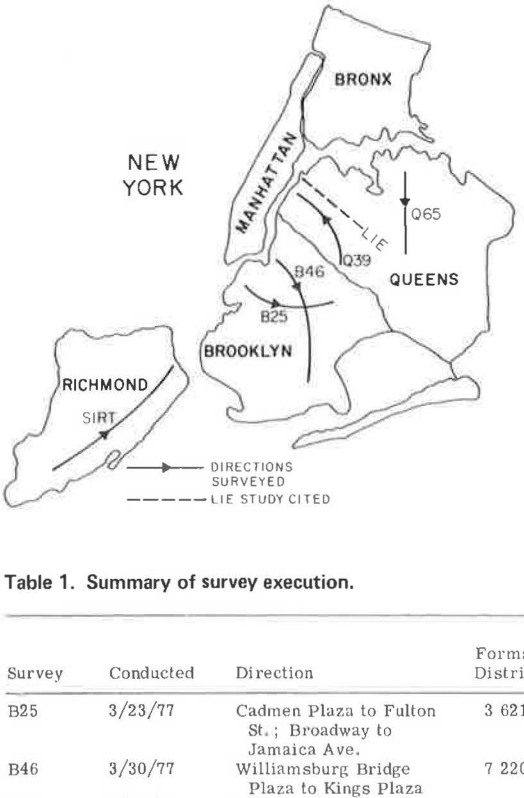
Except for premium services (express bus and LIE), school trips make up 22-28 percent of trip purposes. Thus, one out of every four riders is going to school. Work trips account for about two out of every four riders.

A review of the trip purposes by occupation reveals the following regarding most frequent purpose:

1. About 3 out of 10 trips made by retired persons and housewives are for shopping, 2 out of 10 are for social purposes, and 2 out of 10 are for medical purposes;
2. About 8 out of 10 trips made by students are for school and 1 out of 10 are for shopping; and
3. About 8 out of 10 trips by workers were specifically for work.

The purposes of those miscellaneous trips not included in this listing were diverse.

Figure 1. Location of the five services surveyed.



Occupation and Income

The occupational distribution of users of the surveyed services matched closely those in the source population, considered in light of the destinations available along the route. Except for the lowest-income groups, who are underrepresented on the services, the ridership also reflects the income distribution of the source population.

Gender

Bus transit is startling in that it is dominated by females. This impression is confirmed by data and accentuated by comparison with other modes. Figure 2 shows the male-female gender split on the several routes and modes considered. The pattern is as follows:

1. Local buses have 60-80 percent females,
2. Express bus has close to a 50-50 split,
3. SIRT has 60 percent male, and
4. LIE traffic is dominated by male users.

SIRT is the sole rail transit line on Staten Island and is directed to Manhattan-bound traffic.

Why are there so many female riders on transit services? The data do not allow a conclusive deduction. Nonetheless, some deductions may be drawn from the following observations.

In the counties studied, zero or one automobile per household (85 percent of the cases) is the dominant condition. Bus riders actually own more automobiles per household than the source distribution, but two out of three indicate that an automobile is not available for the trip surveyed.

Women earn less than men within each occupation, even when skewing is allowed due to age distribution by gender.

Most riders express a preference for the automobile mode if it were available and feasible. The automobile mode is expensive. Figure 3 illustrates just the incremental costs of the automobile.

A plausible scenario emerges: females dominate the ridership for simple economic reasons. Where they are the sole jobholders in the household, they are less able to afford the automobile alternatives. Where they are the second jobholders in the household or using transit for nonwork purposes, there is generally only

Table 1. Summary of survey execution.

Survey	Conducted	Direction	Forms Distributed	Forms Returned	Response Rate (%)	Survey Method	Comments
B25	3/23/77	Cadmen Plaza to Fulton St.; Broadway to Jamaica Ave.	3 621	360	10	Handout on bus, mail back	Crew of 22
B46	3/30/77	Williamsburg Bridge Plaza to Kings Plaza	7 220	620	8.5	Handout on bus, mail back	Crew of 20
Q39	3/18/76	Maspeth-Ridgewood to Long Island City	2 121	598	28	Handout on bus, mail back	Crew of 20
Q65	3/26/76	College Point to Jamaica	5 476	1058	19	Handout on bus, mail back	Crew of 20
SIRT	11/17/76	Tottenville to St. George	7 236	5908 (4863 filled in)	82 67	Handout at entering station, pickup at exit station	Crew of 125; forms returned for origin and destination if not filled out
Express bus	1/73		6 285	5257	84	Handout and collection on bus	12 out of 31 surveyed in New York metropolitan area
LIE	11/77	Toward New York City, morning peak Toward Long Island, evening peak	15 288	3322	22	Mail form, mail back	Only surveyed peak-hour users

Figure 2. Gender distribution for several modes and routes.

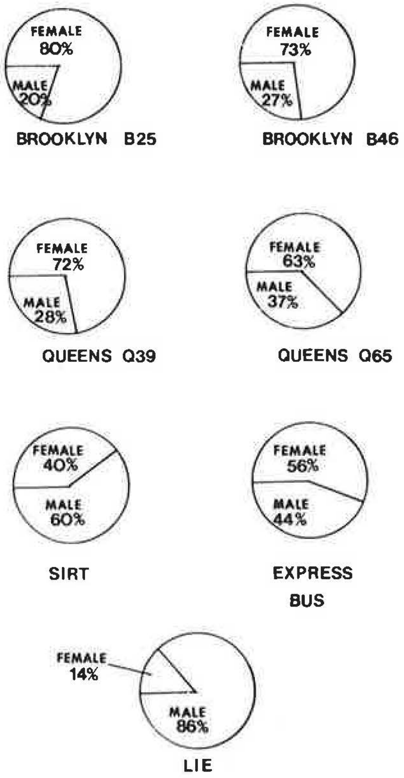
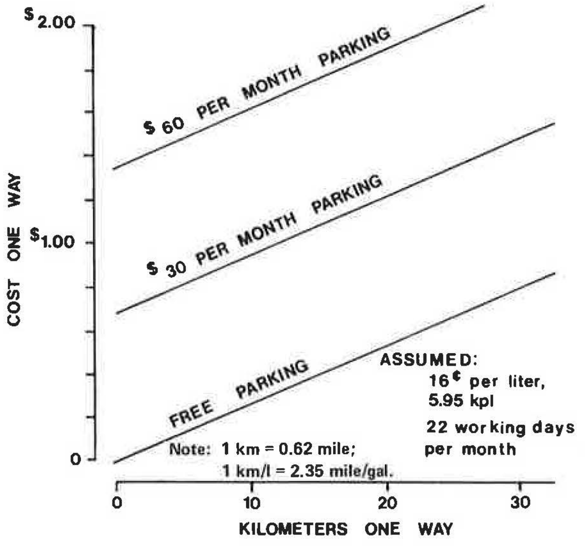


Figure 3. Gasoline and parking cost per one-way trip by automobile.



ASSUMED: 16¢ per liter, 5.95 kpl
22 working days per month

Note: 1 km = 0.62 mile; 1 km/l = 2.35 mile/gal.

one automobile in the household, which is being used by another member or is not affordable to use. The aspect of gender in mode choice was investigated in great detail as a result of these findings. The detailed studies are reported later in this paper.

RELATION TO THE AUTOMOBILE ALTERNATIVE

Two topics of special interest relative to the automobile alternative emerged from the study:

1. Reasons for selecting transit over the automobile mode and
2. Perceived advantages of the automobile over transit and related costs.

Reasons for Selecting Transit

Transit riders were asked the reason they chose transit. The Brooklyn survey asked, "Why not use a car for this trip?" The Queens survey asked, "Why did you choose to use transit for this trip?"

Figure 4 details the results, which are summarized below:

1. The prime reason is that an automobile is not available;
2. Parking problems (not available, too expensive, or too much trouble) are generally the next-most-important factor; and
3. Transit is good is an aggregate of transit is faster and transit is more convenient; it is ranked second in Queens and third in Brooklyn.

The prime reasons for using transit relative to the impracticality of using an automobile (i.e., transit is chosen for negative, not positive reasons) even in the transit-saturated environment of New York City, where the psychological acceptance of transit could be expected to be high.

The "automobile not available" statement was checked relative to the zero-automobile households in the key origin zip codes, the most readily available relevant statistic. Figure 5 shows the relation between the two statistics. It shows the response "automobile not available" to be logical and consistent with the factual information.

The Brooklyn riders were asked specifically for the most important reason for using this bus route (as opposed to other transit alternatives). The nonstudent responses are indicated below:

Response	B25 (%)	B46 (%)
No other transit available	21	49
Comfort and convenience	40	18
Savings in travel time	17	14
Savings in travel cost	13	12
Other	9	7

These were not the order of the responses on the survey form; they are ranked in generally decreasing order for convenience.

Perceived Advantages of the Automobile and Related Costs

The survey of LIE users is of particular interest because these users are often bound for the same general areas as are the transit riders surveyed. The tunnel users are Manhattan-bound; nontunnel users are primarily bound for Queens and Brooklyn, although there is a Manhattan component that reaches Manhattan through Brooklyn via one of several East River bridges.

Asked the principal reason for using an automobile rather than public transportation or other alternative, the response was as follows:

Response	Tunnel (%)	Nontunnel (%)
Car needed during the day	26	20
Convenience worth extra time or money, if any	24	17
Next-best way takes longer	17	23
No other means of making the trip exists	15	23

Response	Tunnel (%)	Nontunnel (%)
Next-best way would cost more	4	6
Other	14	11

Clearly, the need for the car (real or perceived), the convenience, and the time are the substantial factors quoted. Cost is not a major motivation--only 5 percent of the users claim that as the reason for using the automobile (a somewhat obvious result, given that the car is virtually always the most expensive alternate).

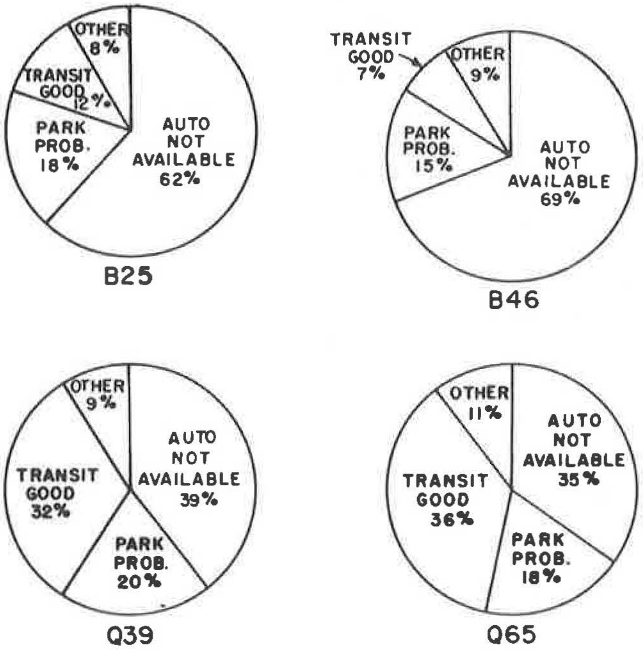
Few work-trip users drive part way, using a transit mode for the remainder of the trip (2 percent of the tunnel users, 9 percent of the nontunnel). Most do not use the vehicle at work (60 percent no use, 36 percent job-related use, 4 percent personal use).

Those who indicate that the next-best way would take longer or cost more were asked for specific amounts. Fifty percent of the respondents judge that the next-best way would take 45 or more minutes. Thus, in response to the cost item, only 5 percent of the total judge that the median (50th percentile) cost penalty would be $40-50/month.

The LIE automobile users encounter substantial expenses. They estimate the median weekly out-of-pocket costs as $15-20. Those who pay for parking pay substantial amounts: The median monthly payment for tunnel users is approximately $75; for nontunnel users, it is $50. Of those who park their vehicles and do not use them, they indicated the following:

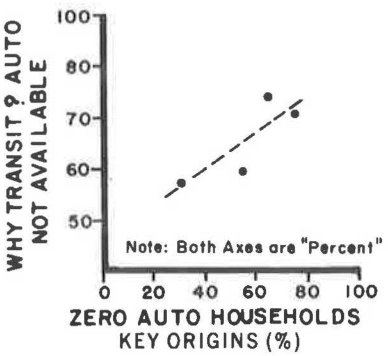
Vehicle Parked	Tunnel (%)	Nontunnel (%)
On street	15	39
In free lot	23	33
In pay lot	62	28

Figure 4. Reasons for selecting transit over automobile.



Note: Nonstudent Respondents

Figure 5. Comparison of automobile availability response with key origins.



EXPRESS BUS EXPERIENCE

The express bus experience offers two important contributions:

1. The modal preferences of the riders and
2. The meaning of comfort and convenience of the riders.

Express buses drew significant ridership immediately on initiation in the city of New York and proved to be both succesful and popular. However, the express bus survey established that 83 percent of the riders were diverted from other public transportation modes. Only 9 percent were drawn from automobile, either as a driver or as a passenger (i.e., a pooled vehicle). Some others were trips not made before, perhaps due to the prior infeasibility of the origin-destination pair in the view of the trip maker.

Figure 6 summarizes the stated reasons for using express bus over the previous method: except for Staten Island, comfort and convenience is selected in 83 percent of the responses. Staten Island is unusual in that trip lengths by public transport are generally much longer than those in other parts of the area. The express bus is the first relatively direct nonwater public transport mode. Note that in all cases travel cost is not a reason for selecting express bus, again because express bus costs more than competing transit modes.

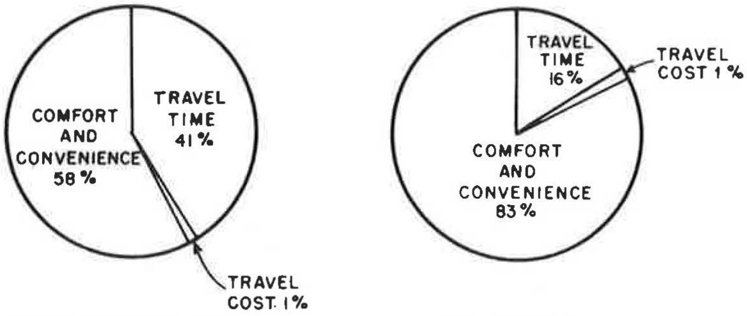
Express bus users who indicated comfort and convenience were asked to select the two most-important factors from a list provided. These factors were later organized into distinct comfort and convenience factors by those who undertook the analysis. Figure 7 summarizes the results:

1. Comfort means having a seat and having air conditioning in the summer; comfort is about two-thirds of the phrase comfort and convenience;
2. Convenience means no transfer, close to destination, reliability of schedule, and convenience of schedule to work;
3. Safety is of greater importance in the off peak than in the peak, where assurance of a seat is of much greater interest; of course, the peak crowd itself provides some security; and
4. Cleanliness and politeness (courtesy of driver) are also elements in the comfort attribute.

This provides some insight into the phrase comfort and convenience, at least as perceived by this rider group.

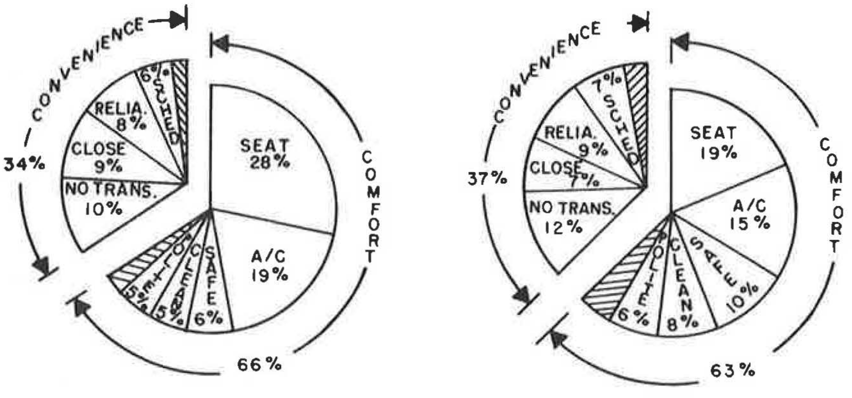
FURTHER INSIGHT FROM THE GENDER PATTERN

In a related effort, one of us developed a microscopic stochastic behavioral implementation model (BIM) and exercised it in a set of case studies (4). One of the cases related specifically to the question of what gender-based model differences must exist in order to conform to the patterns observed above, specifically with regard to work trips.



(A) STATEN ISLAND (B) OTHER

Figure 6. Reasons for using express bus over previous method.



(A) PEAK (B) OFF PEAK

Figure 7. Meaning of comfort and convenience on express bus survey.

The Model and Decision Rule

The model can accommodate a range of decision rules and is suited to specification and modification of parameter values and variable types. A logit model of the following form was selected for our purposes:

$$P_i = J_i e^{-u_i} / \sum_{j=1}^{L} J_j e^{-u_j} \qquad (1)$$

where

P_i = probability of selecting alternative i of L possible alternatives;
J_i = attractiveness of the destination, where J_i is jobs or jobs remaining if the person is home-based, and residences or residences remaining if the person is job-based in his or her decision making; and
U_i = transportation utility of alternative i.

The utility U_i can further be expressed as a function:

$$u_i = \alpha_i + \sum_{j=1}^{P} \beta_k u_i^{(k)} \qquad (2)$$

where

i = an inherent utility of the prime mode on alternative i, referred to as a bias coefficient;
$u_i^{(k)}$ = the kth utility variable that contributes to the measured total utility U_i; and
k = the weight or importance associated with $u_i^{(k)}$.

The mathematical form of the above equation is not unlike that used in the historic macroscopic gravity models (5, 6) and the more recent Urban Transportation Planning System (UTPS) inclusions (7). Its use differs in several intents, however:

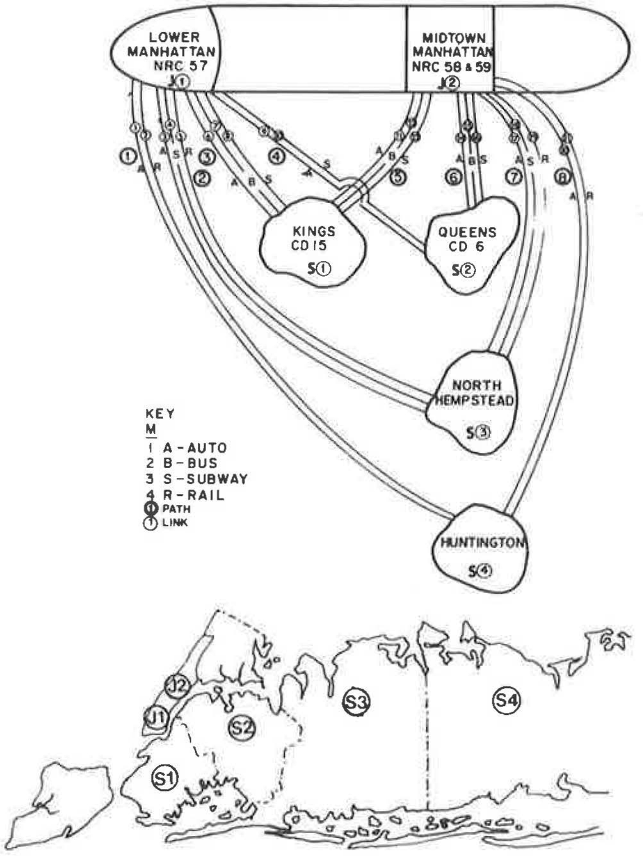
1. The model within which it is to be used is oriented toward the individual's decision process and this form can be so used,
2. The alternatives are to be various paths to the set of feasible destinations and thus imply simultaneous selection of destination and mode,
3. The coefficients J_i are keyed to job or residence opportunities, and
4. The function is used for individuals and is updated in the course of the effort (e.g., J_i may change).

It is recognized that other model forms exist and could be used. Nonetheless, given the available data and the preponderance of the generic form, it was selected for the first implementation.

Case Study

Figure 8 illustrates a set of zones in a corridor between the work centers in Manhattan and several residence zones to the east. Census fourth-count and fifth-count summarizations were available to describe basic characteristics in terms of census tracts or zip codes, respectively. To represent a closed system for modeling, journey-to-work data (8) were used to proportion the total distributions within the zones that were considered. Supplementary data were available for grouping census tracts into convenient aggregations for modeling. The Tri-State Regional Planning Commission supplied files for aggregating census tracts into minor civil divisions or planning districts (9) for residential distributions and nonresidential clusters (NCR) (10) for job-site distributions.

Figure 8. Case study 2: LIE network.



Supplementary data are also available on such matters as gender distributions within occupations (11; 12, tables 173 and 174) and pay differences by gender within occupations (12, table 176; 13).

For the specific variables in the decision model, it was decided to use a mode bias coefficient (A_i), in-vehicle travel time ($IVTT_i$), out-vehicle travel time ($OVTT_i$), and travel cost (TC_i).

These are used to compute the utility, U_i:

$$U_i = A_i + B_1 IVTT_i + B_2 OVTT_i + B_3 TC_i \qquad (3)$$

The coefficients B_1 and B_2 were selected based on occupation and gender.

In a validation check, the correspondence between the predicted base condition and the existing journey-to-work statistics was good.

Replicating the Gender Pattern

Although the overall correspondence just cited was good, it did not extend to the field-observed gender variations. Figure 9 contains a summary of the percentage of males on each of three modes cited and includes the base statistics. Note that no substantial variation is evident, despite the fact that the income variation was explicitly taken into account.

Recall that the utility function is of the form:

$$U_i = A_i + \sum_k b_i u_i^{(k)} \qquad (4)$$

The decision was made to investigate variations in the bias coefficients A_i, which differed by mode, to attempt to explain observed gender-based modal variations.

Figure 9 also illustrates the effect of introducing a bias coefficient of A = -2 and A = -3 for the automobile mode for females only. Note that a negative A leads to a disutility. Clearly, each of the values considered leads to a substantial decrease in the female share of the automobile mode, without any substantial effect on the other modes.

Because of the nature of the model, it is necessary to introduce a specific variation for the bus mode to move toward the observed pattern. This is reported in Figure 10, where values of $A_{\text{bus, female}} = 1$ and $A_{\text{bus, female}} = 2$ are introduced (the base from which they are introduced is the case in which $A_{\text{auto, female}} = -2$). Note that this positive A is a preference.

Figure 10 reports the effect of these last variations, which substantially reduce the male representation on the bus without substantially affecting the other mode patterns. A final case of

$$A_{\text{auto, female}} = -1.5$$
$$A_{\text{bus, female}} = 1.5$$

was introduced to attempt to refine the match to the observations. The result is also reported in Figure 10 and is a rather close match. It would not be appropriate to attempt closer values because of the inherent uncertainty in the exact data values.

It is interesting that the $A_{\text{auto, female}}$ value thus obtained can be translated into an equivalent travel cost increment of approximately \$2/trip (or \$4/day). This can be obtained by transforming the terms in the utility function, one into the other:

$$A_i = b_3 \times TC \qquad (5)$$

and similarly for the other terms in U_i.

It is interesting that a travel-time increment of \$4/day is approximately \$900/working year (229 days), which one may think of as the incremental annual cost of owning a second car (over and above the daily tolls and parking fees that are already taken into account in the explicitly specified travel cost for the given mode for all potential users). The concept that the $A_{\text{auto, female}}$ may be equivalent to purchase of a second car is worthy of note. This is particularly true in the environments tested, where single-car households are by far the most common. Needless to say, this term might not exist in a suburban or rural environment.

One may observe that the coefficient $A_{\text{bus, female}} = 1.5$ is equivalent to an inherent preference for the bus, which has the same utility valuation as 40 min of additional out-of-vehicle travel time. Thus, a bus trip that has 40 min more access time than an available alternative, such as subway, is equally attractive. It does raise a question, which must remain unanswered at this time, as to the motivation of this apparent preference.

Clearly, to explain observations, substantially different valuations are needed for male versus female. This study cannot resolve why those differences exist, or even quantify them in a systematic scientific survey of users. Nonetheless, it has to be observed that these variations must logically exist to explain observed phenomena and that behavioral models must explicitly take the potential for such variation into account.

CONCLUDING COMMENT

The variations herein have special interest because they include observations on several different modes. Regarding the gender-based analysis, the disaggregation by gender and occupation used in this work is not the

Figure 9. Percentage of males on modes with automobile bias: LIE case study.

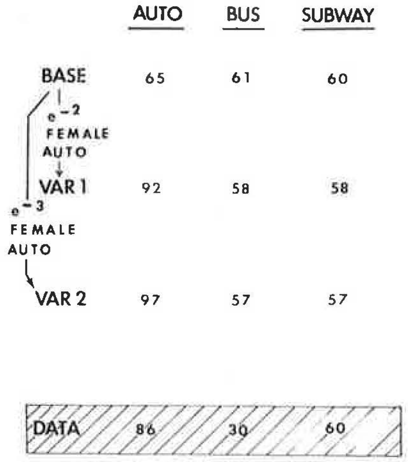
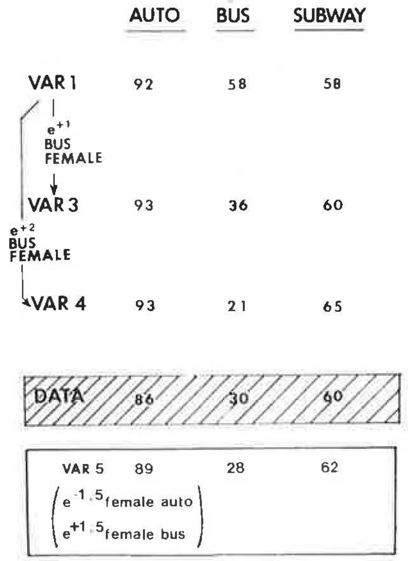


Figure 10. Percentage of males on modes with automobile and bus bias: LIE case study.



only one that could have been used and is not necessarily the most basic. Their use in the transportation field is, however, still relatively novel. One must understand that there are contributing factors under the umbrella headings of gender and occupation that can be explained in terms of the more conventional basic variables. Nonetheless, this disaggregation is useful to investigate policy questions related to what if one or more societal changes were made, such as equal pay, greater representation of women in certain occupations, or equalization of automobile availability.

In this work, the case studies included an investigation of what modifications were needed in the behavior model coefficients so as to conform to existing observations of gender representation in various modes. Other cases were executed that considered the case of equal pay for male and female. This did not introduce any substantial change in mode use by gender.

Clearly, there is an opportunity to investigate a variety of scenarios and to trace their implications. This would include the various what if questions and a consideration of how women value their travel parameters and modal choices relative to those of men.

ACKNOWLEDGMENT

This paper was produced as part of a program of research and training in urban transportation sponsored by the Urban Mass Transportation Administration, U. S. Department of Transportation. The results and views expressed are the independent products of university research and are not necessarily concurred with by the Urban Mass Transportation Administration.

REFERENCES

1. W. R. McShane and others. A Study of Transit Rider Characteristics. Urban Mass Transportation Administration, 1979. NTIS: PB 295-107/AS.
2. J. C. Falcocchio and others. Survey of the Ridership of Express Buses in the City of New York. Polytechnic Institute of New York, Brooklyn, summary rept., May 1974.
3. Survey of the Western Long Island Expressway: 1977-1978. New York State Department of Transportation, Albany, 1978.
4. P. Menaker. A Behavioral Implementation Model. Polytechnic Institute of New York, Brooklyn, Ph. D. dissertation, June 1979.
5. Peat, Marwick, Mitchell and Company. A Review of Operational Urban Transportation Models. U. S. Department of Transportation, Rept. DOT-TSC-496, April 1973.
6. Computer Programs for Urban Transportation Planning. Urban Planning Division, Federal Highway Administration, July 1974. U. S. GPO stock 050-001-00125-0.
7. Urban Transportation Planning System: Users Guide. Planning Methodology and Technical Support Division, Urban Mass Transportation Administration, Rept. UTP.40.74.1.5, July 1977.
8. 1970 Census Journey to Work—Residence to Worksite Summary. Tri-State Regional Planning Commission, census tabulation tapes, New York, 1970.
9. Minor Civil Division Converter File. Tri-State Regional Planning Commission, census tabulation tapes, New York, 1970.
10. 1970 Census NRC Crosstabs File. Tri-State Regional Planning Commission, New York, census tabulation tapes, 1970.
11. 1975 Disclosure Tape: Employment Security Data by Zip Code. Port Authority of New York and New Jersey, New York, 1975.
12. U. S. Census of Population and Housing 1970, NYSMSA. Bureau of the Census, U. S. Department of Commerce, Vol. 1, Pt. 34, March 1973.
13. U. S. Census of Population and Housing 1970: Volume II, Occupation of Persons with High Earnings. Bureau of the Census, U. S. Department of Commerce, 1970 Census of Population Subject Repts., Sect. 7F, June 1973.

Publication of this paper sponsored by Committee on Transit Service Characteristics.

Development and Application of Performance Measures for a Medium-Sized Transit System

William G. Allen, Jr., and Lewis G. Grimm

This paper summarizes the results of a study of service performance measurement and operating guidelines for the Delaware Authority for Regional Transit (DART) system. This fleet of 100 buses serves the Wilmington metropolitan area and is typical in many respects of many medium-sized bus systems across the country. The project consisted of several elements. First, a brief overview was presented of the historical perspectives on transit performance standards and the current state of the art, specifically noting activities at the state and regional level over the past few years. Next, a preliminary set of transit performance measures and operating guidelines was formulated for local review and comment. To assist in the evaluation of the adequacy of the preliminary performance measures and service standards, the draft standards were used to assess DART's existing operations. This assessment was hampered by a number of data inconsistencies, due primarily to the fact that much of the data required had been collected over a period of several years by using different data collection and analysis procedures. Efforts were made to minimize these inconsistencies and, where this could not be done, recommendations were made for improved data collection procedures to eliminate this problem in future years. As part of the service assessment, order-of-magnitude cost estimates were prepared to define the general range of capital and operating investment that would be required by DART to modify its current services so as to be in greater compliance with the proposed service standards and operating guidelines. The last step of the project was the preparation of guidelines to assist local agencies in the implementation of the service standards and operating guidelines and the continuous monitoring of DART's performance relative to these standards. This element of the project addressed the manner in which the current infrastructure for transit planning could be improved and described the appropriate level of detail and methodology for the continual evaluation of DART's performance. A discussion was presented of the basic procedures by which to amend or modify the service standards and operating guidelines.

This paper documents the results of a recently completed study of the development of transit service standards and operating guidelines for the Delaware Authority for Regional Transit (DART) in Wilmington, Delaware. The project was conducted for the Wilmington Metropolitan Area Planning Coordinating Council (WILMAPCO). It represented one aspect of the overall DART planning program, which is intended to increase patronage, improve operational efficiency, and increase regional transit accessibility in the Wilmington urban area. The service standard development project was expected to help fulfill these goals through the achievement of the following objectives:

1. The development of a realistic, comprehensive set of performance measures, service standards, and operating guidelines for current and future transit service;
2. The determination of the existing level of transit service based on these standards and guidelines; and
3. The development of a continuing process of monitoring transit data, updating the standards as necessary, and using this information to improve the quality, efficiency, and cost-effectiveness of transit service in the region.

Consistent with these basic objectives, this project produced a group of transit service standards and operating guidelines that provide a framework for the provision of cost-effective public transportation services throughout the Wilmington metropolitan area.

TRANSIT STRUCTURE IN THE WILMINGTON AREA

DART is an independent authority established in 1969 under the laws of the state of Delaware with the responsibility for the provision of fixed-route, fixed-schedule public transportation services in New Castle County, Delaware. Recently, DART was abolished as an independent authority and renamed the Delaware Association for Regional Transit. The new organization is a subsidiary of the Delaware Transportation Authority, a division of the Delaware Department of Transportation.

DART receives funds from a variety of sources. Urban Mass Transportation Administration (UMTA) Section 5 funds are used for capital acquisition and operating assistance; the nonfederal share of capital costs is provided by the Delaware Department of Transportation. The nonfederal share of operating assistance is provided jointly by the state of Delaware, New Castle County, and the city of Wilmington, although no formal allocation of nonfederal operating assistance funds has been agreed to by the participating jurisdictions. In recent years Delaware's share has been between one-half and two-thirds of the total nonfederal annual financial assistance; New Castle County and Wilmington provide the difference. With the recent reorganization of DART, the state has assumed a larger share of the financial burden. However, the absence of a formal cost-sharing program among the participating local jurisdictions remains a problem in this area, as in many similar areas throughout the country. Indeed, a question of how much money should be allocated for DART was one of the reasons for the initiation of this project, as it was felt that an objective procedure for examining the cost-effectiveness of DART's services would facilitate agreement on an equitable sharing of costs by the local jurisdictions. The DART performance measures and standards are similar in many respects to those developed for Boston and Seattle.

TRANSIT PERFORMANCE MEASUREMENT

As part of their comprehensive state-of-the-art series of planning manuals for urban transportation, the National Committee on Urban Transportation published two manuals in 1958 on procedures for measuring transit service and establishing warrants for new services (1, 2). These manuals have been considered the standard references in transit evaluation for the past two decades. They were written primarily for use by managers to monitor transit operations internally, as this was the extent of performance measurement for many years. In an era of steadily declining profits and increasing public funding during the 1960s, a few systems developed similar measures to gauge their overall performance in providing service.

In 1973, one of the earliest systems of formal public evaluation of transit service was developed by the Pennsylvania Department of Transportation to assist in the allocation of state operating assistance to transit systems in Pennsylvania (3). The available documentation indicates that, after an initial shakedown period, this program is accomplishing most of its goals. A similar system has been developed in California (4). Considerable recent efforts by researchers, administrators, managers, and planners have been directed at the need for transit performance measures and how to develop and use them. Since the early 1970s, several of the more progressive transit authorities have documented their performance measures, service criteria, and route planning guidelines as part of their short-range service planning efforts.

This current interest in performance measurement is largely the result of a key motivating factor that has developed over the past 10 years: the necessity of public ownership and funding of transit. Public ownership came as a result of governmental recognition of transit as a necessary public service. In this regard, much recent state and federal legislation has allocated funds for capital improvements and operating assistance to public transit properties. But, as it has solved some of the problems of transit, government subsidy has created others, not the least of which are accountability, incentive, and control in transit management (5).

Of immediate concern is the possibility that any government subsidy could diminish the motivation to provide the best possible service, unless operations are controlled to some extent by guidelines and performance measures. Such guidelines, applied judiciously, can be an integral part of a program to increase transit efficiency and effectiveness, while at the same time helping to safeguard the interests of the public, who now have a permanent involvement in the provision of public transportation services.

There are several current views on the purpose of transit performance measurement. These are generally divided into two areas: funding and planning. As transit's financial needs increase over time, along with other demands on the local coffers, a mechanism for justifying these needs becomes essential.

Documented performance measures and guidelines also allow transit system management to make more rational decisions about internal resource allocation and provide a means of communicating service policies to elected officials and the general public. Such guidelines also benefit the operations planning function and aid in the development of short-range plans. In fact, transit's input to the transportation systems management (TSM) process can be directly related to actions identified through the application of transit performance standards (e.g., increases in transit operating speed through traffic engineering techniques). Finally, transit marketing efforts can benefit from the information developed from the monitoring of operations and ridership.

Performance indicators are derived from a knowledge of the locally accepted goals and objectives for transit. In this sense, they help to gauge whether or not the service is meeting those goals. To this end, indicators can be used in several ways. A variety of indicators describe various aspects of transit efficiency and effectiveness; some can be used in the evaluation of overall system performance or individual route performance and others in the evaluation of a single transit function, such as maintenance or procurement. Some indicators are particularly useful to those who make decisions about external operations. The importance and utility of any given measure depends on the perspective of those who interpret it.

From a policy standpoint, transit performance indicators reflect much more than the quality or economy of system management. They reflect government decisions that directly or indirectly affect transit operations, local operating conditions, and local transit use patterns. For example, fare policies established by local decision makers greatly affect financial performance indicators. Regional wage differentials affect cost-efficiency indicators. Disproportionate peak-period transit use affects labor productivity and vehicle utilization indicators.

Clearly, peer group comparisons of transit performance should not be made strictly on the basis of indicators. Similarities and differences of various communities and transit operations should be carefully considered, along with potential differences in data element definition, when making such comparisons.

On the other hand, indicators can be a valuable aid to managers in the comparison of performance on different routes in a single system. For example, routes may be ranked on the basis of indicators such as farebox revenue and operating cost, passengers per vehicle hour, or passenger kilometers per vehicle kilometers. Another valuable use of indicators is to trace changes in system performance over time. Indicators can facilitate the analysis of cost and ridership trends and changes in trip patterns and overall service levels (6).

Selection of how many and which transit data to use as performance indicators is somewhat dependent on the individual system involved. Several generally accepted criteria for the selection of these performance indicators are that they should be

1. Related to a stated system objective,
2. Easily understandable and definable,
3. Unbiased and objective,
4. Measurable from available data,
5. Methodologically correct (i.e., properly separating input and output measures), and
6. Acceptable to the parties involved.

Performance measures or indicators are the means by which broad system evaluation criteria are measured. These indicators need to be related to hard numbers. Therefore, each indicator is quantified relative to a service standard that is a benchmark against which existing and proposed services may be assessed. Standards can be defined in various ways, including the following:

1. Historical transit industry par values, where these exist,
2. Guidelines from other transit properties that have them,
3. Average figures from comparable transit systems,
4. Dynamic standards expressed in terms of DART system averages or distributions of DART operating data, and
5. Independent judgment of transit operations and planning professionals.

No single technique is completely adequate, but a combination of all of them is usually found satisfactory due to the variety of data elements considered in standards development.

The particular use of performance indicators and standards is a direct function of the main purpose of evaluation. For example, the primary purpose of the Pennsylvania Department of Transportation Operating Guidelines and Standards Program is to assist the state in allocating operating assistance funds to the various transit properties in Pennsylvania. Therefore, each eligible system is evaluated on an objective scale and assigned a grade based on its performance, generally on an annual basis. The operating assistance share is then determined from the performance for the current year. This method allocates funds as a reward for relatively good or improving service. A mechanism has also been developed to fund those systems that are not performing relatively well in order to help them to meet the standards.

Most statements of transit service standards and operating guidelines are associated with individual systems, however. This internal monitoring can be periodic or continuous, as part of an ongoing service evaluation process. It can be formalized by the development of a permanent framework of committees and a specific review process, as is the case in Massachusetts, or it can be done on a less structured basis, as in Dade County, Florida.

The level of evaluation is another key consideration. Transit performance can be measured in relation to the transit industry as a whole, in relation to individual routes within a system, in relation to the various areas or area types served, in relation to the affected party (i.e., rider, operator, or community), and in relation to various functions internal to the system such as operations, scheduling, or maintenance (6).

The final (or perhaps the initial) consideration in the development and use of service standards is their relationship to funding. It is imperative that service guidelines and standards be determined in conjunction with financial policy, both within the transit operation and between the operation and the levels of government involved with funding. It is pointless to set standards that are not attainable within available financial resources; standards should reflect what is attainable. There may, of course, be alternative policies and standards to meet specific contingencies, such as a drastic gasoline shortage, and there may be service objectives that exceed the adopted standards, which would be made effective were additional funds to become available (6).

To be sure, the achievement of some objectives (such as on-time performance) may not lend itself to a direct cost determination. Meeting such standards becomes not only a matter of resources but of management's use of these resources. However, the financial impacts of meeting most standards can be ascertained. For areas whose governments have quantified their desire for transit in terms of financial commitments, like the Regional Transportation District in Denver, service guidelines and standards have become the tool to help shape the service in the proper direction.

STUDY PRODUCTS

The major product of this study was a set of transit performance measures and associated service standards, operating guidelines, and warrants for new service that can be used to objectively assess the efficiency and effectiveness of public transportation services in the Wilmington metropolitan area. The term "performance measure" was used in this project to denote a descriptive item that could be measured in numerical terms or assessed in qualitative terms. Thus, on-time performance is such a measure. Service standards are the benchmark values associated with each performance measure. If DART's operation was deemed to be better than a particular service standard, the system was said to meet or exceed that service standard. Failure to meet the standard results from performance worse than the threshold value.

In addition to the measurable performance indicators, a series of warrants for new service and operating guidelines were also developed. These are also benchmarks but are not used to measure the system's performance as such; instead, they are intended to be used as rules of thumb to assist in providing transit service in the area. One example is the warrant for placement of bus shelters, which permits an objective examination of where passenger waiting shelters should be located based on the number of persons waiting at a particular bus stop and the frequency of service provided at that stop.

The total group of performance measures, warrants, and service guidelines was organized into three basic sections: service design guidelines, operational performance measures, and ridership social-environmental measures. The service design guidelines addressed the overall structure of the route system, including the classification and location of bus routes, regional accessibility via public transportation services, and warrants for new and improved fixed-route, fixed-schedule services. These guidelines are basically indicators of the organization and the extent of bus routes in the Wilmington area. The operational performance measures include management-related measures for the examination of transit operational efficiency (e.g., maintenance standards) and some consumer-related quality of service items, such as percentage of buses with operating air conditioners. These are a general measure of management policy and resource-allocation techniques. The ridership social-environmental measures describe the use of the system and are thus indicators of transit effectiveness, that is, how effective the transit service is in attracting and keeping ridership.

The three different perspectives were deemed necessary in order to properly address the varying concerns of the transit rider, DART management, and the regional community in the improvement of public transportation services. DART management is concerned with all three sectors, riders are primarily concerned with the quality of service, and the community at large is interested in the effectiveness of the transit service.

The transit performance measures and standards were based on experience in other urban areas, available data on DART's operations and of the operations of similar-sized transit systems, and generally accepted transit industry guidelines. Although the literature review indicated that few hard-and-fast transit performance measures and standards exist, the indicators developed during the course of this study represent current thinking in the area of transit performance measurement. We should note that there is more industrywide agreement on what the indicators should be rather than what the standards should be.

For example, many transit properties use the statistic of vehicle kilometers per road call as a measure of the adequacy of their maintenance program, but considerable debate exists over what minimum value should be used as this service standard since it depends on many factors, over a number of which (such as roadway maintenance) the transit operator has little or no control. The indicators and service standards developed during the course of this project should not be regarded as permanent but as a first step in the evolutionary process of critical self-examination by DART. In addition, care must be used in measuring DART's performance (or indeed the performance of any transit system) against these indicators. These service standards and operating guidelines should be used in the proper context, at the correct level of detail, and with operational data as recent and accurate as possible. Most importantly, all indicators are not of equal performance. The priorities assigned to each performance measure should reflect a compromise of the collective views of transit riders, the transit manager, and the community in order to en-

Table 1. Summary of DART peak-period on-time performance.

		Percentage of Peak-Period Buses		
Route No.	Name	Early	On Time	Late
A-1	Bellefonte	17.2	75.9	6.9
A-2	Concord Pike	30.4	52.2	17.4
A-5	Foulk Road	21.1	73.7	5.2
A-10	Delaware Avenue	40.0	60.0	0.0
A-11	40th & Washington Street	18.5	81.5	0.0
A-12	Boulevard	40.9	50.0	9.1
A-14	Centreville	30.0	70.0	0.0
A-15	Harvey Road Express	0.0	100.0	0.0
B-1	Mill Town Road	57.1	42.9	0.0
B-5	Newport	33.3	50.0	16.7
B-6/7	Newark/Prices Corner	23.1	69.2	7.7
B-8	West 5th Street	73.3	26.7	0.0
B-12/14	Crossgates/Hockessin	35.7	64.3	0.0
B-16	Newark Express	38.9	61.1	0.0
AB-3	Kynlyn/S. Harrison	31.3	65.6	3.1
AB-4	Cleveland Avenue/Vandever	31.9	68.1	0.0
C-3	Castle Hills	10.0	80.0	10.0
C-6/7	Stratford/Llangollen	7.1	85.7	7.1
C-15	New Castle/Dunleigh	31.5	63.2	5.3
Systemwide average		31.1	64.1	4.8

sure a proper emphasis in the assessment of the performance of an individual transit system. In any event, the indicators and guidelines developed during this project are listed below:

1. Service design guidelines—service classification system, spatial guidelines, directness of service, route layout, frequency of service, temporal guidelines, bus stops, passenger shelters, and warrants for new service;
2. Operating performance measures—speed, layover time, load factors, schedule adherence, service dependability, safety and security, complaints, passenger amenities, transit information systems, and productivity; and
3. Ridership social-environmental measures—ridership measures, social standards, environmental standards, and system impact standards.

With appropriate modifications for individual systems, these could form the basis of a generalized transit performance measurement program for most medium-sized transit operations in the United States or Canada.

ASSESSMENT OF CURRENT DART OPERATIONS

Once the initial set of proposed performance measures, service standards, and operating guidelines had been developed, an evaluation was made of DART's current operations to define the degree to which existing services were in compliance with the proposed standards. An important use of this assessment was in measuring the reasonableness of each of the proposed standards; that is, whether or not they were either too lenient or too demanding in terms of the current operations and should thus be modified to present a more realistic target for the improvement of services in the near term. Any such assessment should represent only one element of a comprehensive evaluation of any transit system. Also, due to differences in data collection procedures and vintage of transit data, there may be a number of inconsistencies associated with any such assessment. Thus, it is important to clearly define an implementation and monitoring program that will minimize such data variability in the future.

One important use of an on-going assessment of transit operations is a definition of any changes required to meet the adopted standards and guidelines. Included with the DART evaluation was an estimate of the capital and operating costs required to modify the services as necessary. As part of this assessment, a determination of the order-of-magnitude costs was prepared to assist the affected local agencies in future transit service planning and evaluation.

The actual assessment of DART's operation used data collected during the fall of 1977 and the spring of 1978. An individual assessment was prepared for each of the proposed transit service standards and operating guidelines on either a route-by-route or systemwide basis, as appropriate. Each assessment consisted of a short definition of the indicator or guideline, a description of the measure for each of the service standards, a brief discussion of the context of the performance measure, and an assessment of how the DART operation complied with that proposed service standard. An example assessment is presented below for the service standard dealing with schedule adherence.

Schedule adherence is a measure of whether or not transit vehicles run according to the published timetables. The appropriate variable is on-time performance, which was defined in this study as the percentage of runs that pass a given checkpoint no earlier than the scheduled time and no later than 5 min after the scheduled time. Schedule adherence is traditionally an important measure to transit system users. A 1975 DART ridership survey indicated that it is considered very important to transit-dependent and non-transit-dependent passengers (7). A high degree of on-time performance results in shorter waiting times, fewer missed buses, and an overall increase in rider confidence in the system. Good schedule adherence results from well-written schedules, close street supervision, and bus operator cooperation. An example of service standards is as follows:

1. Leave times at the terminal points of a route will be exact. Approximate leave times will be determined at intermediate points (i.e., major street intersections and activity centers) along a route.
2. No trip will leave either a terminal point or an intermediate point ahead of the scheduled leave time.
3. On time is defined as 0-5 min late.

Schedule adherence is also closely related to load factors, maintenance standards, and other operating conditions such as layover time provisions. The table below presents the schedule adherence standards that were employed in the assessment.

	Minimum Percentage of Service on Time		
Time of Day	0-15 Min Headway	16-30 Min Headway	Infrequent Service (headway > 30 min)
Peak hours	75	85	95
Off-peak hours	80	95	95
Weekend	80	95	95

Table 1 presents the results of a March 1978 peak-hour schedule adherence check in comparison to the proposed schedule adherence standard. Table 1 indicates that almost one-third of the actual runs operated on that day ran ahead of the posted time schedule. The late runs (e.g., those more than 5 min late) were relatively low in number (only 4.8 percent of total runs) but the extremely high percentage of early runs (31.1 percent of the total operations) is a major problem.

The table below presents the on-time performance summarized by frequency of service based on schedules in effect on December 11, 1977. At that time no peak-hour routes had headways greater than 30 min.

	0-15 Min Headway (%)		16-30 Min Headway (%)	
Performance	Actual	Standard	Actual	Standard
Early	30	0	34	0
On time	65	75	63	85
Late	5	25	3	15

Based on this analysis, the schedule adherence is slightly worse for the less-frequent routes, which is the opposite of what normally occurs. Early runs are intolerable on lower-frequency routes since a missed bus may mean a wait of 30 min or more for the next bus.

CHARACTERISTICS OF THIS PROJECT

One of the characteristics that sets this study apart from similar ones is the size of the transit system. DART is a medium-sized operation that has 82 scheduled peak-hour buses out of a total fleet of 100. It is, thus, one of the smallest transit systems to have sponsored the formal development of transit performance measures and operating guidelines. The acceptance of study results illustrates that the concept of using performance measures for internal system reexamination is applicable for even small systems, as long as they are presented and used in the proper context. This is different from the perspective of the statewide service standards and operating guidelines developed in New York, Pennsylvania, and California, which are intended to assist the respective state departments of transportation in the allocation of financial assistance to a number of operations in each state. Used in that manner, systemwide measures of productivity and cost-effectiveness were required, whereas the DART study has developed both systemwide indicators, as well as more detailed route performance standards. DART is currently the only fixed-route, fixed-schedule system operating in Delaware.

Perhaps the most unusual feature of this project was the high degree of citizen participation in a rather technical transportation planning study. Recent experience in the Wilmington metropolitan area indicated that the more successful transportation and land use planning activities had included substantial citizen involvement throughout the duration of the project. At first, it was not thought that a technical study of this nature would be appropriate for significant citizen input. However, after discussions among WILMAPCO, other involved local agencies, and the consultant, it was decided to initiate a major citizen participation effort specifically including current transit users (both captive and noncaptive), as well as representatives of the elderly, handicapped, and minority populations within the DART service area.

The success of this activity far exceeded initial expectations, as the membership of the citizen's transit advisory committee (CTAC) maintained a high level of involvement throughout the entire project. CTAC members carefully reviewed all interim study products with a view toward making the project documentation as useful and understandable as possible for both laypersons and elected local officials who would have to make decisions as to the level of future transit investments in the region. The viewpoints of current transit system users and of the elderly and handicapped representatives on the committee provided valuable input to the project and enabled transit management and planning staff to obtain a much better understanding of the actual needs of DART's existing clientele.

The possibility always exists that a study will become only another report that will wind up on the shelf of the client agency with few, if any, of the recommendations being implemented. It was thus encouraging that several of the specific project recommendations were implemented while the study was still under way. For example, a number of the assessments of current DART operations identified minor problems with the printed timetable that were corrected during DART's semiannual schedule revision process. The result was that an improved level of service on a much more easily remembered schedule was provided in several areas, with plans for all other DART routes to be subjected to similar detailed schedule revision during the next two years.

In addition, since the assessments allow for an estimate to be made of how much change would be required in the level of service to meet the proposed standards, the results of this project were used during the course of recent labor negotiations and better enable DART management to explain why changes in scheduling and driver assignment were necessary. The successful implementation of the standards is indicated by the fact that all the participants in the project accepted the spirit of the guidelines and are working toward full implementation of the service standards. The Delaware Transportation Authority is in the process of formally reviewing and endorsing the standards for future DART operations planning activities.

An especially noteworthy aspect of this project was the degree of interaction among the staff representatives of the metropolitan planning organization (MPO), the regional transit authority, the state department of transportation, local governmental agencies, and the citizen's transit advisory committee. This opportunity to conduct an objective evaluation of the system and to help provide a direction for future transit investment in the region enabled the involved parties to work together more effectively than in past projects. In this particular instance, the function of the consultant was not only to provide technical input but also to serve as a mediator between the divergent viewpoints of the various participating agencies.

CONCLUSIONS

The primary conclusion of this study is that it is possible, practical, and worthwhile for a medium-sized transit system to develop and use a set of performance measures and service standards. The utility of performance measurement as a planning, management, and public relations tool is not restricted to systems that have several hundred or more vehicles. Smaller systems may even have an advantage in that many of the same people who helped develop the standards are heavily involved in the daily operation of the system. In the case of the DART system, these performance measures are more likely to be implemented because relatively less red tape exists.

Another conclusion is that some of the more technical aspects of transit operating standards can indeed be understood by citizens and planning and policy-oriented officials when presented in the proper manner. This includes carefully defining all terms, making liberal use of graphics and tables for explanation, and providing concrete examples of the implications of meeting or not meeting the standards. The review of the appropriateness of various service standards proceeded much more easily after the advisory committees were shown how well DART currently operates and were given a rough estimate of the cost to meet each standard.

A third conclusion is that a considerable amount of prior planning should go into the development of performance measures and standards. This planning should answer several interrelated questions, such as the following:

1. What are the intended purposes of these performance measures?
2. What are the specific objectives for transit service?
3. Is there a useful financial, planning, and operating data base available? and
4. Are the funding sources sufficiently predictable that it is realistic to attempt to meet the standards?

The experience of this study is that the value of having indicators and standards is directly related to the degree of positive response to these questions. Without some answers to the questions developed before or during a study, performance measures are surely useless.

The final conclusion is that it is preferable for the development of indicators and standards to be undertaken at the local level, ideally beginning with transit management. The DART study was funded by the local MPO (WILMAPCO), with the cooperation of DART and close coordination with the Delaware Department of Transportation. The local initiative approach has the advantage of being more sensitive to local objectives and concerns, more precise, and more useful as a management tool than are state or federally mandated evaluations. Statewide programs, by definition, are authorized by state legislatures for the primary purpose of funding allocation. Although statewide evaluations may serve this function admirably, they are not generally as helpful to the individual systems.

ACKNOWLEDGMENT

This paper was adapted from the draft final report, the Development of Service Standards and Operating Guidelines for the Delaware Authority for Regional Transit, prepared by Barton-Aschman Associates for WILMAPCO. We would like to thank Gordon Schultz of Barton-Aschman for his input throughout the study. A special note of appreciation is due Robert Taylor of the Delaware Department of Transportation, Raymond Malenfant of WILMAPCO, and Steven Welch of DART. These individuals provided invaluable guidance and assistance and are in large part responsible for the success of this project. This study was sponsored by WILMAPCO as part of its unified transportation planning work program and was financed in part by a Section 9 grant from UMTA. The information and opinions expressed in this paper are ours, and we assume full responsibility for the paper's contents.

REFERENCES

1. Measuring Transit Service. National Committee on Urban Transportation, Public Administration Service, Chicago, IL, Procedure Manual 4A, 1958.
2. Recommended Standards, Warrants, and Objectives for Transit Services and Facilities. National Committee on Urban Transportation, Public Administration Service, Chicago, IL, Procedure Manual 8A, 1958.
3. Operating Guidelines and Standards for the Mass Transportation Assistance Program. Bureau of Mass Transit Systems, Pennsylvania Department of Transportation, Harrisburg, 1975.
4. E. Fuller. Performance Measures for Public Transit Service. Division of Mass Transportation, California Department of Transportation, Sacramento, 1978.
5. W. G. Allen, Jr., and F. DiCesare. Transit Service Evaluation: Preliminary Identification of Variables Characterizing Level of Service. TRB, Transportation Research Record 606, 1976, pp. 41-47.
6. Proceedings of the First National Conference on Transit Performance, Norfolk, Virginia, September 1977. Urban Consortium; Public Technology, Inc., Washington, DC, Jan. 1978.
7. Harland Bartholomew and Associates. Greater Wilmington Short-Range Transit Improvement Program: Update. Wilmington Metropolitan Area Planning Coordinating Council, 1976.

Publication of this paper sponsored by Committee on Transit Service Characteristics.

Diagnostic Tools in Transit Management

Subhash R. Mundle and Walter Cherwony

Historically, transit management had to rely on a technique known as peer-group comparison to identify strengths and weaknesses in the performance of their system. In this technique, performance indicators for the system under study are compared with the average performance of systems that have similar characteristics. This method, though useful, is deficient in that it does not totally reflect the differences in operating characteristics or environment among transit properties. This paper presents a diagnostic tool for comparing performance among transit systems by suggesting a method to eliminate deficiencies in the traditional approach. The paper suggests that combined uncontrolled and controlled comparisons be used to identify relative strengths and weaknesses in performance. The uncontrolled comparison is the traditional approach in which system performance is compared with average performance of the peer systems. The controlled comparison is performed by comparing the actual performance with the expected performance. The expected performance is calculated from models that can be developed from the experience of the peer systems. This paper presents a case study in which uncontrolled and controlled comparison concepts were used to identify strengths and weaknesses of 11 bus depots in the New York City Transit Authority. The paper presents 10 transportation and maintenance performance indicator models that were used to calculate the expected depot performance. The models were developed through stepwise multiple regression analysis of the New York City Transit Authority's actual operating statistics for fiscal year 1977. The paper also discusses how the uncontrolled and controlled comparisons were subsequently used to set priorities among depots for remedial action. The application of the performance comparison technique discussed in this paper to smaller systems would require comparison of the system's performance with that of other similar transit properties.

The limited availability of public funds to underwrite transit deficits and the increasing gap between operating cost and passenger revenue have caused transit managements throughout the nation to strive for improvements in transit system efficiency, effectiveness, and productivity. To accomplish this objective, a two-step process is often employed by transit managers: (a) the identification of problem areas and (b) the development of an action plan to remedy these deficiencies. In view of the limited time and resources available to transit managers, the first step is of utmost importance if efforts are to be focused on problem areas of highest priority. For this reason, there is a need for simple, easy-to-use diagnostic tools to quickly pinpoint transit deficiencies.

The traditional approach to the assessment of transit performance is to compare the performance of a transit system under study with that of transit properties that have similar characteristics. Typically, this peer group comparison is carried out for a variety of performance indicators in which the results of the system under study are contrasted with the average for the peer group. This method is helpful in providing an analytical framework; however, it is deficient in that it does not totally reflect the difference in operating characteristics or environment among transit properties. In essence, this simple comparison technique is uncontrolled since it does not account for inherent differences among systems.

To remedy this situation, a supplemental diagnostic tool is proposed in which the performance of a transit system is measured not only in absolute terms against a peer-group average but also in relative terms with respect to where the system performance should be, given its operating environment. This latter comparison technique is termed a controlled comparison since it attempts to reflect inherent differences among transit properties.

This paper presents a case study of this diagnostic tool applied to the bus operations of the New York City Transit Authority (NYCTA) as part of an overall organization and management study of the entire Metropolitan Transit Authority (MTA). The analysis was performed for 11 depots and 10 performance indicators to identify depots that are deficient in performance and the nature of those problems. Although the New York situation is unique in that the system is sufficiently large that comparisons can be made among depots, the approach is readily applicable to smaller transit properties where the comparison is made with peer-group systems.

CONCEPT DEVELOPMENT

Traditionally, transit system performance has been assessed by using a simple peer-group comparison technique. A variety of indicators are specified to measure various apects of transit performance. Peer-group systems are selected on the basis of similarities among the system under study and the peer-group properties in terms of factors such as fleet size, geographical region, and demographic characteristics of service territory. In the New York City case study, the peer group is merely the 11 depots that constitute NYCTA bus operations. As depicted in Figure 1, the comparison is made between each depot's performance relative to the system or peer-group average. Depot results above the system average would suggest superior performance and results below the average would indicate a deficiency. The problem with this simplified comparison technique is that it does not account for differences in operating characteristics and environments among the depots. For example, a depot that serves suburban Staten Island might exhibit superior fuel economy results, but a Manhattan depot that serves a densely developed portion of New York City might score well below the system average. However, these results might be misleading because they would not reflect the impact of bus operating speed. Thus, the Manhattan depot might be performing better than could be reasonably expected and the Staten Island depot could be performing worse than expected. To rectify this situation, the peer-group comparison is expanded to include a comparison of actual performance relative to expected performance (i.e., controlled comparison).

As shown in Figure 2, the reference line for this comparison is a 45° line rather than a horizontal line, as in the uncontrolled comparison. The actual performance would be the same as that used in the uncontrolled comparison; however, the expected value would be determined from regression analysis of the peer-group depots. The disadvantage of using only a controlled comparison is that it does not relate performance relative to the system. For example, a depot that is worse than expected but better than the average would not be considered a priority location for remedial action, although performance should be improved.

Since each comparison technique provides only part of the information needed for a diagnostic tool, system performance should be assessed by both methods. By combining both procedures, each depot's performance can be categorized into four possible outcomes, as depicted in Figures 3 and 4. The implications of each of the four categories of results are as follows:

1. Better than average and better than expected—a depot that has better performance than the systemwide average as well as actual performance better than expected; the two comparisons are compatible and suggest superior performance;
2. Better than average but worse than expected—a depot that has better performance than the systemwide average but the actual value is less than expected; these results suggest good performance with room for improvement;
3. Worse than average but better than expected—a depot that performs below the systemwide average but better than expected; these results of the controlled comparison suggest satisfactory performance in the face of difficulty; and
4. Worse than average and worse than expected—a depot that scores poorly both in terms of systemwide average and expected value; these results suggest poor performance and the need for further analysis and improvement.

Placement of each of the 11 NYCTA depots in one of these four categories results in the identification of depots that require remedial action for each indicator. By quantifying the results of both comparison techniques, the priority order for remedial action of the depots can be set in order to attain maximum benefits with limited financial resources.

METHODOLOGY

The application of the conceptual framework discussed above requires development of performance indicator statistics for the uncontrolled comparison and the development of performance indicator models so that expected performance can be calculated for the controlled comparison. (These models were designed for U.S. customary units only; therefore, values are not given in SI units.) The first step was the development of a number of transportation and maintenance performance indicators for NYCTA depots that were developed from available operating statistics for fiscal year 1977.

As shown below, 19 operating statistics were considered in the analysis:

1. Operating speed,
2. Passengers per bus mile,
3. Passengers per bus hour,
4. Annual passengers per operator,
5. Operators per bus,
6. Annual bus hours per operator,
7. Operator pay hours per bus hour,
8. Annual bus miles per operator,
9. Annual miles per bus,
10. Ratio of peak bus requirements to base bus requirements,
11. Average fleet age,
12. Spares ratio,
13. Buses per mechanic,
14. Annual bus miles per mechanic,
15. Annual bus hours per mechanic,
16. Bus miles per quart of oil consumed,
17. Bus miles per gallon of fuel consumed,
18. Bus miles per maintenance-related road call, and
19. Facility age.

Figure 1. Uncontrolled comparison.

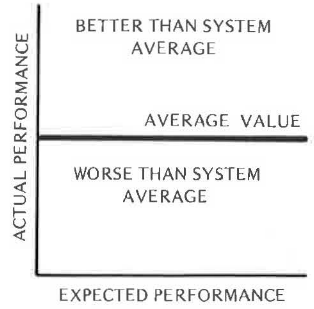
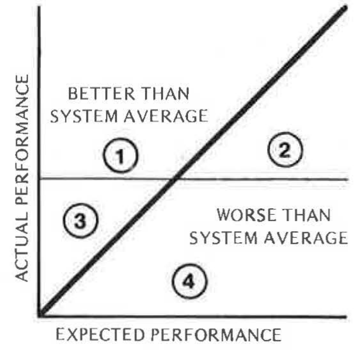


Figure 2. Controlled comparison.

Figure 3. Combined comparison.



However, only 10 performance measure were used in the analysis. As shown below, five indicators each were selected to assess transportation and maintenance performance:

Transportation Indicators	Maintenance Indicators
Operating speed	Buses per mechanic
Operators per bus	Annual bus miles per mechanic
Operator pay hours per bus hour	Bus miles per quart of oil consumed
Annual bus miles per operator	Bus miles per gallon of fuel consumed
Annual miles per bus	Bus miles per maintenance-related road call

In the next step, stepwise linear multiple regression analysis was used to examine the relationship between each performance indicator and those factors or variables that influence it. The regression analysis procedure may be visualized by examining the relation between operating speed and passengers per bus mile, as shown in Figure 5. The scatter of points for each operating depot indicates that a perfect relationship between the two variables does not exist. However, a straight-line graph can be found that best fits the data points. Figure 5 indicates the line of best fit for the NYCTA depots. It is obvious from this example that only one variable is necessary to describe operating speed. In the case of a more complex relationship, more than a single factor may have a role in the determination of the performance indicator. For example, the formula that defines buses per mechanic for the depots relies on both the ratio of peak to base bus requirements and average fleet age.

In selecting a final set of relationships for each of the 10 transportation and maintenance performance indicators listed above, the following guidelines were established:

1. All variables can be easily obtained from NYCTA's normal data collection procedures,
2. Equations should be relatively simple to apply in terms of included variables,
3. Formulas should be logical both in terms of the sign of each coefficient and the variables included, and

Figure 4. Possible performance results.

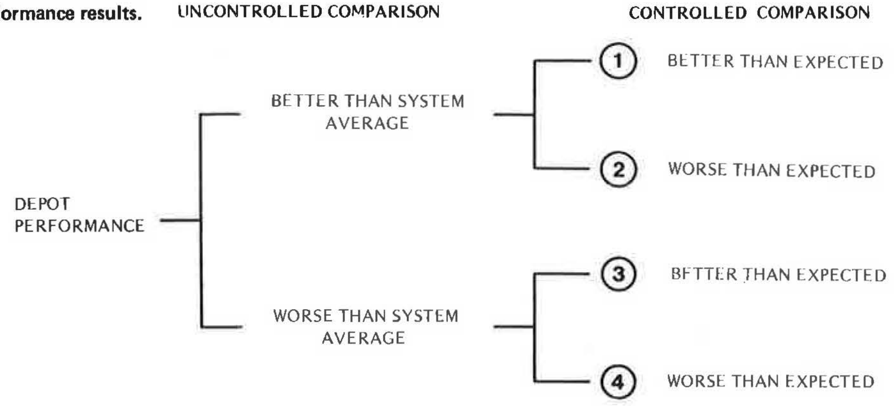
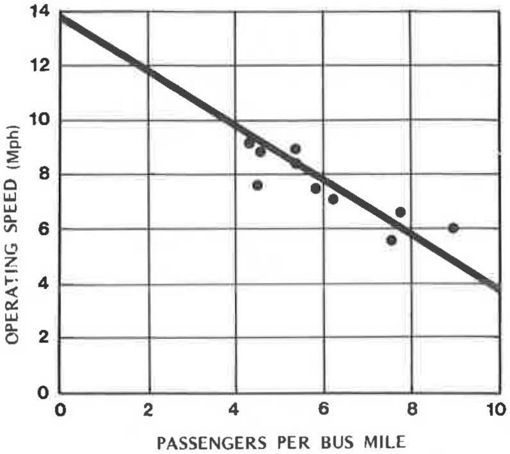


Figure 5. Relationship between operating speed and passengers per bus mile.



4. Relationships quantified should represent a reasonable fit of the data.

Several iterations of the statistical analysis were performed to arrive at the relationships that achieved the greatest satisfaction of the items cited above. It is recognized that numerous diverse factors influence the performance indicators and that quantification of formulas with only one or two explanatory variables represents a simplification. Further, some potential explanatory variables do not lend themselves to mathematical quantification.

As a result of the statistical analysis, mathematical models that describe the relationships for 10 performance indicators were quantified. The models developed provide a reasonable means of predicting expected performance, while making due allowance for the differences in depot operating conditions. These relationships, although not perfect, are relatively easy to derive and can be used on a continuing basis to monitor changes in the level of performance for each measure. The following formulas were used in the controlled peer-group comparisons to predict expected values of the performance indicators.

Transportation Indicators

The formula for operating speed employs only a single explanatory variable. The inverse relationship between operating speed and passengers per mile is obvious, since passenger use, which measures patron boardings and alightings, influences the number of stops.

$$\text{Operating speed} = 13.829 - 1.007 \text{ passengers/bus mile} \quad (1)$$

Two explanatory variables are required to describe operators per bus, which is inversely proportional to the ratio of peak/base bus requirements and directly proportional to miles per bus. As the peaking characteristics of the system increase, the number of drivers per bus decreases and approaches one. Conversely, a system that exhibits a uniform demand throughout the day would enable each bus to be driven by more than a single driver. As the miles per bus increase, the number of drivers per bus also increases; that is, greater use of capital resources causes the need for more drivers. Logically, the more miles that a system operates, which in part is a function of speed, the greater the number of operators needed to provide service.

$$\text{Operators per bus} = 2.149 - 0.357 \text{ ratio of peak bus requirements to base bus requirements} + 0.242 \times 10^{-4} \text{ annual miles/bus} \quad (2)$$

The NYCTA formula relates operator pay hours per bus hour to a single variable—peak-to-base bus ratio. This relationship reflects both the diseconomies associated with establishing the labor requirements based on the peak demand requirement as well as the restrictions and penalties established in the collective bargaining agreement (e.g., spread time, guaranteed time, and minimum straight runs).

$$\text{Operator pay hours per bus hour} = 1.466 + 0.074 \text{ ratio of peak bus requirements to base bus requirements} \quad (3)$$

The number of miles each operator can drive annually is a direct function of the speed at which the bus travels. Higher speeds produce greater mileage statistics; lower speeds translate into fewer miles. In part, this relationship reflects the desire by transit management to provide the same number of hours to each driver to the maximum extent possible.

$$\text{Annual bus miles per operator} = 1289.105 + 1481.040 \text{ operating speed} \quad (4)$$

The statistical analysis for annual miles per bus suggested that two explanatory variables are appropriate. The first, peak/base ratio, is inversely related to the miles per bus. Obviously, with more buses in service for only a limited time period (morning and evening rush hours), the buses accumulate fewer miles. The other variable that influences the number of miles per bus is operating speed. Not surprisingly, higher speeds translate into more miles.

$$\text{Annual miles per bus} = 13\,094.820 - 5876.167 \text{ ratio of peak bus requirements to base bus requirements} + 3187.265 \text{ operating speed} \quad (5)$$

Maintenance Indicators

The NYCTA formula for buses per mechanic has as independent variables both peak and base bus requirements and average fleet age. Because maintenance activities are a function of the number of buses as well as their utilization, the buses per mechanic are directly proportional to the ratio of peak to base requirements. On the other hand, average fleet age is inversely proportional to this performance indicator. Older buses, which are more prone to mechanical failures, require more maintenance employees. Conversely, newer coaches, which should experience fewer mechanical problems, reduce the number of mechanics.

$$\text{Buses per mechanic} = 3.154 + 0.168 \text{ ratio of peak bus requirements to base bus requirements} - 0.094 \text{ average fleet age} \quad (6)$$

The formula to describe annual bus miles per mechanic consists of operating speed and average fleet age.

$$\text{Annual bus miles per mechanic} = 121\,882.700 + 5602.496 \text{ operating speed} - 9490.317 \text{ average fleet age} \quad (7)$$

Oil consumption is inversely proportional to average fleet age in that older buses require more oil per mile.

$$\text{Bus miles per quart of oil consumed} = 202.707 - 11.441 \text{ average fleet age} \quad (8)$$

The NYCTA formula for bus miles per gallon of fuel consumed consists of a single independent variable—operating speed. Experience with most vehicles indicates that higher speeds cause greater fuel economy. Further, higher speeds imply less stop-and-go operation and a reduction in idling time.

$$\text{Bus miles per gallon of fuel consumed} = 1.729 + 0.208 \text{ operating speed} \quad (9)$$

Two variables were included in the equation for bus miles per maintenance-related road call—average fleet age and buses per mechanic. Frequency of breakdowns is inversely proportional to vehicle age, which is logical because older buses are more prone to failure. The model also includes the number of buses per mechanic. As fewer mechanics are available to perform maintenance duties, the miles per breakdown increase. Conversely, fewer buses per mechanic translates into an improved road-call experience.

$$\text{Bus miles per maintenance-related road call} = 7047.614 - 301.010 \text{ average fleet age} - 895.973 \text{ buses per mechanic} \quad (10)$$

The 10 performance indicators and relationships listed above were used in the controlled and uncontrolled comparison of NYCTA depots. For the controlled comparison, expected depot performance is calculated by substituting appropriate values of independent variables in the models discussed. For example, expected depot operating speed is calculated by substituting the passengers per mile statistic for that depot in the operating speed model.

DISCUSSION OF THE RESULTS

In the New York MTA management study, performance of the NYCTA bus depots was evaluated by using the uncontrolled, controlled, and the combined comparisons for each of the 10 indicators discussed. Performance indicators simply highlight the strengths and weaknesses of the operation. Once the weaknesses are identified, steps can be taken to determine the causes and to remedy the deficiencies. In the interest of brevity, the discussion of only one performance indicator—operating speed—is presented here to illustrate the usefulness of uncontrolled and controlled comparisons, followed by the priority order of depots for remedial action (1).

The uncontrolled and controlled comparisons of depot operating speed are illustrated in Figures 6, 7, and 8. From the uncontrolled comparison, in which depot operating speed is compared with the system average (Figure 6), we see that the Castleton Avenue Depot exhibits the highest operating speed and that the 126th Street Depot exhibits the lowest. This is not surprising since Castleton Depot provides suburban service on Staten Island and 126th Street Depot operates mostly in Manhattan. The controlled comparison, in which the actual depot operating speed is compared with its expected value (Figure 7), indicates that the 126th Street Depot exhibits much better actual performance than its expected value, whereas the Crosstown Depot exhibits far worse actual performance than its expected value. The uncontrolled and controlled comparisons by themselves do not indicate those depots that need remedial action.

The combined comparison (presented in Figure 8) indicates the depots that exhibit similar operating speed performance. Those depots that indicate actual performance below system average (as well as below their expected values) obviously warrant further investigation for remedial action. From Figure 8, it can be observed that, even though the operating speed at the 126th Street Depot is below the system average, the comparison with its expected value indicates a satisfactory performance. Several NYCTA depots, such as Crosstown, East New York, Fifth Avenue, and Freshpond, indicate performance below system average as well as below their expected values. These depots evidently constitute a peer group for improvement in their operating speed performance.

The comparisons, similar to the ones discussed above, were prepared for each of the five transportation and maintenance indicators. The uncontrolled and controlled comparisons were then used to establish management priority to remedy transportation and maintenance deficiencies in NYCTA depots. The priority of depots is determined by ranking the uncontrolled and controlled performance results. The composite uncontrolled ranking for transportation and maintenance functions is prepared in two steps. In the first step, depots for each of the five transportation and maintenance indicators are ranked in a straight ordinal fashion from 1 to 11, where 1 represents the best actual performance and 11 represents the worst actual performance. In the second step, the cumulative score for the depots is again ranked from 1 to 11, where 1 represents the depot that has the least cumulative score for five transportation or maintenance indicators and 11 represents the depot that has the highest cumulative score. The uncontrolled rankings of depots for transportation and maintenance functions are given in Table 1.

The controlled ranking of depots is performed in two steps, similar to the uncontrolled ranking. However, in the controlled ranking, the depots are ranked based on the percentage difference between actual and expected performance (i.e., 1 = the depot that exhibits the highest percent better performance and 11 = the depot that exhibits the highest percent worse performance). The controlled rankings of depots are also shown in Table 1. The rankings shown in this table were used to establish management's priorities for implementing remedial steps at the depots. From the priority scheme given below, it is evident that the comparison of uncontrolled and controlled ranking was helpful in identifying performance deficiencies that would otherwise have gone unnoticed.

Management Priority	Transportation Performance	Maintenance Performance
Top	East New York	Jamaica
	126th Street	East New York
	Crosstown	126th Street
	Castleton	Castleton
Second	Fifth Avenue	Fifth Avenue
	Jamaica	Crosstown
	Freshpond	Ulmer Park
Third	Flushing	Queens Village
	Queens Village	Flatbush
	Flatbush	Freshpond
	Ulmer Park	Flushing

For example, under uncontrolled comparison, Castleton Depot ranked on top of all the other depots in transportation and maintenance functions, which indicates superior performance. However, the controlled ranking indicated that there was significant room for improving performance at this depot. Consequently, Castleton Depot was assigned top priority for remedial action. Therefore, both uncontrolled and controlled comparisons are needed to correctly identify deficiencies that are not evident from either of the methods alone. It should be pointed out that the comparison technique and ranking scheme discussed here can also be used to determine the priority of depots for individual performance measures.

Figure 6. Uncontrolled comparison of depot operating speed—actual performance versus system average.

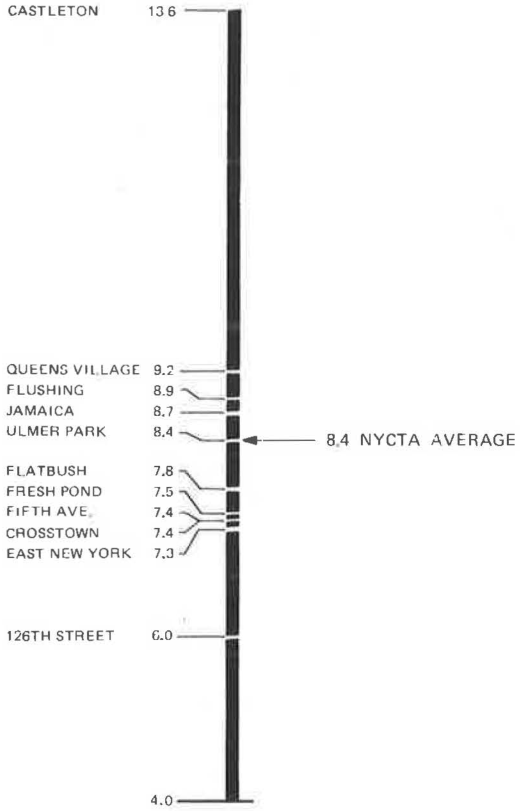
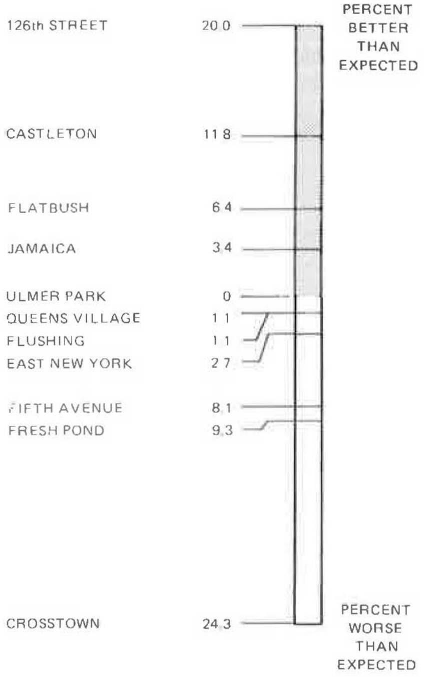


Figure 7. Controlled comparison of depot operating speed—actual versus expected performance.



CONCLUSIONS

The methodology used to compare transportation and maintenance performance of NYCTA depots allows comparison of each depot performance with the system average, as well as with its expected value. Uncontrolled and controlled comparisons provide a useful technique because it not only takes into consideration the interdependency between different indicators but makes an allowance for unique operating characteristics of each depot.

This methodology can be easily adapted to compare performance of depots in other larger transit systems that operate out of multiple operating facilities. Of greater significance is that the technique can also be used to compare performance of one system with the performance of peer systems to identify deficiencies.

The use of only a single comparison technique (uncontrolled or controlled) does not provide sufficient information to diagnose problem areas and to develop a program of priority remedial action; instead, both comparison procedures must be employed. The suggested

Figure 8. Combined comparison of depot operating speed.

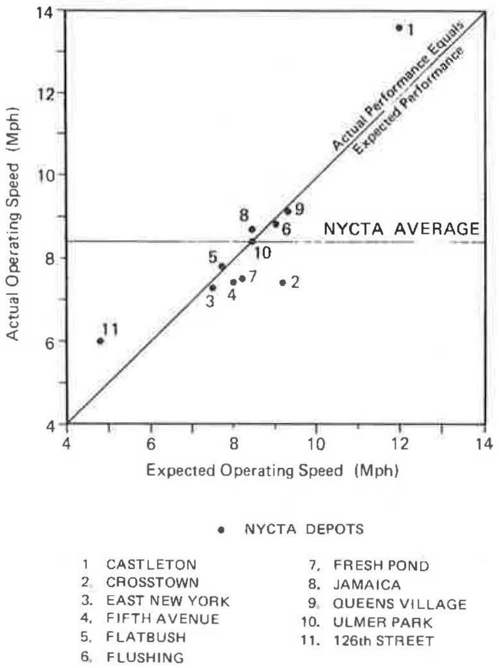


Table 1. Depot ranking for transportation and maintenance performance.

	Transportation Performance Ranking		Maintenance Performance Ranking	
NYCTA Depot	Uncontrolled	Controlled	Uncontrolled	Controlled
Castleton	1	8	1	8
Crosstown	8	11	9	6
East New York	10	10	8	10
Fifth Avenue	9	7	4	7
Flatbush	4	2	2	3
Flushing	5	4	3	1
Freshpond	6	5	6	2
Jamaica	6	6	10	11
Queens Village	3	3	7	4
Ulmer Park	2	1	5	5
126th Street	11	9	11	9

diagnostic tool in this paper is easy to apply and should be used by transit managers on an ongoing, continuous basis to monitor and improve performance.

REFERENCE

1. Task IV: Surface Peer Group-Cost Center Analysis, New York MTA Management Study, Draft Final Report. Transportation Consulting Division, Booz, Allen, and Hamilton, Inc., Philadelphia, PA, May 1979.

Publication of this paper sponsored by Committee on Transit Service Characteristics.

Portfolio Model of Resource Allocation for the Transit Firm

David C. Prosperi

An agency resource-allocation model is presented that allows a fuller understanding of performance by linking aggregate performance indicators with diaggregated measures of route-level activity. The evaluative framework is based on the portfolio-choice model of financial management. In this model, aggregate return and risk parameters are found by examining the resource-allocation pattern and ridership levels for individual routes. Although this is primarily an economic utility-maximizing approach, the model parameters were calculated for two time periods and compared in an evaluation of resource reallocation. Before and after levels of service and ridership counts for 41 routes operated by the San Diego Transit Corporation provide the inputs to the modeling effort. Results show that the average return increased with minimal risk impacts in the post-reallocation period, indicating a better resource-allocation package. The relationship between resource allocation and the aggregated average return is thus made explicit and the change in aggregate indicators viewed directly as a function of management and operational considerations at the route level. Finally, the model or method of analysis is evaluated in terms of both its conceptual and measurement procedures.

An evaluative framework for overall transit performance is presented that employs the concepts of resource allocation and economic returns for individual routes. Prior conceptualizations and analyses of transit performance have focused on aggregate measures of inputs, produced outputs, and consumed outputs (1). Route-level analysis of performance is a relatively new phenomenon and critical evaluation of route-level activity is a young enterprise. Route-level demand models have appeared only in the past few years (2, 3) and route-level performance evaluations are still in the stage of ad hoc development (4-6). The purpose of this paper is to formalize and demonstrate an internal resource-allocation model that allows fuller understanding of performance by linking aggregate systemwide performance indicators with disaggregated measures of route-level activity.

The framework is based on the portfolio-choice model of financial management. A portfolio is a collection of activities to which resources can be allocated. The portfolio-selection probelm is to choose investment levels for individual activities so as to maximize a utility function defined over both the expected value of the collected returns of individual activities and the total risk associated with achieving those returns. Thus, the approach allows the derivation of an aggregate systemwide indicator of performance based on an analysis of route-level activity. In the transferral of the model from financial management to an evaluative tool for transit managers, investments are resources assigned to routes, individual activity returns are measures of the patronage of transit routes, the expected value of the portfolio is a weighted average of route investment and return levels, and risk is a measure of the variation in the mean expected return.

TRANSIT PERFORMANCE: ORGANIZATIONAL AND ROUTE ANALYSIS

Transit performance is becoming synonomous with the terms efficiency and effectiveness. In a recent paper, Fielding and others identified nine preliminary performance indicators that focus on these twin evaluative criteria (1). Sticking with standard definitions, efficiency is the ratio between produced output and the amounts of input required to produce them. Effectiveness is the degree to which outputs are consumed or used and a relative measure of output quality. Cost or resource effectiveness is a composite measure and is defined as the ratio of consumed outputs to costs or input magnitudes. As such, this last performance concept serves as an overall indicator of the performance of the entire service provision process, including both the production and the consumption of services.

Fielding, Glauthier, and Lave suggest three general areas of use for their set of performance indicators. They are as follows:

1. Management uses, including the identification of activities within a system in which achieved indicator values are above or below some norm, the comparison of activities of one's own agency against indicator values achieved by similar properties, internal monitoring of production processes, and the stimulation of discussion among transit operators and key personnel;
2. Evaluation of suborganization performance, including the development of objectives and auditing procedures for improvement of activities, route evaluation, and the facilitation of labor negotiations; and
3. Inputs into public policy, by focusing discussion on a common set of issues and criteria.

Although these are all universally considered as legitimate uses of performance measurements and evaluation, specific procedures to accomplish such objectives have yet to be developed at a generalized and transferable level.

Two broad approaches to route-level performance evaluation are in practice in the transit industry. The first approach concentrates on service standards, both with regard to design (e.g., route frequency, headways, and growth potential) and operations (e.g., schedule adherence, travel speeds, accidents, complaints, and lost runs). In a recent study of several western transit properties, Glauthier and Feren (7) described how these standards are related to transit goals and objectives and reviewed within-house methodologies wherein surveillance of design and operational variables led to improvement in route performance.

The second approach to route-level performance focuses on the economic and financial ratio. These ratios may serve, alternatively, as indicators of the relationship between produced output or consumed output and inputs, or between consumed outputs and produced outputs. The most common form found in transit-evaluation literature is the revenue/cost (R/C) ratio. Other measures, such as passengers per kilometer or subsidy per passenger also are employed in various situations. The basic procedure for the individual route R/C ratios is to disaggregate an agencywide cost function into individual cost functions for routes (8). At a conceptual level, as much detail as desired is possible, although practical difficulties usually result in somewhat simplified models. This methodology has been applied and discussed by both Cherwony (9) and Nelson and Nevel (10).

What is important at this juncture is how these route-level economic and financial ratios are used by transit managers and what interpretation can be given to them. Cherwony, for example, employs the concept of cost centers. The general approach of comparing route-level financial ratios assumes that routes are in competition with one another for resources. As such, the analysis of route performance is placed within the overall context of economic resource-allocation problems. Moreover, since such competition occurs with a budget constraint, the evaluation system leads to a zero-sum outcome (i.e., if more resources are placed on a route, another route must have its resources depleted). Following the principles of resource allocation formulated in microeconomics, Cherwony correctly asserts that it is possible to increase the aggregate R/C ratio of a transit property by shifting resources from relatively less cost-effective routes to those that have relatively high cost-effective ratios. Such shifting of resources may be done even though all the component parts of a system are operating in a deficit situation—a frequent happenstance in not-for-profit systems.

Nelson and Nevel have used disaggregated R/C ratios for a different purpose. R/C ratios were defined for route segments and mapped for each route of South Bend (Indiana) Transit. Visual inspection of the resulting map enabled the authors to suggest service modifications to loop ends. Although not intended as such, their approach serves as an important development in the spatial analysis of both revenue generation and costs. With the emergence of equity as a third important and distinct evaluative criterion, the Nelson and Nevel approach could yield fruitful results in future studies.

This brief review of transit performance and management is concluded with several observations. First, the performance criteria currently in use are mostly economic and socioeconomic in nature, even though the provision of transit is largely publicly provided. The use of economic indicators for what is essentially a social service arises from the general concern for cost or resource effectiveness in the public sector. Second, although economic indicators are employed by evaluators, there is no overall economic framework that guides the provision of service systemwide. Economic notions are used in a descriptive instead of a prescriptive manner. There is a noticeable paucity of theoretical models and their application to guide resource allocation. Finally, discussions of how aggregate performance measures are related to disaggregated measures are virtually nonexistent. The model presented below seeks to help fill this void.

RESOURCE ALLOCATION AND THE GENERAL PORTFOLIO APPROACH

The link between aggregate performance indicators and route-level activity lies in the rules of resource allocation as formulated in the theory of the firm. Resource-allocation rules are developed within traditional microeconomics to help decision makers achieve optimality in aggregate production efficiency. In the real world, optimal allocations are generally impossible for two reasons: (a) equity constraints that mandate diversified allocation patterns and (b) the relative permanence of system infrastructure (either for reasons of sunk capital or for maintenance of political support). Thus, the best one can usually do is an evaluation study to identify resource reallocations that are in the proper direction (i.e., that move towards an optimal distribution).

The objective function of resource allocation (assuming for the moment that optimal allocation patterns are possible) is the maximization of economic return from use of scarce inputs. In traditional production situations, marginal analysis is employed to determine whether a scarce resource is equally productive in all output-generating processes. The rules are founded on the concept of opportunity cost. When the marginal-analysis equality conditions do not hold across all processes, the producer is missing an opportunity to increase aggregate return by reallocating resources from activities that have smaller marginal returns to ones that have larger marginal returns. It is within this vein, for example, that Cherwony fosters movement of resources from less cost-effective routes to more cost-effective routes.

The usefulness of traditional resource-allocation rules as a general framework of economic analysis is weakened by the context of perfect certainty in which they are formulated. The notion of certainty violates real-life observation, both in its basic assumption that outputs are perfectly predictable and in its implication that, with a linear production function and a budget constraint, a decision maker will optimally choose the project or process that offers the highest rate of return. Thus, a decision maker will invest all of his or her resources in the best alternative. This obvious failure of traditional resource-allocation rules to cope with empirical aspects of real-life situations paved the way for new approaches to resource-allocation analysis, including the use of nonlinear production functions or modes of analysis.

The portfolio-selection model provides an alternative method of assessing resource-allocation problems. The model is well documented in the literature on financial asset management (11-13). The analytical problem may be posed as a general question: Given a set of economic return and variance estimates regarding future outcomes on individual securities, what is the optimal portfolio an investor should select? In other words, How much of an investor's resources should be allocated to each security? There are two crucial distinctions between the portfolio-selection model and the resource-allocation models of traditional microeconomics:

Figure 1. Preference map for risk-averse decision makers.

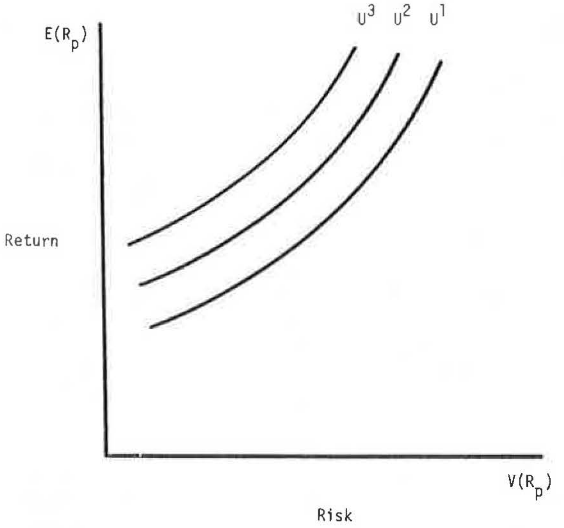
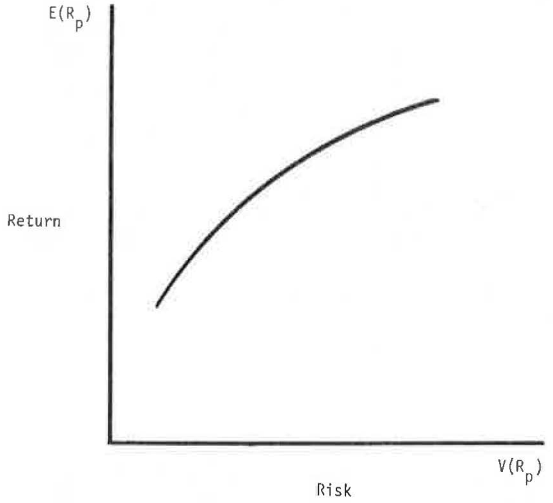


Figure 2. Efficient set of portfolios.



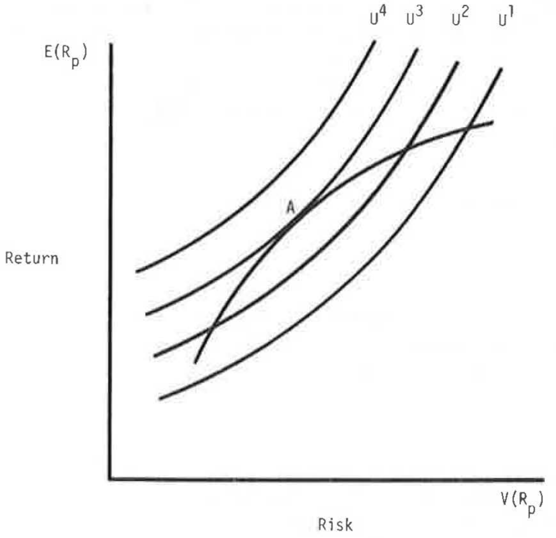
1. Instead of making estimates on future outcomes solely in terms of point estimates, additional information is incorporated that accounts for the extent to which subsequently realized outcomes may differ from those predicted; and

2. As a result of the explicit consideration of an uncertainty factor in the decision-making model, a two dimensional return-risk criterion emerges in place of simple maximization of aggregate economic return.

Theoretical Development

The selection of an investment portfolio by a decision maker or system manager relies on the theory of choice under uncertainty as portrayed in the axioms of the behavioral theory of expected utility and the traditional microeconomic utility-maximizing approaches (14). In theory, decision makers possess a utility function defined over possible values of aggregated expected return $[E(R_p)]$ and aggregated expected risk $[V(R_p)]$. Furthermore, decision makers are assumed to be averse to risk (i.e., to get the decision maker to take on more risk, he or she must be compensated with greater expected returns). Given this assumption, the utility function of the decision maker is an increasing function of aggregate return and a decreasing function of aggregate risk. Therefore, any indifference curve in risk-return space with respect to portfolio opportunities must be positive sloping. For given and fixed utility levels (i), therefore, a series of convex indifference curves (U^i) can be drawn to represent the trade-off between aggregate return and risk. By convention, utility increases in a direction upward and to the left (see Figure 1).

Figure 3. Equilibrium and optimality.



Portfolios can be evaluated relative to others on the basis of two parameters: aggregate expected return and aggregate risk. The key variables that give rise to the different values of the portfolio parameters are the levels of investment in each of the assets. Different investment patterns in the set of assets will yield different values of aggregate return and risk that, when plotted in risk-return space, represent the set of portfolio opportunities. By using the assumption of risk-averse decision makers, it is possible to define, via the dominance principle (no portfolio with the same or higher aggregate expected return has a lower aggregate risk), a specific subset of all feasible portfolios termed the efficient set. The optimal portfolio will be chosen from this subset. Graphically, the efficient portfolio subset will lie somewhere along the upper left boundary of the set of feasible portfolios. Thus, in risk-return space, the set of efficient portfolios traces a positively sloping concave curve (see Figure 2).

The choice situation is stated in terms of the economic concepts of equilibrium and optimality. Resolution is achieved in the normal manner by selecting a portfolio that maximizes decision-maker utility. The optimal portfolio is thus found by a tangency between the efficient-set curve and the highest indifference curve attainable, illustrated by point A in Figure 3. The concepts and graphical resolution here are similar to classic utility-maximizing approaches, except for the labeling of the variables. The fundamental difference is that an optimal portfolio is a set of investments among the assets rather than a single investment in a single asset.

The Mean-Variance Empirical Model

In practice, the parameters of interest in portfolio selection are found by the mean-variance model. Here, the weighted mean (a central tendency measure) of the individual security returns serves as a measure of the collected expected return of the portfolio, while the variance of the aggregated weighted mean indicates the riskiness of the portfolio by measuring the possible dispersion of the actual collected return.

The expected return of a portfolio $[E(R_p)]$ composed of p securities is

$$E(R_p) = \sum_{p=1}^{P} w_p \, e(r_p) \qquad (1)$$

where

$e(r_p)$ = expected return of individual security p,
w_p = proportion of total funds invested in security p, and
$\sum_{p=1}^{P} = 1$

Thus, via the statistical concept of the weighted sum of random variables, the portfolio's expected return is simply the weighted average of the expected return of its individual components. The weights are the proportions of resources allocated among the individual securities.

The variance $[V(R_p)]$ of a portfolio composed of p securities is given as

$$V(R_p) = \sum_{p=1}^{P} w_p^2 \sigma_p^2 + \sum_{p=1}^{P} \sum_{q=1}^{Q} w_p w_q r_{pq} \sigma_p \sigma_q \qquad (2)$$

where

σ_p^2 = variance of the individual securities;
σ_p, σ_q = standard deviation of the individual securities;
r_{pq} = correlation coefficient between securities p and q, $p \neq q$; and
w_p, w_q = proportion of funds invested in securities p and q.

The variance or riskiness of a portfolio is obviously a more complicated function. It depends on three elements: (a) the variance of the individual securities, (b) their correlation coefficients, and (c) the proportions invested.

The explicit consideration of individual activity variances and the calculation of total risk is the major differentiating characteristic of the portfolio approach. As can be readily observed from the specific risk equation presented, total risk is divided into two parts: riskiness of individual securities and riskiness accrued via the interrelationships among individual security variances. Analysts who use the method in financial management have focused on the second part of the risk equation (14). The effect of these interrelationships on the total riskiness of the portfolio should be intuitively clear—the higher the positive correlation between securities, the higher the variability of the portfolio return. From this mathematical certainty, analysts have developed the portfolio-diversification principle, which states that overall risk is mainly determined by these interrelationships rather than the individual variance of securities. In empirical situations, this is a matter to be determined. If the first element is the larger contributor to total risk, risk reduction is continued by further stabilizing each individual security. In the other case, risk reduction is achieved most efficiently by investing in securities whose returns are uncorrelated.

THE APPROACH AND MODEL APPLIED TO TRANSIT PERFORMANCE

The portfolio-theory approach and its operational form (the mean-variance model) serve as a vehicle to link an overall resource-effectiveness performance measure for a transit firm to an analysis of resource effectiveness at the route level. Specifically, an aggregate performance indicator is computed by considering both the route-resource-allocation package (levels of inputs or service) and the resulting economic returns to individual routes (ridership). The quintessential question is how to arrive at a particular resource-allocation package that achieves the highest expected return on the route portfolio (i.e., weighted average of individual route returns) while minimizing the level of risk.

In this paper, the perspective of optimality is replaced by a focus on a typical planning evaluation analysis. The portfolio parameters associated with a before resource-allocation package and its resulting returns are compared with an after pattern of inputs and ridership levels among routes.

The before-and-after approach to comparing portfolios can lead to four possible outcomes, only one of which is clearly better. An unambiguously better portfolio is one whose expected mean value is higher than in the former case and whose total risk is lower than in the previous case. Conversely, an unambiguously worse portfolio is one that has a smaller expected return and a larger variance than the existing resource-allocation pattern. The indeterminate cases—higher return, higher risk and lower return, lower risk—are unclear results even though they are consistent with the axioms of the underlying preference theory of the model. In these cases, additional analysis and interpretation must be brought to bear to ferret out the consequences of the resource reallocation.

An Application

An initial application of the method to transit management is developed. The central thrust herein is on forming preliminary definitions and measures for the parameters of the model and to illustrate the usefulness of the approach in transit resource-allocation planning. For demonstration purposes, several simplifications have been made in operationalizing key variables.

Monthly operating reports for 41 routes operated by the San Diego Transit Corporation were obtained for the period between July 1977 and June 1978. The reports contain information on route lengths, number of bus trips, and average passengers per trip. The before-and-after comparison of portfolio parameters is examined by partitioning the data set into two six-month time periods. This division corresponds to a major route rescheduling effort implemented in January 1978 (15).

Calculation of Variables and Parameters

The variables and parameters to be calculated are minimal in number and complexity. The major data requirement is time series data for route trips and passengers. Four variables and two parameters are to be computed. (The model was developed for U.S. customary units only; therefore SI units are not given.) The variables are

1. Proportion of resources assigned to individual routes in each time period (w_p^t, t=1,2),

Table 1. Resource allocations.

Route	Percentage of Resources		Direction of Change		
	Time Period 1	Time Period 2	Up	Down	Same
1	1.201	1.204	X		
2	2.069	2.035		X	
3	2.300	2.282		X	
4	3.923	4.099	X		
5	6.252	7.111	X		
6	3.031	3.073	X		
7	5.555	5.476		X	
8	5.806	5.249		X	
9	4.716	4.997	X		
10	0.623	0.588		X	
11	2.246	2.365	X		
12	0.255	0.272	X		
13	7.931	7.856		X	
14	0.866	0.865			X
15	5.472	5.123		X	
16	1.497	2.008	X		
17	0.097	0.115	X		
18	6.024	5.814		X	
19	2.004	1.859		X	
20	5.696	5.587		X	
21	2.419	2.275		X	
22	5.023	4.920		X	
23	0.586	0.582			X
24	4.552	4.594		X	
25	2.208	2.191		X	
26	2.515	2.483		X	
27	2.744	2.674		X	
28	1.029	1.013		X	
29	0.247	0.244			X
30	0.424	0.780	X		
31	0.271	0.267			X
32	0.243	0.240			X
33	0.248	0.244			X
34	1.353	1.163		X	
35	0.199	0.199			X
36	0.756	0.750		X	
37	0.369	0.370			X
38	2.512	2.512			X
39	1.900	1.789		X	
40	1.527	1.528			X
41	1.236	1.234			X

Table 2. Returns and standard deviations for individual routes.

Route	Time Period 1		Time Period 2	
	Estimated Return	Estimated SD	Estimated Return	Estimated SD
1	49.41	1.09	46.28	0.87
2	50.01	1.31	50.81	1.75
3	52.96	2.53	52.53	2.48
4	62.11	3.57	63.48	2.85
5	60.63	2.04	58.20	1.16
6	38.96	1.86	39.18	1.98
7	54.38	2.40	57.03	2.81
8	24.48	1.32	42.15	8.00
9	67.38	3.47	64.56	1.94
10	31.28	1.14	31.10	1.60
11	35.05	3.23	33.25	1.87
12	14.43	1.02	15.43	0.89
13	46.98	1.42	48.36	1.28
14	29.05	1.20	30.18	0.96
15	30.63	0.69	32.78	1.25
16	11.70	1.10	11.41	1.02
17	19.68	8.88	25.61	5.53
18	33.11	1.65	33.43	1.76
19	31.63	2.62	33.76	3.26
20	61.20	1.45	61.88	1.69
21	32.81	0.92	35.36	0.95
22	81.31	0.81	84.00	4.55
23	20.20	0.66	20.71	1.38
24	58.63	3.49	58.88	1.35
25	26.76	1.63	29.80	1.14
26	19.68	1.84	19.18	1.09
27	26.38	3.05	30.06	1.78
28	30.58	0.85	32.01	2.86
29	13.90	0.75	13.86	0.75
30	9.83	1.04	7.06	1.94
31	21.58	2.34	24.00	1.25
32	8.85	1.12	8.35	0.80
33	16.51	1.51	14.10	0.99
34	21.36	2.72	22.40	0.87
35	18.75	1.00	20.68	1.34
36	17.15	2.70	15.68	2.74
37	3.83	0.23	4.51	0.24
38	25.98	3.21	23.25	2.46
39	28.71	1.33	29.65	2.17
40	15.73	0.65	16.60	0.90
41	18.56	1.98	20.81	1.94

2. Estimated return in ridership for each route in each time period $[e(r_p^t),\ t=1,2]$,
3. Variance associated with each route's return $(\sigma_p^{2t},\ t=1,2)$, and
4. Covariation of route returns among all pairs of routes.

The return and risk parameters of the portfolio are defined as above in Equations 1 and 2.

The proportion of resources allocated to each route in each time period can be initially determined as follows. First, the total monthly number of trips is multiplied by the route length and added over the months in the time period, yielding total route revenue vehicle miles:

$$\mathrm{RVM}_p = \sum_{m=1}^{6} \mathrm{trips}_p \times \mathrm{length} \qquad (3)$$

Then, assuming the cost per revenue vehicle mile to be constant throughout the system, the proportion of resources allocated to each route is simply its six-month total divided by the sum of individual route totals in the time period:

$$w_p^t = \mathrm{RVM}_p^t / \mathrm{RVM}^t \quad t=1,2 \qquad (4)$$

The calculated allocation packages, defined by Equations 3 and 4 for the two time periods, are illustrated in Table 1. Visual inspection of the table shows that 11 routes received relatively more service in the second time period than in the first six months and that the relative level of service was cut on some 20 routes subsequent to the change in scheduling.

For individual routes, the expected return in each time period $[e(r_p)]$ is measured in terms of the number of passengers per run. The specific values are obtained by calculating the mean of six monthly averages for the consumption data reported in the original data:

$$e(r_p) = \left(\sum_{m=1}^{6} \mathrm{passengers/run}_p \right) / 6 \qquad (5)$$

This calculation is performed for all routes in both time periods.

The individual route risks associated with the return variable for each route is calculated as the variance about the mean found by Equation 5, for each time period. The values for individual route returns and variances are displayed in Table 2. Note that the variances are, for the most part, extremely small, which indicates temporal stability of the route return estimates.

The final variables to be calculated are the correlation coefficients between the within-period temporal pattern of returns among routes. The Pearson product moment coefficient (r), which measures the relationship between return values for routes, is computed for each pair of routes in both time periods.

Given the original data and the derived variables necessary for analysis, the portfolio parameters in each time period are calculated by employing Equations

1 and 2. The mean value of the portfolio in the first six months was 42.98 passengers/run; the associated portfolio variance was calculated to be 1.013 24. In the second time period, after the reallocation of resources, the mean value of the portfolio was 44.62 passengers/run, a 4 percent increase. However, the new variance was estimated to be 2.647 38, an increase of 161 percent. The direction of change in both of the aggregate portfolio parameters is in accordance with anticipated results, given the supposed preference for risk aversion described in the earlier theoretical development of the model structure.

The improvement in the mean return value at greater risk is an ambiguous result, as defined above. One interesting numerical result that emerges from the two sets of calculations may serve to add some judgmental evidence pertaining to the appropriateness of the second resource allocation. If we combine the return and risk portfolio parameters in each time period into an overall range (i.e., 41.97-43.99 for the first six months and 41.98-47.26 for the postshift allocation), it is straightforward to observe that the lower limits are almost exactly equivalent. The interpretation of this finding is that the reallocation of resources is favorable for the combined reasons that the mean value of the second portfolio is greater and, at worst, the second portfolio would give rise to an equvialent return under similar conditions that the initial allocation pattern would command. The advantage of the second resource allocation then is that, by concentrating resources in routes that have higher individual returns, the mean return of the portfolio is increased at little or no improvement in the worst possible risk situations.

EVALUATION OF THE METHOD

The major advantages of the portfolio-choice approach are fourfold:

1. The method provides for a fresh perspective on transit management and performance by focusing not only on returns but also on risks associated with different allocation strategies,
2. It provides a vehicle to link aggregate performance measures to detailed analyses of activities,
3. It allows the resource-allocation question to be placed in a more satisfactory conceptual framework than is possible under standard microeconomic theory, and
4. It allows the analyst and decision maker to incorporate additional information into the decision process and thus provides for better decisions.

The major disadvantage of the approach is the amount of data required to calculate the variables and parameters of the model. This problem is not as serious as might be thought after an initial reading. The cost-effectiveness approaches reported in the literature and the various costing models that are now being used demand as much if not more data. Given that data are becoming more available in disaggregate form, the application of portfolio-choice theory represents only a marginally more complicated effort.

Regarding future research needs, two areas may be identified. The first focuses on the resource-allocation packages that would be needed to achieve a target aggregate R/C ratio for the firm. Research here would be involved in finding the optimal allocation of resources to routes if the operator sets, say, a goal of achieving a 0.50 fare-box recovery ratio. Sensitivity analysis could be performed on various target ratios. Here, the analysis would focus on how the allocation package among routes changes as a function of the target ratio. The second area of future research that might provide useful insight is to examine the spatial impacts of different resource-allocation packages. Since routes, by definition, have a spatial component, it is directly possible to analyze the consequences for the area of different resource-allocation strategies. In perspective, the two research ideas outlined here can be readily incorporated into a larger strategy of evaluation including both the economic (resource allocation) and social (spatial) aspects of transit provision.

REFERENCES

1. G.J. Fielding, R.E. Glauthier, and C.A. Lave. Performance Indicators for Transit Management. Transportation, Vol. 7, 1978, pp. 365-379.
2. R.W. Schmenner. The Demand for Urban Bus Transit: A Route-by-Route Analysis. Journal of Transport Economics and Policy, Vol. 10, 1976, pp. 68-86.
3. J.H. Shortreed. Bus Transit Route Demand Model. TRB, Transportation Research Record 625, 1977, pp. 31-33.
4. The Five Year Plan: Operational, Capital and Financial Program, Fiscal Years 1978-1982. Southern California Rapid Transit District, Dec. 1976.
5. T.S. King. A Rational Approach to Planning: Tri-Met's Criteria for Service. Transit Journal, Vol. 1, 1975, pp. 23-26.
6. Massachusetts Bay Transportation Authority and Tidewater Transportation District Commission. Bus Service Evaluation Procedures: A Review. Urban Mass Transportation Administration, Rept. UMTA-MA-09-7001-79-1, March 1979.
7. R.E. Glauthier and J.N. Feren. Evaluating Individual Route Performance. Transit Journal, Vol. 5, 1979, pp. 9-26.
8. Arthur Anderson and Company. Bus Route Costing for Planning Purposes. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1974.
9. W. Cherwony. Cost Centers: A New Approach to Transit Performance. Transit Journal, Vol. 3, 1977, pp. 70-80.
10. K.E. Nelson and W.C. Nevel. Cost-Effectiveness Analysis of Public Transit Systems. Traffic Quarterly, Vol. 33, 1979, pp. 241-252.
11. J. Fried. Forecasting and Probability Distributions for Models of Portfolio Selection. Journal of Finance, Vol. 25, 1970, pp. 539-555.
12. H.M. Markowitz. Portfolio Selection. Journal of Finance, Vol. 7, 1952, pp. 77-91.
13. W.F. Sharpe. Portfolio Theory and Capital Markets. McGraw Hill, New York, 1970.
14. E.F. Fama and M.H. Miller. The Theory of Finance. Holt, Rinehart and Winston, Chicago, 1972.
15. Five-Year Plan Update. San Diego Transit Corporation, San Diego, March 1978.

Publication of this paper sponsored by Committee on Transit Service Characteristics.

Evaluating Potential Effectiveness of Headway Control Strategies for Transit Systems

Mark A. Turnquist and Steven W. Blume

Holding strategies for control of headways between transit vehicles are often considered as a means of improving the reliability of transit service. This paper describes simple tests that can be used to identify situations for which control is potentially attractive. These tests depend only on a simple measure of headway variability and the proportion of total passengers who will be delayed as a result of the holding strategy. Thus, this analysis provides transit operators with a simple screening model to evaluate potential effectiveness of controls.

Headway control has been proposed as one way to improve the reliability of transit service. By reliability we mean the ability of transit to adhere to schedule or to maintain regular headways and a consistent travel time. This ability is important to both the transit user and the transit operator. To the user, nonadherence to schedule results in increased wait time, makes transferring more difficult, and creates uncertainty about arrival time at the destination. To the operator, unreliability results in less effective utilization of equipment and personnel and reflects itself in reduced productivity and increased cost in the system's operations.

A study of the potential effectiveness of various strategies for control of unreliability in transit services is thus a vital element in the search for ways to improve transit productivity and efficiency. Such control strategies have important implications for both planning and management of transit systems. Control strategies may be divided into two basic groups: planning and real time. In general, the distinction is that planning strategies involve changes of a persistent nature. Examples include restructuring of routes and schedules, changes in the number and location of stops, or provision of exclusive rights-of-way. On the other hand, real-time control measures are designed to act quickly to remedy specific problems. These actions have immediate effects but seldom exert any influence on the general nature of operations over a longer time period.

Several real-time strategies for correcting service disruptions have been discussed in the literature. A good summary of the state of current knowledge in this area has been provided by Abkowitz and others (1). One commonly considered control strategy is the holding of selected vehicles at control points along a route to regularize headways between successive vehicles. That is, a vehicle that arrives at the control stop too close behind the preceding vehicle would be deliberately delayed to make the headway between these vehicles more nearly equal to the scheduled headway.

The major incentive for making headways more regular is to reduce waiting time of passengers who board at or beyond the control point. If passengers arrive at a stop without regard to the schedule of service (i.e., randomly), a well-known formula [see Welding (2)] gives the average wait time as

$$E(W) = [E(H)/2] + [V(H)/2E(H)] \tag{1}$$

where

$E(W)$ = average wait time,
$E(H)$ = average headway between vehicles, and
$V(H)$ = variance of headway.

Thus, making headways more regular (i.e., reducing the variance) serves to reduce average wait.

On the other hand, the major costs of such a policy are borne by passengers who are already on the vehicle, since they are delayed when the bus is held up. Thus, the implementation of a holding control strategy involves making some passengers better off at the expense of others. At a minimum, if control is to be effective, it must reduce aggregate waiting time by more than it increases aggregate in-vehicle time (possibly allowing for some differential weighting of these two elements of total trip time).

The purpose of this paper is to provide some basic rules of thumb to indicate the conditions under which a holding strategy might be effective. By implication, we also wish to describe those situations in which such a strategy is not likely to be effective. These rules of thumb are based on relatively modest data requirements about the route and, hence, should be useful in making basic planning decisions about whether or not to implement such a control strategy on a given route.

PREVIOUS ANALYSIS

An article by Barnett (3) has provided several important ideas for the work contained here. He formulated a model based on a simple discrete approximation to the probability distribution of vehicle arrival times at bus stops. Based on this simple model, an optimal holding strategy can be derived to minimize the total delay to all passengers who use the route. The resulting strategy depends on (a) the mean and variance (or standard deviation) of the headway distribution, (b) the ratio of average vehicle load at the control point to average number of boarding passengers at subsequent stops, and (c) the correlation between successive vehicle arrival times at the control stop. This last information is a measure of the degree of bunching or pairing of vehicles on the route: A route on which vehicles have bunched in pairs would have a large negative correlation between successive headways because a very short one (between two paired vehicles) will be followed by a very long one (between bunches). Statistical estimation of this correlation is difficult, however, because of the small sample sizes available and the notorious unreliability of the estimators of covariance.

The objectives of this paper are to analyze holding strategies by using a more general probability model of vehicle arrival times at the control stop and to shed some additional light on the question, Under what conditions is control likely to be of value? Specifically, we wish to allow a transit operator to address this question without detailed knowledge of the covariances between successive vehicle arrival times at stops, as this information is seldom available.

Our approach is to use a general model of the probability distribution of headways between successive vehicles and then examine two simple cases that provide approximate upper and lower bounds on the potential benefits of a holding strategy. By doing this, basic conclusions can be reached regarding situations in which control is likely to be beneficial and those in which it is not.

We will examine a holding strategy that holds each early vehicle (i.e., each vehicle preceded by a short headway) until the headway preceding it reaches a minimum allowable value (h_{min}). The structure of the analysis is to find the value of h_{min} that minimizes total delay to passengers (including both wait time and in-vehicle delay). This optimal value of h_{min} will be denoted h^*_{min}. Once h^*_{min} is found, those situations for which control is advantageous can be identified.

UPPER BOUND ON EFFECTIVENESS OF HOLDING

Control of headways will make the greatest reduction in total delay when headways alternate (i.e., short, long, short, long). This happens on routes where vehicles are influenced substantially by the operation of the vehicle in front of them. For example, this would tend to be the case where loading delays are relatively more important than traffic congestion in determining overall vehicle operating speed. Routes in which pairing is prevalent would be of this type. In such a situation, holding a vehicle to lengthen a short headway also serves to reduce the long one that follows. Thus, the variance of headways is reduced by a greater amount for a given delay to the held vehicle than if short headways might be followed by another short headway.

The extreme case is when the observed sequence of headways alternates between two discrete values. In this case, the sum of any two consecutive headways is a constant. That is, if one headway is 2 min too short, the next one must be 2 min too long. By the same argument, if the second headway is 2 min too long, the third must be 2 min too short, and so on. In a statistical sense, successive headways are perfectly correlated, so that knowledge of one headway implies knowledge of the entire set. For this case, headway control will have maximum benefits.

If we denote the scheduled headway by $\bar{H}$ and the magnitude of the deviation by x, the marginal probability density function for headways before control is given by p(H):

$$p(H) = \begin{cases} 0.5 & H = \bar{H} - x \quad (2a) \\ 0.5 & H = \bar{H} + x \quad (2b) \end{cases}$$

For the probability distribution of headways described by Equations 2a and 2b, the expected headway is $\bar{H}$ and the variance is x^2. The control action lengthens the short headways to a value $h_{min} = \bar{H} - \rho x$, where $0 \leq \rho \leq 1$.

We will define an optimal holding strategy to be one that minimizes total delay to passengers. Total delay is expressed as

$$T = \gamma E(D) + (1 - \gamma) E(W) \qquad (3)$$

where

T = total delay to all passengers,
E(D) = expected delay to passengers already on board the vehicle,
E(W) = expected wait time for passengers arriving at or beyond the control stop, and
γ = weighting constant to reflect the relative number of passengers already on board to those waiting to board at subsequent stops.

The expected delay to passengers already on the vehicle is simply the average length of time a vehicle will be held. If we assume that passengers arrive at stops at random times, Equation 1 can be used to determine expected wait time.

A holding strategy that minimizes T will be defined by the value h^*_{min}. Because $h_{min} = \bar{H} - \rho x$, we can find h^*_{min} by finding the optimal value of ρ. Note that after control the headway distribution is given by

$$p'(H') = \begin{cases} 0.5 & H' = \bar{H} - \rho x \quad (4a) \\ 0.5 & H' = \bar{H} + \rho x \quad (4b) \end{cases}$$

This distribution has expected value still equal to $\bar{H}$, but has variance $\rho^2 x^2$. This reduces wait time for passengers yet to board to

$$E(W') = (\bar{H}/2) + (\rho^2 x^2 / 2\bar{H}) \qquad (5)$$

The delay to passengers already on the vehicle is equal to $(1 - \rho)x$ if the vehicle is held. Since the probability of a short headway is 0.5, the expected in-vehicle delay is

$$E(D) = 0.5(1 - \rho)x \qquad (6)$$

By substituting Equations 5 and 6 into Equation 3, we obtain total expected delay as

$$T = 0.5\gamma(1 - \rho)x + (1 - \gamma)[(\bar{H}/2) + (\rho^2 x^2 / 2\bar{H})] \qquad (7)$$

To find the optimal value of ρ, we can differentiate the expression for T with respect to ρ, and set the result equal to zero.

$$dT/d\rho = -0.5\gamma x + (1 - \gamma)\rho\,(x^2/\bar{H}) = 0 \qquad (8)$$

This implies an optimal value for ρ:

$$\rho = 0.5\gamma x \bar{H} / (1 - \gamma)\, x^2 = [0.5\gamma/(1 - \gamma)]\ (\bar{H}/x) \qquad (9)$$

The resulting value for h^*_{min} is then

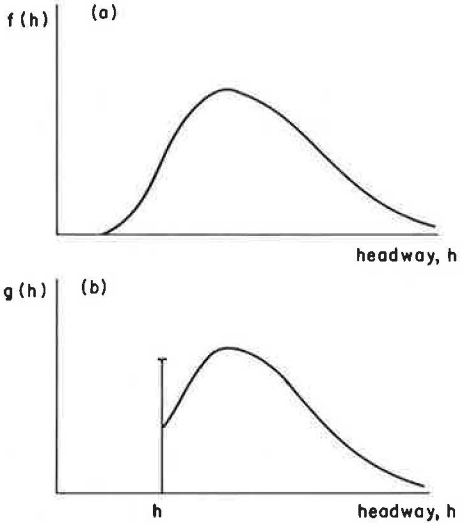
$$h^*_{min} = [(1 - 1.5\gamma)/1 - \gamma]\ \bar{H} \qquad (10)$$

For control to be effective, we must have $h^*_{min} > \bar{H} - x$; that is, the optimal minimum headway after control must be greater than the short headways before control, or it does not pay to control at all. This means that we must have $\rho < 1$, which implies that we must satisfy the condition $x/\bar{H} > 0.5\gamma/1 - \gamma$. However, recall that the variance of the headway distribution before control was x^2. Thus, the quantity $x/\bar{H}$ is simply the coefficient of variation (standard deviation divided by mean) of the headway distribution. Thus, for control to be effective, the coefficient of variation of the headway distribution must exceed $0.5\gamma/1 - \gamma$. If it does not, the optimal value of ρ is 1, which implies no control.

This condition, then, provides a simple test for potential effectiveness of a control policy. It is based on two simple pieces of information: (a) the coefficient of variation in the headway distribution and (b) the relative proportion of riders who are already on board the vehicle to those who are yet to board at subsequent stops.

It must be kept in mind that this condition is derived for the best possible case (i.e., when successive headways are perfectly correlated). Thus, if the condition is not met, we can be confident that control will not be effective. However, we must look more closely at situations for which the condition is met because the actual

Figure 1. Headway distribution (a) before control and (b) after control.



situation may be less favorable to control than is reflected in this model.

LOWER BOUND ON THE EFFECTIVENESS OF HOLDING

In order to establish a lower bound on the effectiveness of holding, we will examine the opposite extreme case, which corresponds to the situtation in which headways between successive vehicles are statistically independent. This means that knowledge that a given headway is short gives us no additional information about the probable values for the next headway. Such a situation would arise, for example, when traffic conditions have a much greater effect on vehicle operations than does the loading time at stops. In this case control will be less effective because we have no guarantee that by lengthening a short headway we are also reducing a long headway. We might be simply reducing another, already short, headway. This case of independent headways thus provides a lower bound on the effectiveness of control strategies, which will allow us to further refine our evaluation of situations likely to be favorable for control.

We assume that the distribution of headways before control is applied is described by a cumulative distribution function [F(h)] with a density function [f(h)]. The effect of the control strategy is to make all headways less than some value (h_{min}) equal to that value. The distribution of headways before and after control is shown in Figure 1. There is a nonzero probability that the headway will take on the discrete value h_{min}, and for values of $h > h_{min}$, there is a continuous density function.

The expression for the distribution of headways after control is applied can be derived by considering a sequence of two successive headways after control, which we will denote H'_{i-1} and H'_i. The probability that $H'_i \leq h$ depends on both the headways H_{i-1} and H_i before control of vehicle i (if any), as well as the value of the minimum allowable headway (h_{min}). On one hand, $H'_i \leq h$ if $H_{i-1} > h_{min}$ (and thus not changed by the control strategy) and $H_i \leq h$. If $H_{i-1} \leq h_{min}$, it becomes $H'_{i-1} = h_{min}$ (after control), and the ith headway is shortened. In this case $H'_i \leq h$ if the sum of H_{i-1} and H_i before control of i was less than $h_{min} + h$. Of course, because the control policy enforces a minimum headway, the probability is that $H'_i \leq h$ will be 0 for $h < h_{min}$. These statements can be summarized in the form of a cumulative probability distribution function [G(h)] as shown in Equation 11:

$$G(h) = P(H'_i \leq h) = \begin{cases} 0 & h < h_{min} \\ P(H_{i-1} > h_{min})P(H_i \leq h) \\ \quad + P(H_{i-1} \leq h_{min}) \\ \quad \times P(H_{i-1} + H_i \leq h + h_{min} | H_{i-1} \leq h_{min}) & h \geq h_{min} \end{cases} \tag{11}$$

where P(·) denotes the probability of the event described by (·). From the distribution function in Equation 11, we can obtain (at least in theory) the probability density function [g(h)] for $h \geq h_{min}$, shown in Figure 1b, by differentiating with respect to h. The probability that $h = h_{min}$ is given by $G(h_{min})$.

As in the previous case of perfectly correlated headways, our analysis proceeds by solving for the optimal value of h_{min} and then using this to describe the conditions for which control is potentially beneficial. The process of finding the optimal h_{min} involves trading off reductions in wait time (due to reduced headway variance) against in-vehicle delays due to holding of vehicles.

The variance of headways after control can be written as shown in Equation 12:

$$V(H') = [h_{min} - E(H')]^2 G(h_{min}) + \int_{h_{min}}^{\infty} [h - E(H')]^2 g(h)dh \tag{12}$$

The rate of change of this variance with changes in h_{min} is

$$(d/dh_{min}) V(H') = 2G(h_{min})[h_{min} - E(H')] [1 - (d/dh_{min})E(H')] \tag{13}$$

For relatively small values of h_{min}, we can argue (to a first-order approximation) that changes in h_{min} will not affect the mean headway. Thus, $(d/dh_{min})E(H') \simeq 0$. While this approximation is not strictly accurate, a good case can be made that an operator is unlikely to implement a control policy that increases mean headway significantly. This would have negative impacts on vehicle productivity and also on passenger wait and travel time. Thus, the magnitude of control delays applied is likely to be small, and hence the approximation is a reasonable one. For small values of control delay, we can also approximate $G(h_{min})$ by $F(h_{min})$. These two approximations allow us to obtain the result in Equation 14:

$$\begin{aligned}(d/dh_{min})E(W') &= (d/dh_{min}) \{[E(H')/2] + [V(H')/2E(H')]\} \\ &\simeq (h_{min} - \overline{H}/\overline{H}) F(h_{min})\end{aligned} \tag{14}$$

This provides the ability to evaluate (approximately) the marginal rate of reduction in waiting time as the minimum allowable headway increases. Total delay (T) will be minimized when the marginal rate of reduction in waiting time is just equal to the marginal rate of increase in in-vehicle delay.

The delay incurred by passengers already on a vehicle that is held is given by:

$$D = \begin{cases} h_{min} - H & H < h_{min} \quad (15a) \\ 0 & H \geq h_{min} \quad (15b) \end{cases}$$

From this, we can derive the expected delay, as shown in Equation 16:

$$E(D) = \int_0^{h_{min}} (h_{min} - h) f(h)dh = h_{min} F(h_{min}) - \int_0^{h_{min}} hf(h)dh \tag{16}$$

Figure 2. Areas of potential usefulness for headway control.

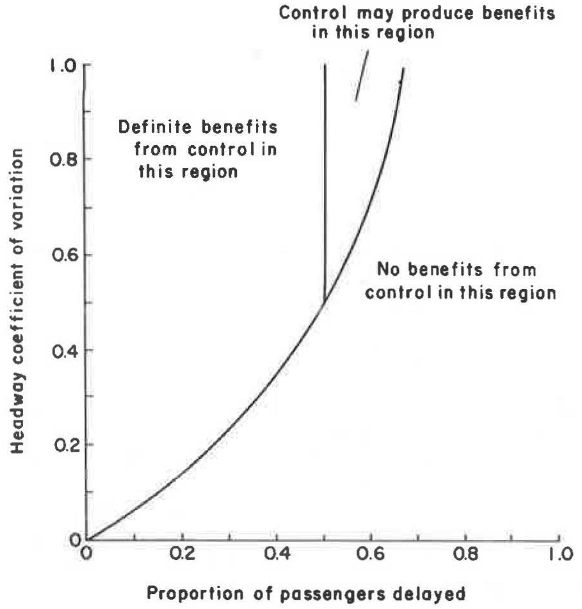
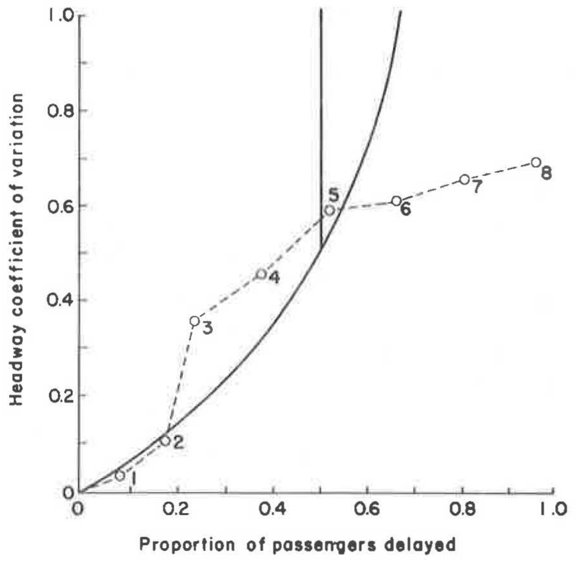


Figure 3. Trajectory of stops along a bus route.



The marginal change in expected delay is then

$$(d/dh_{min})\,E(D) = F(h_{min}) + h_{min}\,f(h_{min}) - h_{min}\,f(h_{min}) = F(h_{min}) \tag{17}$$

By using the expressions in Equations 14 and 17 we can then solve for an optimal value of h_{min} by setting $dT/dh_{min} = 0$.

$$dT/dh_{min} = \gamma(d/dh_{min})\,E(D) + (1-\gamma)\,(d/dh_{min})\,E(W') = 0$$

$$\Rightarrow \quad \gamma F(h_{min}) + (1-\gamma)[(h_{min} - \overline{H})/\overline{H}]\,F(h_{min}) = 0$$

$$\Rightarrow \quad h_{min} = [(1-2\gamma)/(1-\gamma)]\,\overline{H} \tag{18}$$

Since $\gamma > 0$, $h_{min} \leq \overline{H}$. Because we must also have $h_{min} > 0$, this solution is only valid if $\gamma < 0.5$. We can summarize this as shown in Equations 19a and 19b, which give the expression for the optimum value of h_{min}, denoted h^{**}_{min}.

$$h^{**}_{min} = \begin{cases} [(1-2\gamma)/(1-\gamma)]\,\overline{H} & 0 \leqslant \gamma < 0.5 \quad (19a) \\ 0 & 0.5 \leqslant \gamma \leqslant 1 \quad (19b) \end{cases}$$

The condition for which nonzero holding is beneficial ($0 \leq \gamma < 0.5$) provides important information on the situations in which holding strategies are potentially useful, in the same way that the condition $x/\overline{H} > 0.5\gamma/1 - \gamma$ from Equation 9 did.

IMPLICATIONS OF THE ANALYSIS

These two pieces of information can be combined, as illustrated in Figure 2, to yield a convenient representation of situations for which headway control is likely to produce benefits for passengers—those for which it is unlikely to be worthwhile and those for which more careful analysis is required. By analyzing the two extreme cases of independent headways and perfectly correlated headways in detail, we can bound the regions of effectiveness for a class of headway control strategies, as shown in Figure 2. For situations in which control produces benefits under both extremes, we can be fairly confident that it will be beneficial. On the other hand, there are situations in which control does not appear to be desirable under the best of circumstances; hence, control in these situations is unlikely to be useful. There remains one reasonably small region in which control would probably produce benefits on routes where vehicles are substantially influenced by the vehicles in front of them but not on routes where vehicles move relatively independently of one another. For situations in this region, more detailed and specialized analysis is required.

A major implication of the result shown in Figure 2 is that it is wise to control a route at a point where relatively few people are on the vehicle and relatively many are waiting to board at subsequent stops, in order that the value of γ be small. Generally, this means that the control point should be located as early along the vehicle's route as possible. However, reliability problems worsen as one proceeds along a route. If dispatching at the route origin is effective, the headways will be reasonably regular at the early stops along the route, which implies that the coefficient of variation will be small. At stops further along the route, however, the coefficient of variation in headways will tend to be larger. Thus, the decision of whether or not to implement a control strategy is tied to identification of a logical control point along the route.

Each stop along a route will have a particular headway distribution (with implied coefficient of variation) and value of γ associated with it. Thus, each stop could be plotted as a point in the space defined by these two variables, as shown in Figure 3. Then, by looking at the trajectory of the route relative to the boundary values, the transit operator can make a decision about whether or not to control the route and, if so, where. For example, for the route illustrated by Figure 3, control at stop 3 might be worthwhile, but at stop 8 it is unlikely to be beneficial.

It is also illuminating to examine the form of the optimal holding policy for the two extreme cases analyzed here, as illustrated by Equations 10 and 19. Note first that, in both cases, the magnitude of the optimal minimum headway is dependent on the scheduled average headway, but not the variability of headways. Thus, determining a policy on minimum headways to be enforced for each extreme case is quite simple and requires very little data and only simple analysis.

Second, note that the optimal minimum headway is always smaller if successive headways are independent than if they are negatively correlated. This follows logically from the fact that a given amount of delay is less beneficial when headways are independent. Thus, we would expect the optimal delay to be smaller.

Clearly, as previously demonstrated by Barnett (2), precise setting of an optimal strategy for a given situation requires knowledge of the covariance between successive headways. However, our analysis, based on much more general models of headway distributions than he used, indicates that the range of possible values is not large, at least for small values of γ (for which control is likely to be most beneficial).

The models described here make several simplifying assumptions in order to make the analysis relatively tractable. For this reason they should be viewed primarily as screening models, whose purpose is to identify situations in which decisions are relatively clear-cut and to distinguish those situations for which further analysis is likely to be required. While measures of benefits (reduction in total delay) can be derived from the models presented here, those estimates are likely to be less useful than the identification of regions of potential benefits because the models omit several important factors. More detailed simulation studies of selected situations have been reported by Bly and Jackson (4), Koffman (5), and Turnquist and Bowman (6). Such models incorporate considerably more detail about specific routes and can be used to estimate actual benefits from control much more precisely.

SUMMARY AND CONCLUSIONS

The major point of the analysis in this paper is that basic and important decisions regarding headway control can often be made by using only limited statistics about system operation. The essential data on which fundamental decisions can be based are the coefficient of variation in the headway distribution and the relative proportions of passengers who are on board the vehicle (and will be delayed) and passengers who have yet to board (and will benefit from reduced wait time). By using this rudimentary information, a transit operator can make preliminary decisions regarding whether or not headway control is likely to produce benefits or whether further analysis is required.

The models are based on the assumption that passengers arrive randomly through time at bus stops; therefore, use of the results should be limited to situations for which that is likely to be true. In most cases, this means that average headways should be 10 min or less. For routes on which average headways are longer than 10 min, an analysis that includes a more sophisticated representation of passenger arrivals is necessary. Examples of such passenger arrival models are discussed by Jolliffe and Hutchinson (7), Turnquist (8), and Turnquist and Bowman (6).

This analysis has had nothing to say regarding the costs of implementing the control strategy. Our objective has been simply to illustrate situations in which headway control is likely to produce positive benefits. Of course, the decision to implement such a control system would involve evaluation of the costs as well as the benefits. We have described a rather general concept of a headway control strategy. Details of the implementation of such a strategy are likely to vary greatly from property to property, and hence the cost of implementation is likely to vary greatly as well. The transit operator may be able to generate relatively good cost estimates for a particular system but is likely to be much more uncertain regarding the potential benefits of the controls. The analysis in this paper should provide useful information in that regard.

ACKNOWLEDGMENT

This work was supported in part by a contract from the Office of University Research of the U.S. Department of Transportation.

REFERENCES

1. M. Abkowitz, H. Slavin, R. Waksman, L. Englisher, and N. Wilson. Transit Service Reliability. Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, Rept. UMTA-MA-06-0049-78-1, Dec. 1978.
2. P. I. Welding. The Instability of Close Interval Service. Operational Research Quarterly, Vol. 8, No. 3, 1957, pp. 133-148.
3. A. Barnett. On Controlling Randomness in Transit Operations. Transportation Science, Vol. 8, No. 2, 1974, pp. 102-116.
4. P. H. Bly and R. L. Jackson. Evaluation of Bus Control Strategies by Simulation. Department of the Environment, Transport and Road Research Laboratory, Crowthorne, Great Britain, Rept. 637, 1974.
5. D. Koffman. A Simulation Study of Alternative Real-Time Bus Headway Control Strategies. TRB, Transportation Research Record 663, 1978, pp. 41-46.
6. M. A. Turnquist and L. A. Bowman. Control of Service Reliability in Transit Networks. Office of University Research, U.S. Department of Transportation, Rept. DOT-RSPA-DPB-50-79-5, March 1979.
7. J. K. Jolliffe and T. P. Hutchinson. A Behavioral Explanation of the Association Between Bus and Passenger Arrivals at a Bus Stop. Transportation Science, Vol. 9, No. 4, 1975, pp. 248-282.
8. M. A. Turnquist. A Model for Investigating the Effects of Service Frequency and Reliability on Bus Passenger Waiting Time. TRB, Transportation Research Record 663, 1978, pp. 70-73.

Publication of this paper sponsored by Committee on Transit Service Characteristics.

What Public Transportation Management Should Know About Possible User Reactions, as Shown by the Example of Price Sensitivity

Werner Brög and Otto G. Förg

This paper shows that the decrease in demand for urban public transportation if fares are increased can only be predicted accurately by studying the users of public transportation and the possible ways they might react to fare increases. In order to do this, the paper categorizes the users of public transportation according to factors that influence their demand for public transportation. Those persons who have alternative modes of transportation available to them are identified and divided into groups that are sensitive or not sensitive to price increases. These different factors make it possible to estimate the decrease in demand for public transportation if fares are increased. Furthermore, the paper stresses the importance of two groups of persons who do not directly cause a decrease in demand for public transportation. The first group are those persons who have no alternative to the use of public transportation but would be severely hurt economically by fare increases. The second group are those persons who would complain about price increases but would continue to use public transportation. The former group gives useful information on the public's ability to pay, and the latter group indicates the probable public reaction to a fare increase.

It is no longer possible for public transportation to be totally self-supporting. This means that decisions concerning the price structure of such transportation systems must be determined according to norms other than those of the market economy. Since the urban public transportation system is, by definition, a system designed for transportation of the masses, the price for this form of transportation should necessarily lie within the range of the greatest possible profit, on the one hand, and the ability of the users of the transportation system to pay, on the other hand.

Due to inflation, however, the urban public transportation system will be forced to increase its rates periodically. However, herein lies the danger (in contrast to many products of the free-market economy) that the demand for transportation will be reduced by two important types of urban public transit system patrons:

1. Those patrons who are financially incapable of paying higher fares (i.e., those persons for whom the public transit system has a social responsibility to provide transportation) and
2. Those persons who are convinced that the urban public transit system is the sensible alternative to the use of private transportation.

The latter group of persons can only be induced to use the public transportation system by making this system an attractive alternative. Since this group of persons chooses its mode of transportation after rational consideration of the advantages of various alternatives, substantially increased fares could cause members of this group to decide not to use the urban public transportation system.

If one wishes to determine how a sensible compromise can be made, one must do empirical studies. A direct causal relation between price of transportation and demand for transportation cannot be determined, since the price of transportation is only one of the factors one takes into consideration when choosing a mode of transportation. For instance, if one has only a single means of transportation available, one must use this whether or not one approves of the price for the fare (and other aspects of the transportation system).

This means that, if one wishes to do relevant research on reduced use of public transportation that results from fare rises, study of current and potential patrons of the urban public transit system is necessary. Those characteristics of the patrons that determine their price elasticity must be filtered out from those characteristics that only help to describe the individual personalities.

USERS OF THE URBAN PUBLIC TRANSPORTATION SYSTEM

The patron of the urban public transportation system does not act purely rationally according to financial capacity and deduce a logical mode of transportation accordingly. Therefore, studies that restrict themselves only to the economic aspects of the situation are as incomplete as all research methods that attempt to explain complex interrelationships one-dimensionally.

In order to analyze the use patterns of the patrons of the urban public transportation system, one needs to use a method that can identify individual causal relationships in their entire complexity. Only such a construction would make it possible to determine the relevance of transportation within the individual framework of decision making and to depict the dependence of the whole on this one dimension. Then one can predict with a high degree of accuracy how overall demand would be affected by changing one dimension.

Explaining Transit Behavior

An approach was developed that views the actual transit behavior of the individuals studied as primary. Types of routes traveled and completion of routes traveled are the initial building blocks that allow one to become familiar with the character of the patron of the urban public transportation system. According to reason for traveling (daily or occasional trips), day traveled (weekday or weekend), and destination (inner city or suburbs) one can differentiate between totally different types of urban public transit system patrons.

For those routes traveled, one must determine precisely why the urban public transit system was used. The broad spectrum of individual reasons can be summarized in the following analytical questions:

1. Is there another vehicle present that might be used for the trip?
2. Are there certain constraints that make use of an alternative mode of transportation impossible?

3. Does the individual take price of fare into consideration when choosing a mode of transportation?
4. How expensive does the individual consider the alternative mode of transportation to be?
5. Is the individual willing to use the alternative?

The answers to these questions show a pattern of individual situations that define the actual latitude that individuals have in choosing alternative modes of transportation. The reason for the individual choice of mode of transportation for the routes traveled with the urban public transit system is depicted in Figure 1. This figure uses negative selection to depict the factors that are momentarily relevant for decision making. One can only see the entire pattern of the situation for those for whom the last dimension (which explains the choice of mode of transportation) is given. This is sufficient to deal with the problem posed here, for increases in rates only affect the actions of those who actually have an alternative to the urban public transportation system. If one wishes to view potential changes caused by other factors, one must expand on the dimensions of this summary in order to be able to determine in how many cases other factors would affect the situation as a whole, since other constrictions limit one's freedom to choose alternative modes of transportation.

But we would like to emphasize another important point: This graph shows the individual's freedom to choose an alternative for one specific day only. One cannot make any statements that have general validity

Figure 1. Model of behavior on day of random sampling.

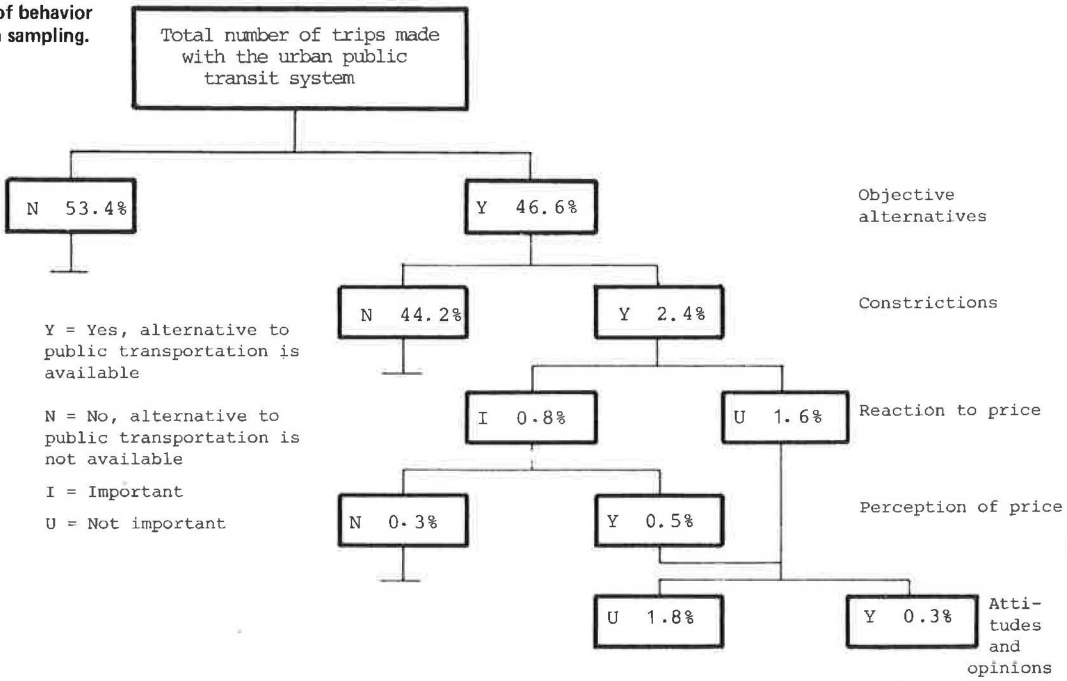
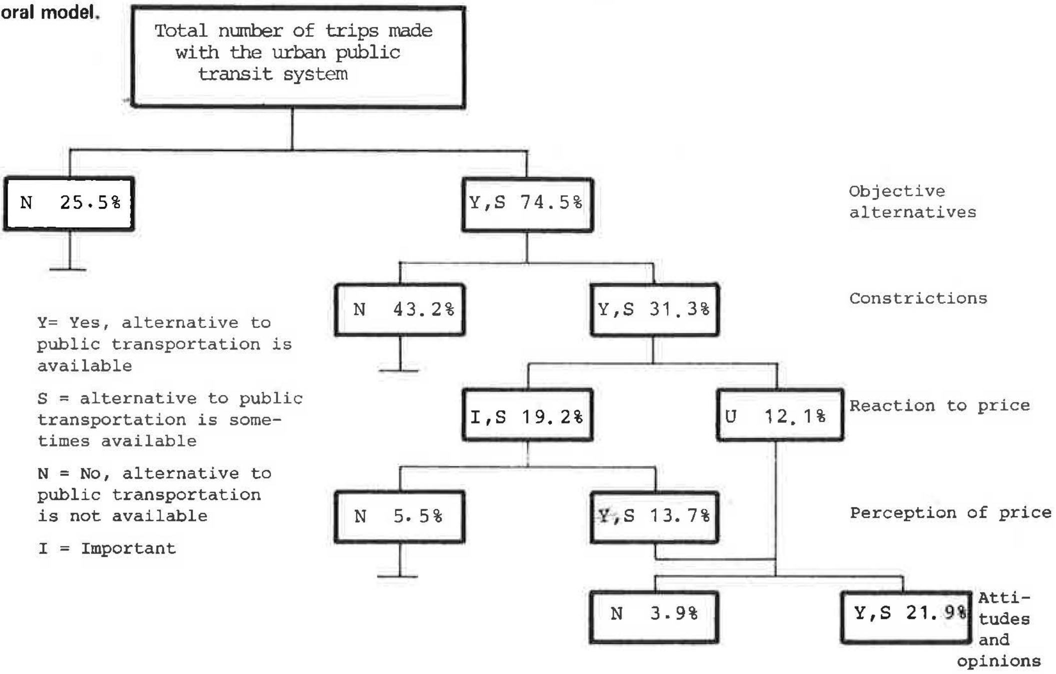


Figure 2. Behavioral model.



when discussing the immensely smaller proportion of persons who have no alternative to the use of the urban public transportation system, since the potential behavior of patrons is determined by changing constrictions. One might, for example, be able to use a car one day and not the next because the car is being repaired. A household can organize the activities of its members according to different modes. If one takes into consideration that individual freedom of action is a sensitive construct that can easily change, the quantitative distribution found in Figure 2 results.

Analysis of Price Elasticity

Figure 2 shows which patrons of the urban public transportation system have the available alternative of using a different mode of transportation if they desire to do so because of increased fares.

But the patron has another possible way of reacting to fare increases. He or she can refuse to use the urban public transportation system or attempt to reduce use of public transportation. However, these options are only open to those patrons who are not traveling for work or educational purposes because it is not a feasible alternative for them. Thus, in order to determine actual behavior related to transportation, exact knowledge of activities of urban public transit systems must be the cornerstone for all reflections. If one relates price elasticity of the urban public transit system patrons to their individual circumstances, one can deduce the decline in public transportation use that can be expected if fares are raised by differing amounts. This can be seen in Table 1.

We would like to emphasize the importance of the following. Price elasticity cannot be universally expressed by means of a coefficient. The curve of reduced trips is obviously not linear. Rather, it is characterized by sensitivity thresholds. This insight is not surprising; all products are characterized by price thresholds. This is also the reason why purely commercial enterprises generally tend to set prices just below the upper threshold (1).

According to the type of patron, use of mode of transportation is changed or use of public transit systems is decreased. Both of these two ways of reacting are frequent. But one must also consider the fact that reduced use of public transportation is an act of protest and, after a short period of adjustment to the new price, use tends to reach the old level.

Patrons can compensate for fare increases (at least partially) by choosing a more economical type of tariff in the future. This method of reaction can only be identified if the individual patron is the pivotal focus of the study.

These results can be differentiated according to different types of user groups, according to the following (see Table 2):

1. Type of ticket used,
2. Mode of transportation used, and
3. Place of residence of the target person.

Those patrons who buy passes for a specified period of time react most quickly to fare rises since they are forced to pay the highest absolute price increase. The upper limit of the reactions of this group, however, is under that of those who pay for each ticket separately (for the fare rises that were studied).

Persons who buy single tickets respond to fare increases most vehemently if rates are increased by more than 20 percent, since these persons are exposed to the absolute increase in price each time they pay for a ticket. Those persons who use the various special fare rates respond to increases less than average.

The price sensitivity of the users of the more attractive subway system is less than average, especially if these rates are increased by less than 20 percent, reactions to the increase are minimal. The persons who react most strongly to a price increase are those persons who use streetcars.

Table 1. Expected change in transit behavior.

	Expected Decreased Use of Public Transportation (%)			No Expected Decreased Use of Public Transportation (%)		
Price Rise	By Changing	By Reducing	Total	Without Changing Ticket	With Change of Ticket	Total
1-10 percent	0.0-0.7	0.0-0.8	0.0-1.5	98.2-100.0	0.0-0.3	98.5-100.0
11-15 percent	0.7-5.3	0.8-2.6	1.5-7.9	91.6-98.2	0.3-0.4	92.0-98.5
16-20 percent	5.3-6.2	2.6-2.8	7.9-9.0	89.3-91.6	0.4-1.6	90.9-92.0
21-25 percent	6.2-8.2	2.8-5.3	9.0-13.5	84.5-89.3	1.6-2.0	86.5-90.9
26-30 percent	8.2-8.3	5.3-6.1	13.5-14.4	83.7-84.5	1.9-1.9	85.6-86.5

Note: Percentages are based on 2409 trips made with the urban public transit system.

Table 2. Percentage of trips for which change in transit behavior is expected.

	Use Expected to Decrease If Fares Are Increased (%)								
	1-10 Percent			11-15 Percent			16-20 Percent		
Item	Change of Mode	Decreased Use	Total	Change of Mode	Decreased Use	Total	Change of Mode	Decreased Use	Total
Type of ticket									
Single ticket (n = 710)	0.0-1.1	0.0-1.9	0.0-3.0	1.1-3.3	1.9-6.5	3.0-9.8	3.3-6.2	6.5-6.7	9.8-
General pass (n = 810)	0.0-1.2	0.0-0.4	0.0-1.6	1.2-12.2	0.4-0.8	1.6-13.0	12.2-12.5	0.9-3.5	13.0-
Student fare (n = 175)	0.0-0.0	0.0-0.0	0.0-0.0	0.0-2.9	0.0-0.0	0.0-2.9	2.9-2.9	0.0-0.0	2.9-
Senior citizen's pass (n = 286)	0.0-0.0	0.0-0.3	0.0-0.3	0.0-0.0	0.3-3.6	0.3-3.6	0.0-0.0	3.6-3.6	3.6-
Other (n = 428)	0.0-0.0	0.0-0.0	0.0-0.0	0.0-0.0	0.0-0.2	0.0-0.2	0.0-0.0	0.2-0.4	0.2-
Total (n = 2409)	0.0-0.7	0.0-0.8	0.0-1.5	0.7-5.3	0.8-2.6	1.5-7.9	5.3-6.2	2.6-2.8	7.9-
Mode of transportation									
Bus (n = 374)	0.0-0.4	0.0-0.0	0.0-0.4	0.4-3.0	0.0-0.4	0.4-3.4	3.0-3.4	0.4-0.8	3.4-
Streetcar (n = 616)	0.0-2.1	0.0-0.0	0.0-2.1	2.1-17.1	0.0-2.3	2.1-19.4	17.1-17.5	2.3-2.4	19.4-
Subway (n = 386)	0.0-0.7	0.0-0.9	0.0-1.6	0.7-0.9	0.9-4.2	1.6-5.1	0.9-2.0	4.2-4.4	5.1-
Train (n = 1033)	0.0-0.1	0.0-1.5	0.0-1.6	0.1-0.7	1.5-3.1	1.6-3.8	0.7-2.2	3.1-3.1	3.8-
Total (n = 2409)	0.0-0.7	0.0-0.8	0.0-1.5	0.7-5.3	0.8-2.6	1.5-7.9	5.3-6.2	2.6-2.8	7.9-
Place of residence									
City	0.0-1.1	0.0-0.3	0.0-1.4	1.1-7.7	0.3-2.3	1.4-10.0	7.7-8.0	2.3-2.3	10.0-
Outlying region	0.0-0.0	0.0-1.7	0.0-1.7	0.0-0.3	1.7-3.6	1.7-3.9	0.3-2.5	3.6-3.8	3.9-
Total	0.0-0.7	0.0-0.8	0.0-1.5	0.7-5.3	0.8-2.6	1.5-7.9	5.3-6.2	2.6-2.8	7.9-

Within the city, more patrons cease to use the public transportation system than is the case in the more outlying regions.

OTHER ASPECTS OF FARE INCREASES RELATED TO PATRON USE

The effect of increased fares is not restricted simply to the percentage of persons who would cease to use the urban public transportation system if fares were raised. Two other factors must be taken into consideration if one wishes to study this phenomenon comprehensively.

Increased fares may not only induce patrons to turn to other means of transportation, but they might also cause considerable ill will in the population. This can result in a poor climate of public opinion that, although it may not initially affect the use of the public transportation system, might result in a movement that would cause those who thought that the increases were necessary to change their minds. Indirectly, this phenomenon could increase the number of persons who ultimately stop using the system of public transportation. In order to prevent this, a study of price elasticity must take these factors into account in order to determine the types of questions for which answers need to be sought. This would make it possible for the results of the studies to be used as a basis for advising public relations personnel of the urban public transportation system.

The analysis of public opinion led to the results given below.

Proposed Fare Increase	Persons Likely to Comment Negatively on Increased Fares (% of trips)
1-10 percent	0.0-5.3
11-20 percent	5.3-32.1
21-30 percent	32.1-43.5

The percentage of persons who only verbally disapprove of fare increases is considerably higher than the percentage of those who would actually stop using public transportation if prices were increased. The social and economic effects of fare increases on patrons are also important. That is, prior to each fare increase, one must determine whether the patrons are capable of paying the increased fares. The statistics below show that different percentages in fare increases cause a serious number of hardship cases.

Proposed Fare Increase	Persons for Whom Increases in Fares Represent a Real Financial Hardship (% of trips)
1-10 percent	0.0-1.8
11-20 percent	1.8-3.7
21-30 percent	3.7-7.5

Each urban public transportation system patron must be analyzed in regard to these three important aspects of fare rises. Table 3 shows the comparison of the actual tendency to cease use of the urban public transit system, public opinion, and real hardship on an intrapersonal basis. The contradictions become obvious in the following comparisons:

1. Those patrons who stop using the urban public transportation system influence the general climate of public opinion; however, only a minority of these persons are actually confronted by economic hardships due to increased fares;
2. Of those persons who are economically hurt by the fare rises, only the minority stop using public transportation; these patrons tend to be less angry about fare rises than average; and
3. Those persons who express strong opposition to an increase in fares tend to stop using public transportation more than average; however, generally their protest remains verbal.

But the influence of these persons on those who initially approve of fare rises should not be underestimated. The hardship that these persons must bear is minimal.

Patrons Who Cease to Use the Public Transportation System

It is necessary for an urban public transportation system that is planning to increase its fares to know which groups of persons would be likely to stop using public transportation after fares are raised. Therefore, we analyzed the latter group of persons according to a variety of characteristics and according to different percentages of increase (see Table 4). It is obvious that those persons who would stop using public transportation belong to those social classes that are better off. This is especially problematical because it is precisely this group that is of such great potential importance to the future of the urban public transportation system.

Expected to Decrease If Fares Are Increased (%)						Use Is Not Expected to Decrease (%)		
25 Percent			26-30 Percent					
nge of e	Decreased Use	Total	Change of Mode	Decreased Use	Total	Without Change in Type of Ticket	With Change in Type of Ticket	Total
-11.3	6.7-10.7	12.9-22.0	6.2-11.3	6.7-11.3	22.0-22.6	76.3-76.9	1.1-1.1	77.4-78.0
-13.2	0.9-3.5	13.4-16.7	13.2-13.2	3.5-5.6	16.7-18.8	80.6-82.4	0.6-0.9	81.2-83.3
-5.7	0.0-2.3	2.9-8.0	5.7-7.0	2.3-2.3	8.0-9.3	89.6-90.9	1.1-1.1	90.7-92.0
-0.0	3.6-6.2	3.6-6.2	0.0-0.0	6.2-6.4	6.2-6.4	83.5-83.5	10.2-10.2	93.7-93.7
-0.0	0.4-0.4	0.4-0.4	0.0-0.0	0.4-0.4	0.4-0.4	99.6-99.6	0.0-0.0	99.6-99.6
-8.2	2.8-5.3	9.0-13.5	8.2-8.3	5.3-6.1	13.5-14.4	83.7-84.5	1.9-2.0	85.6-86.5
-9.7	0.8-1.1	4.2-10.8	9.7-9.7	1.1-1.1	10.8-10.8	87.6-87.6	1.6-1.6	89.2-89.2
-18.4	2.4-6.1	19.9-24.5	18.4-18.4	6.1-6.1	24.5-24.5	75.2-75.2	0.3-0.3	75.5-75.5
-4.4	4.4-7.3	6.4-11.7	4.4-5.0	7.3-7.3	11.7-12.3	87.4-88.0	0.3-0.3	87.7-88.3
-3.1	3.1-5.7	5.3-8.8	3.1-3.1	5.7-7.8	8.8-10.9	85.5-87.3	3.6-3.9	89.1-91.2
-8.2	2.8-5.3	9.0-13.5	8.2-8.3	5.3-6.1	13.5-14.4	83.7-84.5	1.9-2.0	85.6-86.5
-10.3	2.3-6.1	10.3-16.4	10.3-10.5	6.1-6.3	16.4-16.8	82.3-82.6	0.9-1.0	83.2-83.6
-3.8	3.8-3.8	6.3-7.6	3.8-3.9	3.8-5.9	7.6-9.8	86.4-88.6	3.8-3.9	90.2-92.5
-8.2	2.8-5.3	9.0-13.5	8.2-8.3	5.3-6.1	13.5-14.4	83.7-84.5	1.9-2.0	85.6-86.5

Table 3. Consequences of a fare increase.

	Sensitivity (%)				Hardship Cases (%)				Atmospheric Reaction (%)				
	Reaction Possible If Fares Are Increased By			Reaction Not Expected	Affected If Fares Are Increased By			Not Affected	Reaction If Fares Are Increased By			No Reaction	Total (%)
Outcome	1-10% (n = 36)	11-20% (n = 218)	21-30% (n = 349)	(n = 2060)	1-10% (n = 38)	11-20% (n = 76)	21-30% (n = 156)	(n = 1928)	1-10% (n = 111)	11-20% (n = 679)	21-30% (n = 919)	(n = 1196)	(n = 2409)
Sensitivity													
Reaction possible	100	100	100			1	13	12	18	30	36	0	14
Reaction not expected				100	100	99	87	88	82	70	64	100	86
Hardship cases													
Affected		1	9	7	100	100	100			1	5	8	7
Not affected	100	99	91	93				100	100	99	95	92	93
Atmospheric reaction													
Rejection	77	99	99	33		13	34	44	100	100	100		43
Acceptance	23	1	1	67	100	87	66	56				100	57

Table 4. Characteristics of those who stop using public transit due to a fare increase.

		Trip Reductions If Fares are Increased By				Change of Mode If Fares Are Increased By				Reduced Use If Fares Are Increased By			
Characteristic Analyzed	Average	11-15% (n = 193)	16-20% (n = 218)	21-25% (n = 327)	26-30% (n = 348)	11-15% (n = 127)	16-20% (n = 150)	21-25% (n = 196)	26-30% (n = 200)	11-15% (n = 66)	16-20% (n = 68)	21-25% (n = 131)	26-30% (n = 148)
Sex													
Male	46 percent	89.1	87.0	89.1	91.3	97.8	95.7	95.7	95.7	69.6	69.6	78.3	87.0
Female	54 percent	109.3	111.1	109.3	107.4	101.9	103.7	103.7	103.7	125.9	125.9	118.5	111.1
Average age	42 years	107.1	104.8	102.4	100.0	100.0	97.6	100.0	95.2	119.0	119.0	119.0	111.9
Education													
Grammar school	43 percent	81.4	76.7	74.4	72.1	107.0	93.0	76.7	76.7	23.3	25.6	72.1	58.1
High school	37 percent	127.0	124.3	110.8	118.9	116.2	113.5	105.4	108.1	151.4	151.4	116.2	118.9
College	20 percent	90.0	110.0	135.0	130.0	55.0	90.0	135.0	140.0	170.0	165.0	130.0	100.0
Occupation													
Blue-collar worker	6 percent	83.3	83.3	116.7	116.7	66.7	66.7	83.3	66.7	83.3	83.3	183.3	183.3
White-collar worker	31 percent	141.9	135.5	112.9	112.9	167.7	158.1	132.3	129.0	67.7	71.0	74.2	80.6
Professional	6 percent		16.7	16.7	16.7		16.7	16.7	16.7				
Independently employed	3 percent	800.0	800.0	500.0	500.0	1100.0	1033.3	766.7	766.7				
Not employed	54 percent	50.0	51.9	77.8	79.6	20.4	27.8	55.6	57.4	137.0	135.2	122.2	118.5
Average size of household	2.95 persons	84.4	84.4	86.4	89.8	77.6	78.0	83.1	83.4	98.0	98.3	91.5	98.0
Average household income before taxes	2570 DM	98.9	99.6	97.6	98.2	95.0	96.8	93.7	93.7	106.6	105.8	103.5	103.9
Average personal income of each person before taxes	871 DM	117.2	118.1	113.0	109.3	122.4	124.2	112.9	112.4	108.8	107.7	113.1	106.1
Average monthly budget for public urban transportation	61 DM	88.5	100.0	101.6	101.6	96.7	109.8	113.1	113.1	75.4	77.0	83.6	85.2
Place of residence													
Munich	68 percent	125.0	114.7	122.1	116.2	144.1	127.9	125.0	126.5	85.3	85.3	114.7	102.9
Outlying areas	32 percent	46.9	68.8	53.1	65.6	6.3	40.6	46.9	43.8	131.3	131.3	68.8	93.8

Note: 1 DM = $0.56.

What This Teaches Us

A number of basic insights have been gained by this study that one should take into consideration for future research projects on similar themes. The analysis of price elasticity has to take the specific factors for each application of a special case into consideration. This includes the following:

1. Price of fares prior to the fare rise and
2. Time lapse between the last fare rise and the incipient fare rise (i.e., the relationship of the fare increase to the general rate of inflation within a certain time period).

The analysis of price elasticity must be made by using an appropriate research concept. This research concept must be able to measure the highly complex reactions of the patrons. Studies that deal only with verbalized options and attitudes can only measure public opinion but not actual decline in demand for public transportation. On the other hand, studies that deal only with the social structure can analyze this dimension but not behavioral reactions, opinions, and attitudes. An adequate research concept must take all of these three dimensions into account on an individual basis.

The patron of the urban public transit system must be the pivotal point in any analysis of price elasticity. Studies must become familiar with the contradictions inherent in the individual. One must realize that each individual has his or her own subjective perspective from which to view concrete facts; therefore, he or she acts subjectively.

PERCEPTIONS OF URBAN PUBLIC TRANSPORTATION SYSTEMS

The following discussion deals with the question of how drivers who use individual means of transportation, but who have the option of using public transportation to commute to work, perceive the urban public transportation system.

The way persons perceive the alternatives offered them by the public transportation system consists of various elements, such as the following:

1. Routes available,
2. Time needed to travel a certain stretch,

3. Frequency of public transportation,
4. How long one must wait,
5. Price,
6. Comfort,
7. Safety, and
8. Atmosphere.

It is apparent that, if one is not informed about some of these characteristics, one cannot make use of the urban public transportation system. If one is uninformed on some of the other points, one forms a negative opinion of the system of public transportation. This may result in rejection of use of public transportation by the potential patron.

Our institute has come to certain conclusions regarding the above points in various studies that we have made. We would like to comment here on some of the results that we think are relevant.

In a study for the Federal Ministry of Transport on the alternative use of the urban public transportation system for persons who use private means of transportation to commute to work, we discovered that 25 percent of the 165 persons questioned were not familiar with the current public transportation routes for this stretch. Forty percent of the persons studied knew nothing of the modes of the public transportation system as represented by fare price. Since many of those persons who drive cars could not comment on either of these characteristics, 45 percent of the whole cannot be viewed as urban public transportation system patrons due to the fact that they were simply not informed.

Response	Percent
Not informed about the urban public transportation system	
Not familiar with the routes	6
Not aware of the price	20
Knowledge of neither price nor routes	19
Total	45
Informed on the most important aspects of the urban public transportation system alternatives	55

The same persons were also asked to guess how long it would take them to commute to work if they used the urban public transportation system. These guesses were then analyzed for accuracy by comparing the time given with city maps and public transportation schedules. The result was that the persons guessed, on the average, that commuting would take more than 25 percent longer with public transportation than it actually takes.

Trip Time by Public Transportation	Time (min)
Estimated by those who drive cars	84
Actual time needed for trips	66

It is important to know how these drivers assess the time needed for the different phases of commuting to work with their individual means of transportation and with the alternative public transportation.

These persons believe that when they use public transportation, they spend 75 percent of the time needed to reach their destination in waiting and getting to and from transportation to destination. These same persons assume that, if they use their own cars, they spend only 15 percent of the time needed to arrive at their final destination on activities that are not directly related to driving. When one takes these facts into consideration, it seems to be a bit odd that these persons assume that the actual travel time with public transportation is only half as long as the time they need to reach their destination by car.

	Time (min)	
Estimates of Time Required	With Car	With Public Transportation
For walking to or from parking place or place where public transportation stops	6	28
For looking for a parking place or total time spent in boarding public transportation and changing routes	1	34
Actual travel	41	22
Total	48	84

The results of the above study were confirmed in another study done for the Munich Integrated Transport System in which persons who use individual modes of transportation were asked about the concrete alternatives they had to reach their destination by using public transportation. The percentage by which the 938 persons in this study exaggerated the time they thought it would take to travel by public transportation was almost exactly the same as in the previous study. Therefore, it is possible to view the results of the studies as definitive. These persons also believe that public transportation is considerably more expensive than it is. The same study, on the other hand, showed that this group of persons believed their personal mode of transportation to cost less than it actually does.

Item	Percent
Incorrect estimate of time needed to travel with public transportation	
Relative	28.9
Absolute	36.5
Incorrect estimate of cost of using public transportation	
Relative	12.4
Absolute	22.2

The final figures that we would like to quote from the Munich study clearly show a direct relationship between how well one is informed about different modes of transportation and which mode of transportation one chooses. Persons who actually use the urban public transportation system are much more accurate at estimating how long it takes to travel a certain distance, and they can, with accuracy, state the price for public transportation. On the other hand, it has been proved that those persons who drive their own cars and, thus, do not use public transportation are poorly informed about public transportation.

Comparison	Persons Who Drive Cars (n = 938) (%)	Persons Who Use Urban Public Transit System (n = 2409) (%)
Rate of error in estimating length of time needed to travel with public transportation	28.9	10.4
Rate of error in estimating the price of traveling with public transportation	12.4	0.2

REFERENCE

1. Factors Affecting Public Transport Patronage: The Proceedings of a Symposium Held by the Transport and Road Research Laboratory, Livingston, 14 June 1977. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, TRRL Supplementary Rept. 413, 1978, 65 pp.

Publication of this paper sponsored by Committee on Transit Service Characteristics.

Abridgment

Use of Federal Section 15 Data in Transit Performance Evaluation: Michigan Program

James M. Holec, Jr., Dianne S. Schwager, and Angel Fandialan

In the first application of its kind, the reporting system of Section 15 of the Urban Mass Transportation Act, as amended, is being used to support the development of a straightforward, routine, and comprehensive transit performance evaluation program in the state of Michigan. The methodology developed for Michigan satisfies the complementary needs to account for public funds invested in transit operation and development and to promote the efficient and effective use of these funds in the delivery of transit services. At the same time, the methodology avoids placing an additional burden of record keeping and reporting on individual transit operators. In the rapidly developing field of transit performance evaluation, these features are essential for state and local funding agencies to consider as part of any plans to develop a continuing evaluation program. In this paper, the Michigan program is described, and the features of the program that have general applicability for other areas concerned with transit performance measurement and evaluation are highlighted.

In October 1978, the Michigan State Legislature enacted a law that requires the state transportation commission to report annually on the efficiency and effectiveness of all publicly funded transportation programs within Michigan and to describe the progress of these programs in carrying out plans approved in the preceding year.

The act stipulates that the annual reports on transportation programs be based on information included on forms authorized by the U.S. Department of Transportation. This requirement is intended to ensure that, to the maximum extent possible, existing data will be used and transportation programs will not be burdened by additional reporting requirements. As a result, the reporting system of Section 15 of the Urban Mass Transportation Act of 1964, as amended, was selected as the primary data base for the routine evaluations of transit performance.

As a result of these requirements, a study was initiated to develop and test a performance evaluation methodology for midsized transit systems in Michigan. This paper describes efforts by Michigan to increase the accountability of transit systems and to promote increased efficiency and effectiveness of transit management and operation.

MICHIGAN EVALUATION METHODOLOGY

The evaluation methodology developed for the midsized transit systems in Michigan includes two phases: (a) a diagnostic review of selected indicators of transit efficiency, effectiveness, and other related measures of performance and (b) a detailed evaluation of transit performance in areas suggested for further investigation by the diagnostic review. Figure 1 outlines the evaluation process and suggested use of evaluation results.

The intent of both the diagnostic and detailed phases of the evaluation methodology is to systematically and routinely review transit performance to

1. Increase the understanding of transit operations and performance in Michigan by the state department of transportation, the state legislature, and other interested groups;
2. Facilitate the exchange of information among transit properties, particularly in areas where there are innovative operations and exemplary performance;
3. Monitor the use of public funds for public transit service development and delivery; and
4. Identify opportunities to improve transit performance by promoting more efficient and effective transit services.

Diagnostic Review Element

The first phase of the evaluation methodology involves a diagnostic review of selected performance indicators. The diagnostic review includes two elements: (a) a peer comparison in which the performance of all midsized Michigan transit systems is compared and (b) a time-series assessment in which the performance of each system is assessed against itself over time and the change in performance of other midsized transit systems in the state. The primary objective of the diagnostic review is to identify the performance indicators that merit more detailed examination during the detailed evaluation.

Although there is considerable discussion within the transit industry about the uniqueness of transit systems and, therefore, the limitations of peer comparisons, an important premise of the evaluation methodology is that comparisons need not be avoided but instead should be conducted carefully. The use of a peer comparison does not overlook the differences among transit systems; rather, peer comparisons are intended to assist Michigan in identifying and understanding the differences among the midsized transit systems in the state.

Time-series assessments of transit performance facilitate assessment of performance of a transit system over time in relation to change in (a) operating policy, (b) investment and expansion plans, (c) the economy, (d) characteristics of the community, and (e) the transit system's own performance objectives.

Information from the diagnostic peer comparisons and time-series assessments allows Michigan to obtain an increased understanding about transit operations and identify areas of transit performance that merit more detailed evaluation.

Performance Indicators Used

Considerable research is under way to identify and define indicators of transit performance and assess their merit. Too often, however, evaluation methodologies suffer from the use of more indicators than are appropriate for a routine review of performance and do not organize the indicators in a structured evaluation approach. The indicators developed to support the diagnostic review element of the Michigan performance evaluation methodology avoid these pitfalls by focusing attention on performance measures that relate to and build on each other, allowing the assessment of important components of transit efficiency and effectiveness.

The indicators selected for evaluating midsized transit systems in Michigan have the following characteristics:

1. Provide information on the efficiency, effectiveness, and other performance characteristics of a transit system;
2. Provide information on each midsized transit system by total system, mode, functional area, and object expense class;
3. Are not redundant (i.e., each indicator provides an important new element of information relevant to obtaining a complete understanding of transit performance);
4. Relate to and build on each other, thus allowing the assessment of important components of major indicators; and
5. Are developed from information that is routinely collected and reported by transit operators either to meet Section 15 reporting requirements or to complete the Michigan Department of Transportation grant application for capital and operating assistance for transit services.

The indicator structure for vehicle operating efficiency is shown in Figure 2 to illustrate these characteristics.

When using performance indicators to evaluate a transit system, it is important to remember that indicators provide limited information about performance and should be used carefully. If the value of a performance indicator is above or below an acceptable level, further examination is required to determine whether a problem exists, what the characteristics and impacts might be, and what remedies can be tested as a cure.

Detailed Evaluation Element

The second phase of the evaluation methodology involves the detailed evaluation of transit performance, focusing on those indicators identified in the diagnostic review phase of the evaluation process. The detailed evaluation primarily involves (a) preparing for a site visit with each transit system, (b) investigating the factors that affect transit performance through site visits, and (c) documenting the evaluation findings. A detailed evaluation must be conducted to develop informed conclusions about transit performance.

The objectives of the detailed evaluation phase of the methodology are the following:

1. To gather and report information to explain the factors that affect transit performance;
2. To identify examples of innovative performance that can be shared with the other transit operators;
3. To identify opportunities for improvement in transit performance that can be implemented by the transit system with assistance from the state and federal governments, as appropriate; and
4. To monitor changes in transit performance over

Figure 1. Overview of the evaluation process.

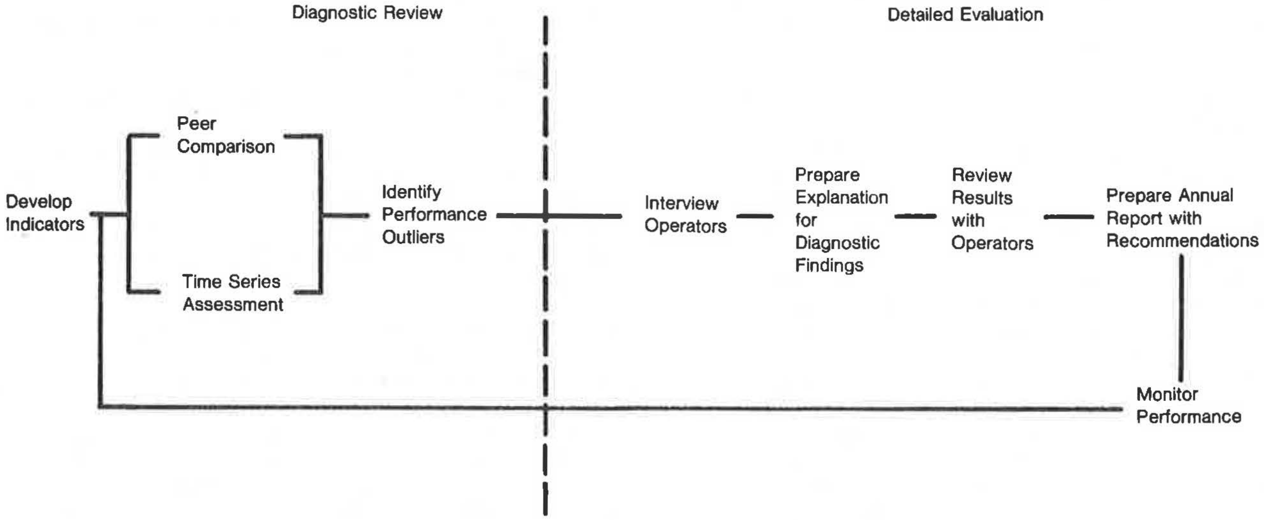
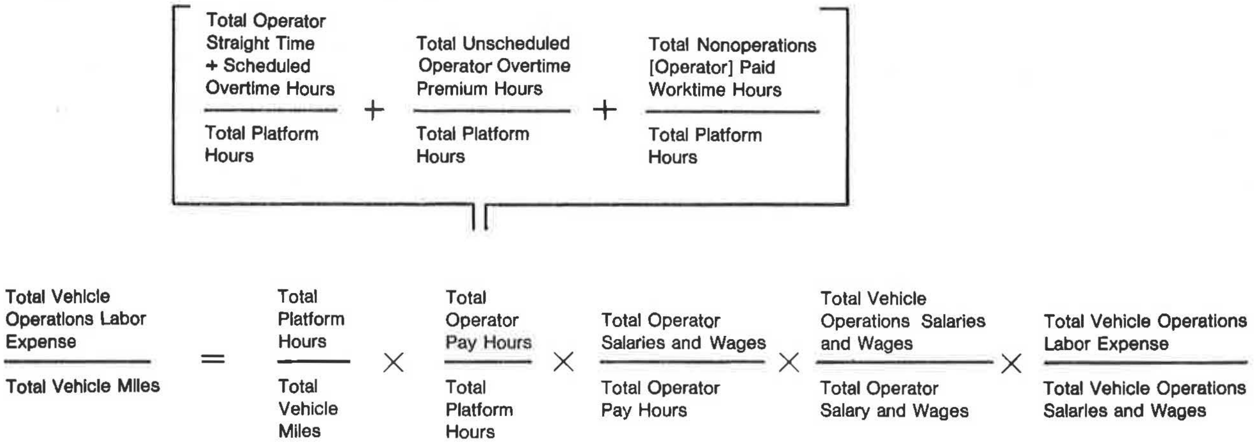


Figure 2. Illustrative indicator structure: vehicle operations labor expense per vehicle kilometer.



time, particularly efforts to improve transit system efficiency and effectiveness.

A performance indicator should be investigated in detail if its value differs significantly from an acceptable standard value. The standard value for the indicator should reflect the performance of a transit system that is similar in size and operating characteristics but performs optimally.

Without specific standards, decision rules need to be developed to identify indicators that deserve further examination. Typical decision rules include the following:

1. The value of the indicator for the current year is significantly above or below the average indicator value for midsized transit systems;
2. The indicator is different from the statewide average in both value and rate of change; and
3. The value of an indicator is significantly higher or lower than the value for the transit system the previous year.

The type of decision rules established will determine the number of indicators examined in depth and will, therefore, influence the resources required to conduct the detailed phase of the evaluation.

Much of the information for the detailed evaluation of a transit system is gathered and validated during site visits with the transit operator. To prepare for the site visits to a transit system, the evaluator conducts a structured review of all of the performance indicators identified for detailed review within each transit system and prepares questions and issues for discussion about these indicators.

The evaluator begins the site visit with an interview of the transit system's general manager. The general manager may initiate detailed discussions about the transit system's performance and the specific indicators or may suggest that the evaluator speak to the assistant general manager or other transit system employees in each functional area. During the site visit, the evaluator may meet with the head of vehicle operations, the maintenance supervisor, and members of the administrative departments, particularly a staff accountant or the transit system's auditor.

Following each interview, the evaluator summarizes general findings as well as findings specific to each indicator. Gaps in information or understanding are noted along with apparent differences in opinion. Follow-up interviews may be necessary to obtain additional information, verify findings, and clarify unresolved issues. Before the site visit is complete, the evaluator briefs the transit system's general manager and seeks comments and insights about the findings and preliminary recommendations.

The primary product of the detailed evaluation phase of the methodology is the documentation of the evaluation results. This report discusses current performance and identifies opportunities for improvement, as appropriate. The evaluation report should serve as a useful management tool for the transit system. Transit system officials should have the opportunity to review, comment, and rebut the evaluation findings before the report is made final or reviewed by outside interests.

The documentation of the detailed evaluation findings should fulfill the following requirements:

1. Briefly describe the transit system;
2. Identify all the performance indicators identified for detailed review in the diagnostic phase of the evaluation and provide a concise explanation of the factors that influence the value and rate of change over time for each of these indicators;
3. Discuss examples of innovative system characteristics that may be shared among the operators;
4. Identify opportunities for improvement and suggest actions for the transit system and state and federal governments, as appropriate; and
5. Report on progress toward improving the efficiency and effectiveness of the system during the past year.

GENERAL APPLICABILITY OF THE MICHIGAN METHODOLOGY

The performance evaluation methodology discussed in this paper was developed for midsized transit systems in Michigan. The methodology, however, is generally applicable for transit systems and for state and local agencies concerned with the ongoing measurement of transit performance. The methodology is attractive because it

1. Is relatively simple, straightforward, and can be routinely applied;
2. Uses readily available data and, therefore, does not place considerable additional reporting requirements on the transit operator;
3. Enhances understanding about transit operations and performance;
4. Facilitates communication between the transit operator and local governing bodies, funding authorities, the legislature, and the public; and
5. Produces information that can lead to improvements in the efficiency and effectiveness of transit performance.

The availability of Section 15 data facilitates the conduct of transit performance evaluation by use of indicators presented in this paper for the diagnostic review element of the methodology.

Although many features of the methodology described have general applicability, transit systems, local governments, states, or other concerns that are considering the development of a transit performance evaluation program should address the following considerations: (a) the audience for the evaluation results, (b) the purpose or use of evaluation findings, (c) the level of detail of the analyses, (d) the frequency of evaluation, and (e) the availability of resources to conduct the evaluations. After these issues are addressed, the evaluation methodology can be tailored to best meet the needs of the participants within the context of each unique state and local environment.

Publication of this paper sponsored by Committee on Transit Service Characteristics.

Systematic Procedure for Analysis of Bus Garage Locations

Frank Spielberg and Marvin Golenberg

The overhead costs of transit operations represent one area in which economies can be achieved. For a large system the costs of putting buses on routes and pulling them off (pull-on and pull-off) and driver relief can be substantial: up to 10 percent of the operating budget for the system studied. These costs are directly related to the route structure and the location and capacity of bus garages. This paper describes a procedure that uses generally available planning data in the analysis of the pull-on and pull-off and relief costs for alternative garage programs. Factors studied include the number of facilities, their location, their capacity, and the routes served from each garage. It is shown that the location of garages in relation to day-base routes is a determinant of relief costs and that the difference in operating costs for alternative programs can approach $1 million per year.

Operation of a transit system involves direct costs associated with the provision of revenue service and overhead costs associated with system management. Revenue service operations have been studied extensively. Transit operators and planning agencies devote continuous effort to determination of how to serve the current and potential transit market with the proper allocation of bus kilometers and bus hours.

Overhead costs, which consist of management, maintenance, planning, marketing, research, fringe benefits, and nonrevenue vehicle operation, are a daily concern of transit operators but have not been studied by the planning profession. A recent study by SG Associates, Inc., addressed two elements of overhead costs as part of an analysis of the efficiency of alternative garage size and location alternatives:

1. The costs of putting buses on routes and pulling them off (pull-on and pull-off costs) and
2. The costs of driver relief.

In the course of this study a procedure was developed for the planning analysis of alternative garage locations. This paper describes the procedure.

Overhead costs represent a significant portion of transit operating budgets. Many elements can be addressed through management practice (e.g., administration, maintenance, and policy) and can be quickly modified if found to be inefficient. Other costs are mandated by union contracts (e.g., reporting time, wash-up time, meal breaks, fringe benefits, percentage of straight runs, spread time penalties, and overtime payments). Management has less flexibility in changing these provisions but can attempt modification, if appropriate, each time a new contract is negotiated. Two cost elements, however, are a direct function of garage location—pull-on and pull-off and driver-relief costs. If a major facility is poorly located with respect to the current or future route structure, management may have to accept added costs for many years.

Relative to the total operating budget, the pull-on and pull-off and relief costs represent, for the system studied, roughly 10 percent of the annual operating budget. In absolute terms, they were just under $32 000/day. The difference among alternative garage size and location strategies was about 10 percent in pull-on and pull-off and relief costs—about 1 percent of the operating budget but an amount of more than $1 million/year. Given the size of the potential saving, the investment in location studies can yield a high rate of return for any transit operator who is considering a new facility.

Throughout the discussion that follows it must be remembered that the costs involved represent only one element in the evaluation of garage location. Other elements include availability and cost of land, surrounding development, access streets, and environmental impacts. All factors must be weighed in the selection of the location for a new facility.

COST ELEMENTS

Pull-on and pull-off costs are defined to include the vehicle kilometers and driver hours from the time the bus leaves the garage until it enters revenue service and the same items from the time the bus leaves revenue service until it returns to the garage. A simplification introduced in the analysis method is the assumption that a vehicle will enter or leave revenue service only at a line terminal. Some properties will carry passengers on trips to and from a garage; however, such trips are often lightly patronized as they do not serve corridors of demand and can result in longer or slower travel. Pull-on and pull-off costs are a subset of deadhead costs, which may also include operations required by interlining.

Driver relief, for this analysis, is defined as the number of pay hours required for a driver to travel from the garage to the relief point. The actual method of computing these pay hours will vary with the labor agreement of each property. For the system for which the method was developed, the pay hours are based on the time required for the driver to travel from the sign-in location (the garage) to the relief point by use of the transit system. Thus, relief hours were a function not only of garage location but also of the route structure and the operating headways at the time the relief was scheduled (typically the day base).

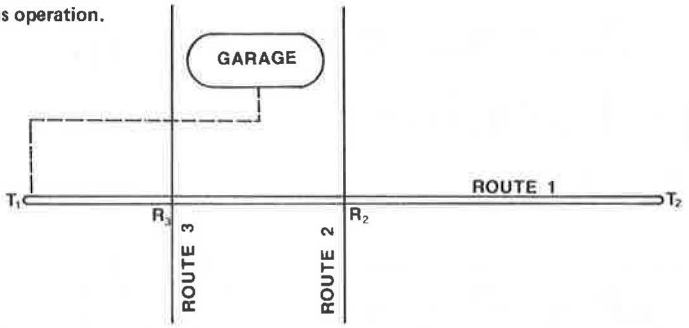
In many instances, relief costs, which were more than one-half of the amount of the pull-on and pull-off costs for all routes other than those that operate a peak-only service, were quite high because the only base-day service that operated near the garage offered infrequent service or because a transfer to a long headway route was required.

Only those relief costs that varied with garage location were considered. Invariant costs (e.g., time spent waiting for the assigned vehicle, check-in, and wash-up) are not included.

The basic structure of the problem is represented in Figure 1. Figure 1 illustrates, in simplified form, the activity of a bus during the day.

The bus leaves the garage in the morning and proceeds to one route terminus (in this illustration, T_1). The distance traveled from the garage to the point where the bus enters revenue service is the pull-on deadhead distance. For a large system that has extensive suburban routes, such as that studied, pull-on distances can exceed 25 km (15 miles) and require up to 1 h of travel

Figure 1. Simplified illustration of daily bus operation.



time. The vehicle operates on the route, as scheduled, until it is time to return to the garage. The return to the garage is the pull-off deadhead distance. Each vehicle operated on a line will have at least one pull-on and one pull-off per day. Trippers, operated only as peak service, will pull-on and pull-off twice each day. In practice, a vehicle will often enter or leave revenue service from a point other than a terminal. For general planning purposes, this can be ignored, as the productivity for such trips is typically low. A vehicle in revenue service from a garage to a suburban park-and-ride lot at 6:30 a.m. may not officially be deadheading, but the effect is the same.

In actual operations the bus may not operate the same route all day; it may be interlined. Such scheduling can increase the efficiency of operations, but it introduces a complexity that cannot be treated at this level of analysis. Further, since interlining operations will vary from schedule to schedule, consideration of such service options is not warranted when determining the sites for storage and maintenance facilities.

One factor that can be accounted for is the maximum distance a bus can travel before it must return to the garage to refuel—about 320 km (200 miles). If the operating characteristics of the line are such that the first vehicles out in the morning will exceed this service limit, then either additional vehicles must be scheduled (with additional pull-on and pull-off costs) or evening peak vehicles, if available, are used for evening service. This latter option is more efficient.

Deadheading operations, such as pull-on and pull-off movements, cost roughly the same per hour and per kilometer as revenue operations yet produce no return. For the system studied these costs were 15.09 cents/km (24.14 cents/mile) and $10.06/h in 1977. They are an essential operating element but must be considered as system overhead to be reduced whenever possible.

A bus can operate all day, subject to fuel limitations, but drivers must be relieved at times specified by work rules, typically after 4-4.5 h. Many contracts specify that such relief must occur at locations that afford the driver a rest room and a place to eat. Further, since the driver will check in at a garage (typically the same garage from which the line is operated), relief points must be accessible by transit from the garage. Two allowable relief points (R_1 and R_2) are illustrated in Figure 1.

Once again an overhead cost—the pay time for a driver to travel from the garage to the relief point—must be considered when finding a site for a garage. In the example, if the travel time from the garage to R_1 and R_2 is the same, the choice of the specific relief point is not critical. However, if route 3 has a 10-min headway and route 2 has a 30-min headway, relief pay time will be reduced by choosing R_2 for relief.

If both routes 2 and 3 had 60-min base headways, relief of drivers on route 1 could involve substantial nonproductive relief time. In fact, for a short route that is located a great distance from a garage, the travel time for drivers to the relief point can be a major cost item in route operations.

A driver does not have to be dispatched from the garage for each relief on a line. Assume that route 1 operates on a 30-min headway. Driver A reports at 6:00 a.m., takes a vehicle, and enters service. Drivers B, C, and D follow at 30-min intervals. After 4.5 h, driver A must be relieved. Driver E travels to R_2 and relieves driver A before 10:30 a.m. Driver A takes a 30-min meal break at R_2 and then relieves driver B. The process continues until each driver is relieved.

Actually, coupling of pieces of work into drivers' runs may result in other patterns, but for planning purposes it can be assumed that the number of relief drivers required per line will equal minimum meal break duration ÷ headway, rounded up to the next integer.

The number of relief movements (to or from the garage) required is a function of

1. Number of vehicles required to provide base service on the line,
2. Minimum break time,
3. Maximum time permitted between breaks, and
4. Hours in the service day.

For the system studied, which assumed operations for 19 h/day, the number of relief movements was estimated as

$$\text{Max}\ [4 \times \text{number of required vehicles},\ 8]$$

For a system in operation, relief points are long established. However, as service is changed or expanded, new relief points that are accessible by transit must be found. Rail stations tend to be excellent relief points since many bus routes will serve the stations, all required facilities are provided, and rapid, short-headway service is provided by rail throughout the day. The locating of garages such that driver relief travel can be over the rail system offers the prospect of considerable efficiency.

The example in Figure 1 might apply for a small to medium-sized transit property, where all routes operate out of a single garage. For larger properties, a route might operate from several garages, each of which results in different pull-on and pull-off and relief costs. Further, each garage has a maximum vehicle storage capacity. Those additional variables introduce added complexity to the problem.

OBJECTIVE FUNCTION

For this study, in which feasible garage sites had previously been determined, the objective was to select garage sites and assign buses to these garages such that pull-on and pull-off and relief costs are minimized subject to certain constraints.

The constraints include the following:

1. Garage storage capacity;
2. All vehicles that operate on a given line in a given period must operate out of the same garage; the additional buses required on a line for peak service were allowed to operate from a different garage than the basic service vehicles;
3. Relieved drivers must report back to the garage from which they originated and relief drivers must be assigned from this garage; and
4. No bus can travel more than 320 km/day (200 miles/day) in revenue service without returning to the garage.

PROCEDURE

The procedure for estimating costs and assigning buses to specific garages consists of the following sequence of steps:

1. Estimate pull-on and pull-off distances and times from both terminals of each line to all garage locations that are logical alternatives for the line;
2. Identify all possible relief points for each line, estimate the travel time from all alternative garages to the relief points, and select that relief point on the line for each alternative garage that yields the least relief cost;
3. Estimate unit costs for one garage-to-line relief movement and for one deadhead, one-way, garage-to-terminal movement;
4. Estimate the number of vehicles and operator reliefs required to operate base service on the line and the number of additional vehicles required to operate the peak-period service;
5. Estimate separately the total cost (pull-on and pull-off and relief) of operating base and incremental peak service for the line out of each alternative garage based on a constant definition of revenue service for the line;
6. Estimate the incremental additional costs of operating the line-period service out of the second-, third-, and fourth-best alternative garage;
7. Prepare a least-cost unconstrained vehicle assignment to the garage location alternative (base and peak vehicles are assigned on a line-by-line basis to the garage, which results in the least pull-on and pull-off and relief costs for the line-service period without regard to capacity of the garage);
8. Develop the minimum-cost constrained vehicle assignment by balancing the unconstrained assignment to the available garage capacity in a manner that adds

Figure 2. Process for analysis of pull-on and pull-off and relief costs.

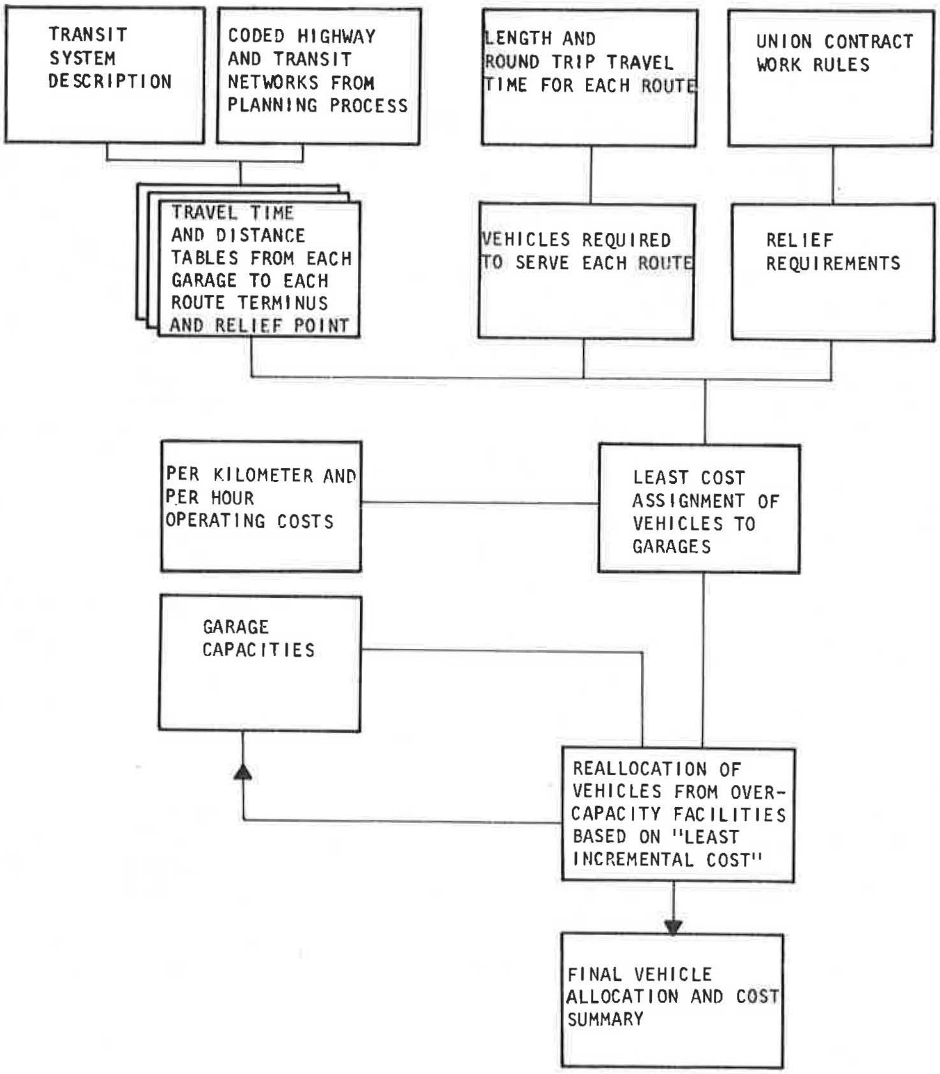


Table 1. Computation of total annual pull-on and pull-off and relief costs.

Base Case	Weekday per Day ($000s)	Weekday Annual[a] ($000s)	Weekday Base per Day ($000s)	Saturday per Day[b] ($000s)	Saturday Annual[c] ($000s)	Sunday and Holiday per Day[d] ($000s)	Sunday and Holiday Annual[e] ($000s)	Total Annual[f] ($000s)
9 garages, retain A and B	33.7	8500	16.7	13.4	737	9.2	534	9 770
9 garages, A replaced by J	34.2	8628	16.7	13.4	736	9.2	533	9 897
8 garages, eliminate B	36.0	9069	17.9	14.4	791	9.9	573	10 434
8 garages, eliminate A	37.0	9346	18.7	15.0	824	10.3	597	10 767

[a] Weekday annual = Column 1 x 252.
[b] Saturday per day = Column 3 x 0.80.
[c] Saturday annual = Column 4 x 55.
[d] Sunday and holiday per day = Column 3 x 0.55.
[e] Sunday and holiday annual = Column 6 x 58.
[f] Total annual = Column 2 + Column 5 + Column 7.

the least additional cost to the unconstrained, least-cost estimate [vehicles are reassigned from over-capacity garages to the next-least-cost alternative garage; the least-cost reassignment also considers multiple garage shifts (the ripple effect) in arriving at the least-cost-capacity-constrained vehicle assignment]; and

9. Determine jurisdictional dedication or apportionment of each line; operating costs of the line based on the minimum-cost constrained garage assignment are allocated to the designated jurisdiction.

Estimates of travel distance and time from garages to line terminals for pull-on and pull-off were developed by using a regional highway map. Distances were estimated by using a map wheel over paths judged to have the minimum travel time at each time period. Travel times were estimated by using average speeds over each type of road. While this study relied on manual effort to determine distances and judgment to determine paths and times, the process could easily be automated by using a precoded highway network and Federal Highway Administration or Urban Mass Transportation Administration software to develop the matrix of over-the-road times and distances from each garage to each pull-on and pull-off point.

Pay time associated with driver relief is computed based on travel time from the sign-in garage to the relief point, assuming travel is over the transit system. A previously coded description of the planned transit system for the design year was used to obtain the appropriate travel times. Midday headways and running times were used, with the assumption of a one-half of headway wait for the initial boarding and all subsequent transfers. Once again, the computation of travel times was carried out manually, but the process could easily be automated by using existing Urban Transportation Planning System (UTPS) programs.

Once all time and distance factors were established, it was only necessary to determine the number of vehicles required to serve each line and the number of driver reliefs in order to compute the costs of these operations for each route from each garage. Figure 2 illustrates the process.

To determine the costs associated with various garage locations the following process was applied:

1. An initial allocation of routes (vehicles) was made, assuming no garage capacity limitations;
2. A table of differences was constructed that showed the difference in cost of serving a route from the least costly garage and all other alternatives;
3. Vehicles were reassigned from over-capacity facilities, in a least-cost-difference order; in this process all vehicles that serve a given line were reassigned as a group; and
4. The process was iterated until all garages were below capacity.

This represents the least-cost solution for a given garage pattern. The process was repeated, a minimal effort as all basic data were assembled, for alternative garage size and location combinations.

FINDINGS

The procedures described represent a systematic method for easy analysis of the cost implications of garage operating patterns for a major transit facility that has multiple garages. For the system studied, the operating cost difference among alternatives is roughly $1 million, which represents approximately 1 percent of the net operating budget and about 3 percent of the cost of nonrevenue service.

Table 1 presents a summary of the cost analysis for four options, including an eight-garage system, a nine-garage system, and elimination or replacement of one or more existing facilities.

For the least costly option, pull-on and pull-off costs constitute 71 percent of the total, and relief costs constitute 29 percent. Excluding services that operate only in peak hours (no relief required), pull-on and pull-off costs constitute 66 percent of the costs, and relief costs constitute 34 percent.

The study also indicated that a shift in operating patterns was called for as the operation changed from a primarily bus-oriented system to one in which the buses acted as rail feeders. Under the new conditions, garages that were once efficient became more costly. The study also indicated a deficiency in vehicle storage capacity in an area in which a garage currently existed but was planned for closure. This led to a suggested reevaluation.

One finding of relevance to many properties is the importance of relief costs in garage location. Unless a garage is located with respect to the operating service such that travel to relief points is facilitated, the cost of relief can rise quickly so that an otherwise efficient garage becomes far less cost effective.

Publication of this paper sponsored by Committee on Bus Transit Systems.

Initial Reactions to a Central Business District Bus Transit Mall in Honolulu

C. S. Papacostas and Gary S. Schnell

The city and county of Honolulu have recently adopted a plan designed to eventually convert a central business district street to a bus transit mall. The first phase of the plan, which was the imposition of turning restrictions on private automobiles, was implemented in February 1979. This paper presents the results of a study that investigated the reactions of the daytime population of the central business district toward the mall and that population's perceptions of the mall's impact on congestion, noise, air quality, safety, convenience, speed, pedestrian circulation, and the general downtown environment. The study was based on an interview survey administered to 170 persons. The major findings of the study were as follows. The mall has caused 26 percent of the automobile users to change their circulation patterns. All factors examined were thought to be enhanced by the mall. Chi-square tests showed that, at the 0.05 level, purpose and arrival time explain the perceptions of congestion and safety impacts but in different ways; mode of travel strongly affects the experience of convenience and speed; the vast majority (85 percent) of the respondents were favorably disposed toward the mall concept. These findings should be useful to urban transportation planners and decision makers because they may represent a shifting of public attitudes toward favoring the preferential treatment of high-occupancy vehicles, in general, and urban bus systems, in particular.

With the recent emphasis on improving the efficiency of existing transportation facilities, more urban areas are applying strategies to enhance the level of service delivered by their bus systems. One such strategy is the dedication of rights-of-way to the exclusive use of buses, both inside and outside major activity centers.

Acting on the recommendations of a study of five alternatives conducted in 1978 by its Department of Transportation Services (1), the city and county of Honolulu adopted a plan that will eventually convert a 0.8-km (0.5-mile) stretch of the central downtown street (Hotel Street) into a two-way bus transit mall. The first step of the plan was implemented on February 13, 1979. This step consisted of the prohibition of private automobiles from turning onto Hotel Street from most cross streets between the hours of 6:00 a.m. and 6:00 p.m., Monday through Saturday.

Approximately three weeks after the turning-movement restrictions went into effect, the Civil Engineering Transportation Program (CETP) of the University of Hawaii conducted a survey to discern the initial reactions of the daytime population of the downtown area toward the transit mall.

STUDY DESCRIPTION

The city and county of Honolulu encompasses the entire island of Oahu (Figure 1). The estimated 1977 de facto population of the island, which includes military personnel, their dependents, and visitors, was 777 000 persons or about 80 percent of the state total. The corresponding density was 503.6 persons/km^2 (1304.3 persons/mile^2) (2).

The most densely populated part of Honolulu is located in an east-west corridor on the southern side of the island. It lies between the ocean to the south and the Koolau mountain range to the north and extends on both sides of the central business district (CBD) (Figure 1).

The CBD has experienced heavy growth over the past decade. Recent estimates of the labor force place the number of jobs there at more than 30 000 (1). The 0.5-km^2 (0.2-mile^2) CBD is bounded by Nimitz Highway to the south, which runs along the Honolulu Harbor, and Vineyard Boulevard to the north (Figure 2). Honolulu's major freeway (H-1) also runs in the east-west direction north of Vineyard Boulevard. North King and Beretania Streets form a major arterial, one-way couplet that traverses the CBD on both sides of Hotel Street, which bisects the downtown area.

Hotel Street is a two-way street approximately 11 m (36 ft) in width. It serves as the major bus roadway in the downtown area. It currently carries a peak-period bus volume of 72-80 buses/h (3).

The survey instrument used in this study was a personal interview questionnaire (Figure 3). The questionnaire was divided into four parts: The first part elicited basic socioeconomic information about the respondents such as age and occupation and travel characteristics such as travel mode and trip purpose. The second part asked whether the mall had an effect on the respondents' CBD travel habits such as trip frequency and choice of mode. The third part requested that respondents assess the effects of the mall on typical impacts such as congestion and air quality, and the last part asked for the respondents' opinion on whether the transit mall should remain in operation.

In order to cover the major segments of the day population of the downtown area, five students who were participating in a university training program funded by the Urban Mass Transportation Administration (UMTA) were instructed to circulate in the general downtown area within a city block on either side of Hotel Street and approach potential respondents randomly during the midday off-peak period. The timing of the survey was selected to coincide with the lunch period in order to ensure the inclusion of representatives of all segments of the daytime population. For example, office workers who drive to work were considered less likely to be encountered during other hours of the day. A total of 170 valid interviews were conducted in this manner during the period from March 7 to March 20, 1979.

RESPONDENT PROFILES

Modal Choice

By nature, the transit mall provides for the preferential treatment of one mode of travel over another. For this reason, the respondent profiles given next make reference to the mode used. Three modal families are included: bus, automobile, and other.

Of the 170 persons interviewed, 52 percent were bus riders, 40 percent were automobile drivers or riders, and 8 percent were users of other modes, including walking. These percentages do not necessarily represent the overall modal split since bus riders and walkers may have been encountered more often than automobile users during the survey that was administered at the street level.

Age and Sex

The calculated average age of approximately 36 was

Figure 1. City and county of Honolulu.

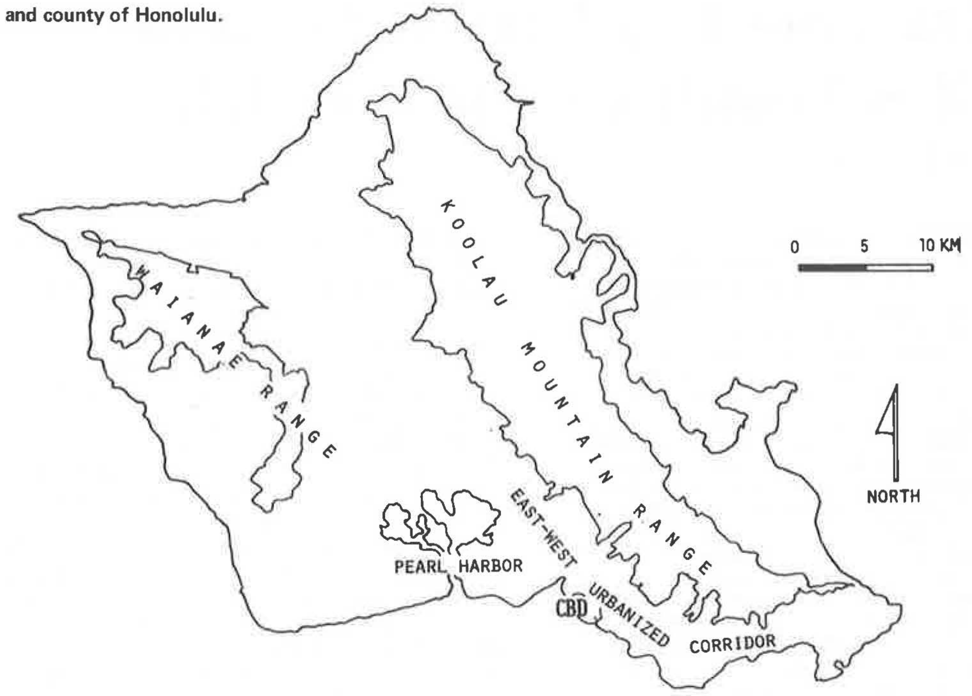
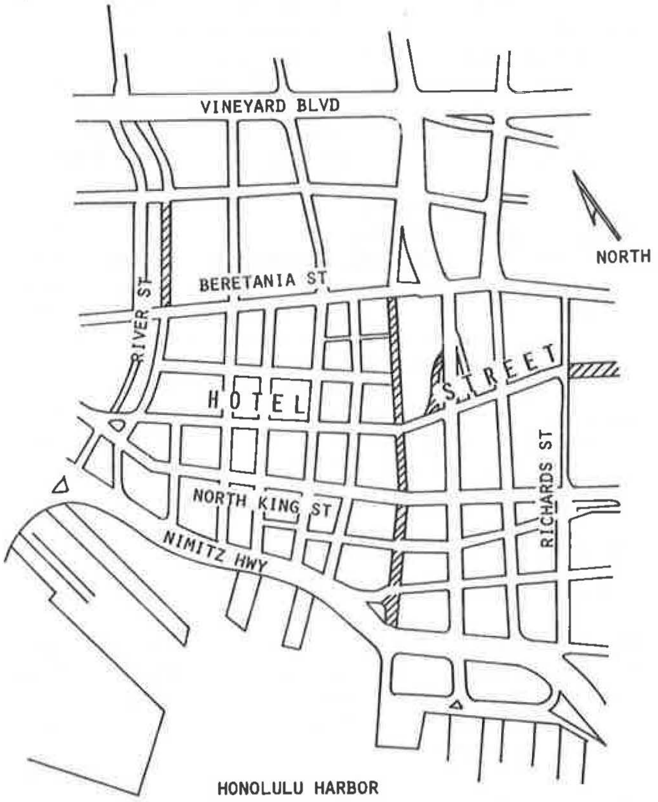


Figure 2. Honolulu CBD.



found to be independent of mode. The sex profiles, however, were found to be different at the 0.05 level of significance on the basis of the chi-square test. The male-female split was 30-70 in the case of bus riders and 65-35 in the case of automobile users. Males and females were equally represented among the users of other modes.

Occupation

Professional or technical and clerical or service workers had an equal share in the automobile sample; each group constituted about 30 percent of the total. By contrast, the representation of these two groups among the bus riders was 10 and 50 percent, respectively. Persons not in the labor force (unemployed, housewives, and retirees) made up about 15 percent of each modal sample. Students constituted approximately 20 percent and other workers about 5 percent of the bus patronage. In the case of automobile use, the last two proportions were reversed.

PERCEPTIONS OF MALL IMPACTS

Changes in Travel Habits

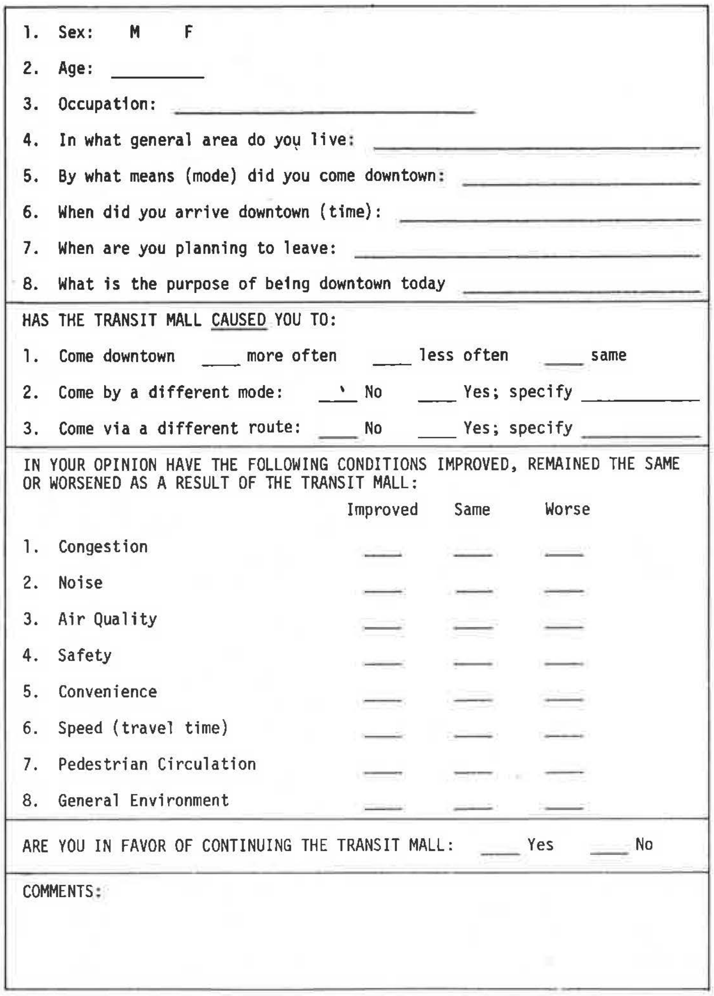
One section of the survey questionnaire asked whether the existence of the transit mall had caused changes in the respondents' travel habits, such as the frequency of travel to the CBD, the mode, or the route used to get there. The only significant change was in the choice of route—26 percent of the automobile users were diverted from Hotel Street to other downtown streets. The reason why no other change occurred is most probably due to the fact that travel on Hotel Street constitutes only a small part of the average overall trip length.

Impact Perceptions

Another section of the questionnaire asked the respondents whether they perceived improvement, stability, or degradation in eight transportation impact areas as a result of implementing the transit mall. The impact areas specified were traffic congestion, noise level, air quality, safety, convenience, travel time (or speed), pedestrian circulation, and the general downtown environment.

Table 1 shows that positive responses exceeded negative replies in each case and that less than 15 per-

Figure 3. Survey questionnaire.



1. Sex: M F
2. Age: ________
3. Occupation: ________________
4. In what general area do you live: ________________
5. By what means (mode) did you come downtown: ________________
6. When did you arrive downtown (time): ________________
7. When are you planning to leave: ________________
8. What is the purpose of being downtown today ________________

HAS THE TRANSIT MALL CAUSED YOU TO:

1. Come downtown ____ more often ____ less often ____ same
2. Come by a different mode: ____ No ____ Yes; specify ________
3. Come via a different route: ____ No ____ Yes; specify ________

IN YOUR OPINION HAVE THE FOLLOWING CONDITIONS IMPROVED, REMAINED THE SAME OR WORSENED AS A RESULT OF THE TRANSIT MALL:

	Improved	Same	Worse
1. Congestion	____	____	____
2. Noise	____	____	____
3. Air Quality	____	____	____
4. Safety	____	____	____
5. Convenience	____	____	____
6. Speed (travel time)	____	____	____
7. Pedestrian Circulation	____	____	____
8. General Environment	____	____	____

ARE YOU IN FAVOR OF CONTINUING THE TRANSIT MALL: ____ Yes ____ No

COMMENTS:

Table 1. Respondent perceptions of the mall's impacts.

Impact	Improved (%)	Same (%)	Worse (%)
Congestion	57	31	12
Noise	32	60	9
Air quality	23	68	9
Safety	56	35	9
Convenience	46	42	12
Speed	38	48	13
Pedestrian circulation	48	46	6
General CBD environment	51	43	6

cent of the respondents perceived any one condition to be adversely affected by the transit mall. The differences between the percentages of positive and negative responses are designated as the weighted ratings of the effect of the mall on each of the eight conditions in the table below.

Impact	Weighted Rating
Congestion	45
Noise	23
Air quality	14
Safety	47
Convenience	34
Speed	25
Pedestrian circulation	42
General CBD environment	45

Use of this difference is equivalent to assigning the values of +1, 0, and -1, respectively, to each positive, neutral, and negative response. According to this value system, four impact areas (safety, congestion, the general downtown environment, and pedestrian circulation) were on the average thought to have experienced the greatest improvement. Note, however, that the first two (safety and congestion) received the highest proportions of positive reactions and the last two (general environment and pedestrian circulation) received the lowest proportions of adverse reactions.

Noise and air quality, the two main concerns that relate to the physical environment, were implicitly placed in the same category by the respondents. These two impacts were perceived most often to be unaffected by the transit mall (Table 1). They also received the lowest weighted rating in the above table.

The remaining two impacts (speed and convenience) received mixed reactions.

Explanatory Variables

Table 2 presents the experimental significance levels computed with the aid of the Statistical Package for the Social Sciences (4) by using the chi-square test. Each cell of the table shows the result of a separate test that compared the responses to the corresponding impact of the respondents who belong to the various categories of the corresponding attribute (see table below).

Attribute	Categories
Mode	Bus
	Automobile
	Other
Sex	Male
	Female
Purpose	Work
	Shop
	Other
Occupation	Professional or technical
	Clerical or sales
	Service
	Other
Age	Under 20
	21-60
	Over 60
Arrival time	Peak period
	Off-peak period

According to Table 2, the comparison of the responses of males and females (i.e., the sex categories) to the impact of safety resulted in an experimental significance level of 0.3432. Only three attributes (purpose, arrival time, and travel mode) were found to affect the responses relating to some impacts at the 0.05 level.

Congestion

Purpose and arrival time were found to affect the perception of the mall's impact on congestion. Workers showed a higher propensity to indicate an improvement in this impact than did shoppers and travelers for other purposes (65.3 versus a combined 43.9 percent). Their corresponding percentages of adverse responses showed a closer agreement (12.6 versus 10.6 percent).

The responses of peak-period travelers differed from the responses of off-peak travelers in all categories. More peak-period travelers felt a favorable effect on congestion (62.5 versus 34.6 percent). The percentage of adverse responses was higher in the case of peak-period travelers (18.8 versus 9.6 percent). Off-peak travelers perceived no change more often than did the rest (55.8 versus 18.8 percent).

Safety

The two variables that were found to affect the respondents' congestion experience (i.e., purpose and arrival time) were also found to affect the assessment of the safety impact that the transit mall conversion wrought. The assessments of safety impacts by respondents from the various purpose and arrival time categories, however, differed from their reactions to the question of congestion effects.

Shoppers cited safety improvements more often (65.2 percent) than did workers (57.4 percent) and travelers for other purposes (46.7 percent). On the other hand, workers were more likely to feel an adverse effect when compared with shoppers (11.7 versus 4.3).

About 30 percent of workers and the same percentage of shoppers agreed that the mall had no effect on safety. The corresponding proportion of those engaging in other activities was 45 percent.

The responses of off-peak-period travelers were almost equally split between the positive and neutral categories (45.3 and 50.9 percent, respectively). On the other hand, peak-period travelers were unevenly divided: 60.3 percent perceived safety improvements, 28.6 percent perceived no change, and 11.1 percent indicated a worsening of safety conditions.

Convenience and Speed

Travel mode was found to strongly explain the respondents' perceptions of the mall's effect on both convenience and speed. A larger proportion of bus riders than automobile users experienced an improvement in convenience (54.1 versus 30.6 percent). On the other hand, the percentage of automobile users who said that they had been inconvenienced by the transit mall (22.6 percent) far exceeded the percentage of bus riders who felt the same way (4.7 percent).

A response pattern similar to that for convenience was detected in the case of speed. About 41.7 percent of the bus riders thought that their speed had improved, but only 3.6 percent of them noticed speed degradation. The corresponding automobile percentages were 31.1 and 27.9, respectively.

Respondents' Views on Mall Continuation

An overwhelming majority (85 percent) of the respondents favored the continuation of the bus transit mall. Moreover, with only a single exception, chi-square tests showed that there was no difference between the overall percentage and the percentages corresponding to the various subgroups in the sample at the 0.05 level of significance.

The exception was in the proportions of bus patrons and automobile users. Although both exhibited a highly favorable disposition toward the continuation of the mall, the 93.7 percent corresponding to bus riders was found to be significantly larger than the 74.1 percent shown by automobile users. Automobile users whose route choice had been affected by the mall exhibited the strongest opposition but, even in this case, the majority (64.7 percent) favored the continuation of the mall.

SUMMARY AND CONCLUSIONS

The survey described in this paper found that the vast majority (85 percent) of the daytime population of Honolulu's CBD were favorably disposed toward the continuation of a bus transit mall in the downtown area. The majority (65 percent) of those automobile users whose downtown circulation patterns were affected by

Table 2. Experimental chi-square significance levels.

Impact	Mode	Sex	Purpose	Occupation	Age	Arrival Time
Congestion	0.5923	0.8535	0.0183[a]	0.1952	0.0628	0.0002[a]
Noise	0.1275	0.6415	0.5919	0.5972	0.1602	0.4576
Air quality	0.7643	0.9985	0.4649	0.3116	0.6724	0.7850
Safety	0.2258	0.3432	0.0158[a]	0.9460	0.6202	0.0313[a]
Convenience	0.0025[a]	0.2748	0.4720	0.8272	0.2495	0.1324
Speed	0.0008[a]	0.1935	0.4202	0.6593	0.7537	0.3126
Pedestrian circulation	0.8983	0.5321	0.1233	0.4538	0.4409	0.1542
General CBD environment	0.2392	0.2879	0.8295	0.8191	0.3789	0.5329

[a] Significant at the 0.05 level.

the mall were also in favor of continuing the bus mall. This group constituted 26 percent of the automobile users interviewed.

Less than 15 percent of the respondents perceived any one of the following conditions to be adversely affected by the mall: traffic congestion, noise, air quality, safety, convenience, speed, pedestrian circulation, and the general CBD environment. The mall's impact on safety and congestion received the highest proportions of positive responses; the impact on the general environment and on pedestrian circulation received the lowest proportions of adverse reactions. Noise and air quality were perceived most often to be unaffected by the transit mall. Speed and convenience received mixed reactions.

Chi-square tests showed that, at the 0.05 level of significance, trip purpose and arrival time explain the differences in the respondents' perceptions of the mall's impact on congestion and safety. The mode of travel to the CBD made a difference in the perceived effects that the mall had on convenience and speed.

These findings should be useful to urban transportation planners and decision makers because they may represent a movement of public attitudes in the direction of favoring the preferential treatment of high-occupancy vehicles, in general, and urban bus systems, in particular.

ACKNOWLEDGMENT

This paper was produced as part of a transportation research and training program funded by the Urban Mass Transportation Administration, U.S. Department of Transportation, and conducted with the cooperation of the Department of Transportation Services, city and county of Honolulu. The results and views expressed are our own and are not necessarily concurred with by either the Urban Mass Transportation Administration or the department of transportation services.

REFERENCES

1. Central Business District East-West Transportation Corridor Study. Department of Transportation Services, city and county of Honolulu, technical rept. RT-78-01, Aug. 1978.
2. The State of Hawaii Data Book. Department of Planning and Economic Development, Honolulu, 1978.
3. Hotel Street Bus Demonstration. Department of Transportation Services, city and county of Honolulu, technical rept. TP-78-02, July 1978.
4. N. H. Nie and others. Statistical Package for the Social Sciences, 2nd ed. McGraw-Hill, New York, 1975.

Publication of this paper sponsored by Committee on Bus Transit Systems.

Recent Experience with Accessible Bus Services

Robert Casey

Fixed-route, standard-sized buses equipped with level-change mechanisms to transport wheelchair or semiambulatory passengers between the ground and the bus floor level are currently in service in 23 locations in the United States. This paper includes a brief description of the services in place and a discussion of experience with their operation. Data are limited due to the newness of many of the services and the fact that few transit operators collect the kind of information that is most useful for evaluation. Available data have been collected to inform planners and operators of future accessible bus services of the policy issues and operational impacts they probably will face and the level of ridership they initially can expect. A few findings can be stated: (a) Lift reliability has improved substantially through the emergence of new lift designs and modifications to existing models and (b) ridership continues to be low, with most transit operators reporting between one and three lift-assisted boardings per day. Most of these trips are taken by a few regular riders. The economic impact varies considerably among operators, depending on the reliability of the particular model of lift operated and whether schedule changes were instituted specifically for implementation of the accessible buses. At current lift-utilization rates, accessible bus service will not significantly affect transit operations.

The past year has seen some major developments in the area of fixed-route accessible bus service that uses standard-sized buses. This type of service has been initiated by 18 more transit authorities, which makes a total of 23 now in operation. Three new level-change devices (most often called lifts) are now being used in service. The Transbus concept, which the U.S. Department of Transportation (DOT) thought would be the solution to fixed-route bus accessibility, received a setback when bid solicitations for the bus produced no respondents. The National Research Council review panel concluded that Transbus, as specified in the solicitation, could not be built without considerable technical and financial risk on the part of the manufacturers (1). Also, DOT issued regulations to implement Section 504 of the Rehabilitation Act of 1973, which (among other provisions) mandated the purchase of accessible buses for every bus ordered after July 2, 1979. These regulations are currently being challenged in court by the American Public Transit Association (APTA).

In spite of the number of accessible bus services that are operational, a wealth of data is still not available. The prime reason for this is that the collection and analysis of the type of information most useful to policy-makers and other transit operators require a substantial evaluation effort, an undertaking that is beyond the fiscal resources available to many transit properties. Consequently, the most detailed information about accessible services will continue to be disseminated through the Urban Mass Transportation Administration (UMTA) Service and Methods Demonstration (SMD) program.

ACCESSIBLE SERVICES

The 23 accessible fixed-route services in operation by December 1979 are summarized in Table 1. The number of accessible buses available at these locations totaled 1239, although only 759 were scheduled to be in service at that time.

In addition, a number of accessible buses are on order from the two U.S. manufacturers of advanced-design buses [Grumman-Flxible and General Motors Corporation (GMC)] and the Canadian manufacturers of new-look buses (General Motors of Canada and Flyer Industries). Accessibility has been and will continue to be achieved primarily in connection with new bus purchases.

IMPLEMENTATION ISSUES

The implementation process requires policy decisions about the following:

1. Whether to institute schedule changes,
2. What to do about reduction in seating capacity,
3. How much driver assistance to permit, and
4. Who will be allowed to use the lift.

A few transit operators made operational changes in anticipation of potential delays in running time caused by wheelchair boardings and alightings. However, the low ridership by wheelchair users has indicated that such action would not appear necessary. Most operators have not made operational changes but are observing the operations closely to determine whether the delays experienced significantly disrupt normal operations on the accessible routes. Changes may be made later if service is seriously affected.

Several operators are concerned by the loss of regular seating capacity due to the provision of wheelchair tiedowns on the accessible buses. The majority of tiedowns take away two regular seats per tiedown position. Four operators have added buses and decreased headways to compensate for this and for the seats lost due to the smaller seating capacity of the advanced-design buses. Extra buses probably would not be added solely due to the loss of seats in the tiedown areas.

Driver assistance to severely mobility-limited passengers is either not permitted or discouraged by nine transit operators. In those situations, potential passengers may be inhibited from using the bus service if they encounter or fear difficulty in using the lift or the tiedown devices.

Seven of the operators have taken the position that only persons in wheelchairs will be allowed to use the lift due to a concern about potential injuries and accident claims from lift standees. In current lift-equipped buses, headroom clearance at the door frame is restricted for persons who stand on the lift.

TRAVEL BEHAVIOR

Surprisingly, accessible-bus-ridership data about wheelchair users are not always available or accurate. Only about one-half of the operators are able to supply what they consider reasonably accurate ridership totals. In spite of the data limitations, ridership figures are presented for every locale for which figures are available.

Ridership Data

San Diego Transit Corporation (SDT) operates accessible bus service with five buses (four scheduled) on two routes. SDT reported an average of 41 wheelchair-passenger trips/month during 1978. In view of the small number of accessible buses used, this is relatively high compared with most of the transit properties. According to SDT, four people are regular riders. The 1978 ridership was about double the monthly ridership totals of 1977. No figures are yet available for 1979.

The Southeastern Michigan Transportation Authority (SEMTA) now has 61 (50 scheduled) buses that operate on nine (seven fully accessible) routes. Virtually all transit

Table 1. Accessible bus service characteristics.

				Accessible Buses[b]				
	Initial Service	Manufacturer		On Prop-		No. of Routes	Fleet Ac- cessibility	Recent Lift Uses
Location	Date	Bus	Lift[a]	erty	Scheduled	Accessible[b,c]	(%)	per Month
San Diego	2/77	GMC[d]	TDT	5	4	2 P	1	41
St. Louis	8/77	Flxible	TDT	157	41	12 P	15	30[g]
San Mateo	9/78	AMG	TDT	24	15	2 P	11	30-40[g]
Detroit (SEMTA)	10/78	GMC[e]	GMC	61	50	7 F, 2 P	18	40-50[g]
Santa Clara	12/78	Gillig GMC[e]	TDT GMC	52 58	21	3 F	27	NA
Fayetteville	12/78	GMC[e]	GMC	6	6	6 P	23	7
Gardena	12/78	GMC[e]	GMC	2	1	1 P	6	NA
Rhode Island	1/79	GMC[e]	GMC	19	15	5 P	9	81[g]
Westchester	3/79	GMC[e]	GMC	105	91	8 P, 2 F	42	NA
Hartford	4/79	Flxible[e]	EEC	155	140	26 P	56	38
Rock Island	4/79	GMC[e]	GMC	7	6	6 P	23	<1
Milwaukee	4/79	Flxible	Vapor	100	62	6 P	17	21
New Haven	6/79	Flxible[e]	EEC	100	90	18 P	83	43
Janesville	6/79	GMC[e]	GMC	10	7	7 P	53	100
Washington	7/79	Flxible	Vapor	150	93	34 P	8	150[g]
Montebello	7/79	GMC[e]	GMC	8	6	2 P	22	0
Ventura	7/79	GMC[e]	GMC	2	1	1 P	5	6
Detroit (DDOT)	9/79	GMC[e]	GMC	163	26	1 corridor F	20	2
Stamford	9/79	Flxible[e]	EEC	35	28	13 F	100	10-15[g]
Seattle	9/79	Flyer	Lift-U	50	27	6 F	7	100+[g]
Colorado Springs	9/79	GMC[e]	GMC	14	0[f]	9 P	37	80[g]
Palm Beach	10/79	GMC[d]	TDT	15	3	1 F	25	28
Los Angeles	11/79	AMG	TDT	28	11	1 P	NA	NA

[a] TDT = Transportation Design and Technology, Inc.; EEC = Environmental Equipment Corporation.
[b] As of November 1979.
[c] P = partially accessible; F = fully accessible.
[d] Retrofitted buses.
[e] Advanced design buses.
[f] Published schedules do not indicate accessible buses.
[g] Estimated.

trips by wheelchair users are made by two persons who transfer from a dial-a-ride service to the accessible fixed-route buses. SEMTA estimates that these two riders take about 40-50 bus trips/month.

The San Mateo County Transit District (SAMTRANS) operates 24 (15 scheduled) accessible buses on two routes. SAMTRANS estimates that ridership by lift users averaged 30-40 one-way trips/month from July to September 1979. Usage has dropped off somewhat since the end of 1978. SAMTRANS' continuing lift-reliability problems have undoubtedly affected ridership. SAMTRANS reports that (a) the majority of lift users are persons in wheelchairs, (b) three wheelchair users take most of the recorded trips, and (c) these three users transfer from a dial-a-ride service to the accessible buses.

The Bi-State Development Agency scheduled 126 of their 157 accessible buses in daily service over the period November 1977-August 1978. However, since the actual availability of accessible buses was generally far short of the required number, Bi-State cut back the number of scheduled buses to 40 in September 1978.

A total of 60 unduplicated users of Bi-State's accessible service were identified during the nearly two-year evaluation period. Only 13 wheelchair users made more than 10 one-way trips on the buses, which represents 82 percent of all trips reported. Ridership has been highly variable, due in part to weather conditions as well as equipment reliability problems and the service cutback. Recent totals have been about 30 boardings/month.

In Fayetteville, North Carolina, the transit operator schedules all six of the accessible buses in service, one on each of the six routes. Since on an average day one of these vehicles is out of service, missed runs are a common occurrence. Since July 1979, when ridership counts were begun, boardings averaged 6/month.

The Rhode Island Public Transit Authority (RIPTA) has had 19 accessible buses (15 scheduled) in service since January 1979. Over the following six-month period, 45 wheelchair trips were recorded, or an average of 8 boardings/month. However, expansion of a one-week count in November would result in a monthly total of 81 boardings. RIPTA indicates that most of the trips counted in the November sample week were taken by three persons.

Milwaukee Transport Services operates 100 accessible buses (55 scheduled) on six routes. Ridership built up to approximately 50 boardings/month during the summer, but dropped off to less than 10 in October and November. They estimate that 80 percent of the trips are made by four or five persons.

Connecticut Transit operates 100 (90 scheduled) accessible buses in New Haven. Each of their 18 regular routes are partially accessible during the peak periods and virtually 100 percent accessible during the off peak. Lift users made an average of 43 boardings/month from August to October 1979.

In Hartford, Connecticut Transit schedules 140 of their 155 accessible buses. All of their routes are partially accessible during the peak periods and almost fully accessible during the off peak. Ridership, which has been somewhat erratic, averaged 38 boardings/month from September to November. Nearly twice that number used the service in August.

In Stamford, Connecticut Transit has operated a fully accessible, 35-bus (28 scheduled), 13-route system since September 1979. November was the first month in which any lift users were carried. The operator estimates that 10-15 lift-assisted boardings were made during that month.

The Janesville (Wisconsin) Municipal Bus System has averaged more than 100 lift-assisted boardings/month from July to November 1979 on their 10 accessible (7 scheduled) buses. One daily rider accounts for almost one-half of the trips. There are four or five other occasional riders.

Rock Island County (Illinois) Metropolitan Mass Transit District has carried only two wheelchair passengers in seven months of operations on their seven (six scheduled) accessible buses. However, except for printed schedules and contracts with social service agencies, there has been little marketing or advertising of the service.

The Washington Metropolitan Area Transit Authority (WMATA) has recently expanded its accessible bus service to 150 (93 scheduled) buses that operate on 34 routes. WMATA estimates 150 wheelchair-user boardings/month based on the known travel patterns of some regular riders.

The Santa Clara County Transportation Agency's recently delivered advanced-design buses have not yet been put into service. At this time, only 21 of their 110 accessible buses are scheduled for three routes. Sample counts were made on a few days in January, February, and March of 1979 on the two routes then in service. An average of 19 lift-assisted boardings were recorded on those days, 9 of them by nonwheelchair users.

Montebello Municipal Bus Lines have carried only one lift user in the early months of operations of their eight (six scheduled) accessible buses.

South Coast Area Transit in Ventura, California, schedules one of their two accessible buses on an hourly headway on their heaviest route. They have averaged 6 boardings/month of wheelchair users.

The city of Detroit Department of Transportation (DDOT) has not yet placed their 122 new accessible buses in service. Consequently, they still operate only 41 (26 scheduled) buses in one fully accessible corridor. In the short period of operation, boardings by wheelchair users have averaged only about 2/month.

Seattle Metro currently schedules 27 of their accessible buses on six routes. Four of these routes began in November. A total of 153 accessible buses will be on the property when delivery of the current order is completed in February. Accessible trolley buses have also been ordered. The six accessible routes have attracted three regular commuters plus a few other occasional riders. Metro estimates that boardings will exceed 100/month.

In Colorado Springs, Colorado Transit Management, Inc., operates 11 or 12 of their 14 accessible buses on their nine routes. However, the accessible bus trips are not yet noted on schedules. In spite of the lack of schedules, the transit operator estimates that wheelchair user ridership totals were about 80 during November.

The Palm Beach County Transportation Authority began accessible bus service in October with 3 of their 15 retrofitted vehicles. One route has been made fully accessible. During the month of October, 28 lift-assisted boardings were recorded. This dropped to 6 in November.

The Southern California Rapid Transit District (SCRTD) in Los Angeles is operating an accessible bus demonstration route that uses 11 scheduled accessible buses. SCRTD has made a special effort to get wheelchair users to ride the line so that they can assess the operational impacts and make appropriate operational changes, if needed, in routes to be made accessible later. During the first few weeks of service, 17 lift-assisted boardings were recorded, 11 of them on one day. However, this is not indicative of ridership to be expected in regular, accessible service.

Summary

As can be seen from Table 1 and the above discussion, ridership by persons who need the lift to board buses has been low. On the other hand, to date, generally only a portion (often small) of the buses at a transit property are accessible, and many of them have been in service for only a few months. The changes in demand in response to increases in the supply of accessible service is a relationship that will be watched with considerable interest. In those instances where the same number of accessible buses has been in operation for six months or more, there has generally not been a continuing increase in ridership.

There are a number of possible reasons for the lack of significant ridership increases. Equipment problems are one of these. Severe winter weather will certainly inhibit travel by wheelchair users. Restrictions placed on driver assistance to wheelchair passengers by some transit authorities may also be a factor that tends to depress ridership. Without the assistance of either the driver or another passenger or companion, some wheelchair users may be unable to use the bus. It is also quite possible that more time is needed for the target population to change their travel habits or patterns and switch to use of the accessible buses.

Very few data are available concerning reasons why more of the target population do not ride the accessible buses. Results from a survey of 60 wheelchair users in St. Louis who had not ridden the accessible buses are presented below; the ratings are based on a scale in which 1 = least important and 5 = most important:

Reason for Not Using Buses	Average Rating
Do not need them—have other transportation available.	3.5
I cannot go out at all without help.	3.5
Lack of curb-cuts near my home or my destination.	3.1
It is too difficult for me to travel on sidewalks or roads to reach the bus stop.	2.8
Bad weather such as rain, snow, or cold.	2.5
Accessible routes do not go near my residence.	2.2
Bus transportation takes too long or is too inconvenient compared to a car.	2.2
Accessible routes go near my home but do not go near my destination.	2.1
Cannot get on the bus lifts very easily.	2.1
Trouble obtaining the schedule of accessible buses.	2.0
Cars parked in bus stop.	1.9
Lifts are unreliable and sometimes do not work.	1.7
Afraid to use—heard bad things.	1.7
The buses are unreliable and do not keep to the published schedule.	1.6
Do not feel safe on the lifts or on the buses.	1.5
Buses are too crowded when I want to use them.	1.5
Do not like going out in public.	1.4

The reasons for nonuse that received the highest ranking were the difficulty of going out at all, the difficulty of getting to or from the bus stops, the availability of another mode of travel, and weather. These answers give evidence that fixed-route accessible buses will only be a feasible alternative for a portion of the target population.

For other segments of the handicapped and elderly populations, the situation is less clear. Little is known about the potential benefits of accessible buses to the elderly and nonwheelchair handicapped. The number of nonwheelchair users who would actually use transit because lifts or other accessibility aids are provided would be difficult to determine.

LIFT EQUIPMENT

Five different wheelchair lifts have been installed by bus manufacturers in standard-sized buses for use in regular fixed-route transit service. The manufacturers of these lifts are Transportation Design and Technology, Inc.; Vapor Corporation; General Motors Corporation; Environmental Equipment Corporation; and Lift-U, Inc. A sixth lift, manufactured by Transi-Lift Equipment Ltd., has only been installed for testing purposes.

Reports from the operators indicate that all lifts have some deficiencies or drawbacks. As deficiencies have surfaced, lift manufacturers generally have been responsive in making modifications to improve performance. As a consequence, the current models are considerably more reliable and usable by the passengers than they were one year ago. It seems reasonable to expect further lift reliability and performance improvements in the years ahead.

Operations

Even though the state of the art in lift technology is advancing, a few transit authorities continue to have difficulty in always providing an accessible bus in accordance with the published schedules. The extent to which this situation prevails will depend, in part, on the number of accessible buses retained as spares. As noted previously, lift reliability has improved considerably. St. Louis no longer reflects the current status of wheelchair lift performance. A couple of transit operators reported that the lifts themselves work quite well but that many of the problems are caused by driver mistakes.

The availability of accessible vehicles can also be affected by their heavy use. These buses are operated more than are nonaccessible buses due to their constant use in both peak and off-peak periods. Consequently, they require more frequent maintenance and repair than do other vehicles. As a result, a larger number of spare lift buses are probably required for schedule adherence than are normally required for the rest of the fleet. This would vary for lifts produced by different manufacturers.

The added workload for maintenance of the lifts has resulted in an increase in the maintenance staff in St. Louis, Milwaukee, Hartford, Washington, Santa Clara, New Haven, Stamford, and Detroit. It seems likely that extra maintenance personnel will be needed by all transit properties that implement any significant amount of accessible bus service.

The potential for through-routing may be restricted or placement of the accessible buses may be complicated if the transit system is only partially accessible. If through-routing is reduced, greater bus service hours should result. It is not known how much through-routing reduction has occurred.

Four transit operators have added buses on accessible bus routes to compensate for seats lost in the wheelchair tiedown positions and due to the smaller seating capacity of the new buses. Buses probably would not be added for the provision of tiedowns alone.

Buses will be delayed whenever wheelchair passengers are carried. The average time for wheelchair users to board and tie down will probably average 2.5-3 min. The average alighting time is about 1 min less than the boarding time. For each wheelchair passenger, the delay would average 4-5 min.

Some transit operators are forced by state laws to operate the narrow, 264-cm (96-in) wide buses. If the lift is installed in the front door of these buses, wheelchair users often find it very difficult to maneuver once inside. This would result in a longer dwell time at the stop. If the lift is installed in the rear door of the bus, the maneuvering problem is removed, since the tiedown positions can be installed directly across from the door. The drawbacks to this lift location are that the driver has to go to the rear of the bus to operate the lift, and maneuvering the bus so that the rear door is adjacent to the curb is not always easy.

Costs

Capital costs incurred will consist primarily of the cost of the lift equipment and wheelchair tiedown devices. The cost of different lifts plus tiedowns on new bus orders prior to the effective date of Section 504 regulations ranged from $8500 to $14 000. Since all buses ordered now must be accessible, the price of the lifts and tiedowns are usually not separated from the total price of the bus and, consequently, are difficult to identify. Retrofits of existing buses would probably cost about twice that of new buses.

In addition to capital costs, accessible bus service will incur increased operating costs. These include the cost of any schedule changes, through-routing reduction, staff time, extra mechanics, driver training, promotion and advertising, accident claims (if any), and extra drivers' pay (if any).

Transit operators who reduce headways to compensate for seats lost in providing wheelchair tiedown positions may incur substantially increased operating costs. SEMTA states that each extra bus costs them $60 000-$70 000/year.

Extra mechanics have been hired by several properties to work on the lifts. Each extra mechanic would probably cost $15 000-$20 000/year.

The amount of driver training given has varied from 0.5 h on lift operation to 10 h, including sensitivity training on the problems of the handicapped. At most properties, drivers are paid extra for attending these training sessions, often at overtime rates. All drivers at each site are normally given this training. Refresher training may also be necessary in some cases.

The amount of promotion and advertising of accessible bus services has varied considerably. Some have used television, radio, and newspapers and have conducted field demonstrations, while a few have done little more than publish schedules that note the accessible bus trips. Bi-State, which had an extensive radio and newspaper advertising campaign, spent approximately $35 000 during the first year of operations.

To date, two operators have reported accident claims due to lift operations. Specific accident-claim data are not available at this time, but most of the claims have been small.

Only in Detroit has lift operation been a major labor issue. The issue was resolved by DDOT when it agreed to pay the drivers $0.50 for each time they operate the lift for a wheelchair passenger. At both SEMTA and DDOT, the minimum layover time was increased by 5 min for accessible buses. The only other known instance where drivers are paid extra for operating accessible buses is in Rhode Island, where they are paid 5 min extra time for cycling the lift before leaving the garage each day.

OTHER IMPACTS

The impact on transit travel time of nontarget group riders will be minimal if wheelchair ridership remains low. Lift operation for two or more passengers during a single bus run would delay other riders and bus operations significantly. This is unlikely to happen with any regularity in the near future.

In order to install wheelchair tiedown positions in the buses, regular seats are usually removed. A level of service change would occur for any riders forced to stand due to the loss of seats.

Accessible bus operations have resulted in a few passenger injuries. The majority of these injuries were caused by persons who fell while boarding or alighting. Four injuries involved wheelchair users.

FINDINGS

The majority of transit authorities that operate fixed-route accessible buses have made no changes in operating procedures or schedules specifically for the implementation of accessible buses. Wheelchair-user ridership has ranged from fewer than one to a little more than three one-way trips per day. Most of the transit trips by wheelchair users are made by a few persons who are fairly frequent riders. Therefore, although accessibility to bus transportation is being improved, the overall mobility of the target population has changed very little as a result of accessible bus service.

Advances in lift technology have improved their performance substantially. As a consequence, lift maintenance expenditures are much less than for earlier models. For most transit operators, the added costs of accessible bus operation will consist principally of the capital cost of equipment, staff planning time, maintenance of lifts, driver training, and promotional expenditures.

The preliminary evidence from recent implementations of accessible bus service indicates that, at current ridership levels, this service concept will not have significant impacts on transit service operations, regular transit riders, or other providers of service to the handicapped. This would not hold true if lift use increases dramatically.

REFERENCE

1. NRC Transbus Study: Part 1—Transbus Procurement. Commission on Sociotechnical Systems, National Research Council, NRC-CSS Special Rept., Aug. 31, 1979, pp. 1-38.

Publication of this paper sponsored by Committee on Bus Transit Systems.

Abridgment

Operational Improvements in a Two-City Bus Transit Corridor

Gary G. Nelson

The Albany-to-Schenectady, New York, bus service corridor is the most heavily used corridor in the Capital District Transportation Authority's service area. The continuing decentralization of activities from the two cores out along the corridor and the increase in service demand has created the need for operational analysis of services to increase their effectiveness. The resultant study emphasized near-term operational improvements that could be implemented within fairly fixed operational support funding. A study of this nature deals with many situations that are site-specific. Of general interest will be the methodology of data collection and analysis and some of the detailed dynamics of a long, relatively high-density bus transit corridor that must consider interactions between line-haul and local services and between transit and the many factors that contribute to route delay. Of 14 problem areas originally enumerated in the study, general analytical conclusions on 7 are given. The corridor demonstrates that many of the operational inefficiencies for which separate right-of-way modes might be proposed manifest themselves at demand levels well below levels at which implementation of such modes is usually considered.

The Capital District of New York is a three-core region that is provided with bus transit service by the Capital District Transportation Authority (CDTA). Of the three regional cores, Troy, Albany, and Schenectady, the heaviest transit usage occurs between the Albany-Schenectady pair over NY-5. Over time, this corridor has experienced considerable out-migration of commercial activity from the cores and this, along with increasing general levels of transit usage, created the need for an analysis of the existing transit service with a view toward increasing its efficiency by means of short-range operational improvements.

The populations of Albany and Schenectady have been stable since 1970, when they were 116 000 and 78 000, respectively. The cores are separated by 24 km (16 miles) straight-line distance over NY-5, which is a four-lane, unlimited-access arterial. Interstate expressway routes also connect the cities but more circuitously. Suburban growth between the two cores along NY-5 has been rapid, since that area constitutes the geographic center of the region. Two major shopping malls are located along the route, one just outside Albany and the other just outside Schenectady (see corridor map, Figure 1). The increasing strip development has also meant an increase in traffic control problems, which has affected running times over the corridor.

Bus transit service over the corridor had evolved piecemeal over time. Two private firms replaced the original trolley service in the 1930s with bus service. The Schenectady local service and Albany-Schenectady through-service was run by a Schenectady firm; the Albany local service was run by an Albany-based firm. These were consolidated into the public operation in the early 1970s, but the through services along NY-5 were still operated from a Schenectady garage. Service adjustments had been made over time to meet the greater demand of the outlying shopping areas, and express service had been added that departed from NY-5 near the Albany end to proceed by expressway to and from Albany.

Since at least 1974, CDTA has been studying a comprehensive revamping of the corridor operations. More immediate problems had always intervened, however, and not until 1979 was a consultant hired to review the corridor and recommend short-range modifications to service.

The corridor is well served and has good all-day usage due to the mixture of office, industry, retail, residential, and school activities. The table below and Table 1 show some usage data. In the table below, headways are for through service with augmentation in core areas.

Day	Ridership	Headway (min)
Weekday	8593	
Morning peak	1721	12
Midday	3159	24
Evening peak	2673	15
Evening	1040	30
Saturday	5469	30
Sunday	1449	60

The Albany end is the most heavily served. Its combination of through and local service yields a 9-min weekday peak headway there, 15 min on through service, and 13 min at the Schenectady end. Curiously, the express service runs only one morning trip to Albany and a 3:00 p.m. trip to Schenectady. The corridor is broken into three fare zones. The base fare is $0.40, $0.60 for core-to-suburb trips, and $0.75 core to core.

The running time of the through buses from core to core is 70 min or a speed of 20 km/h (13.4 mph). The express buses traverse 30 km (20 miles) in 51 min or 35 km/h (24 mph). The local buses on the Albany end run 8.4 km (5.6 miles) in 40 min or 12.6 km/h (8.4 mph).

STUDY ISSUES

CDTA staff were first polled to gain their perceptions of the problems in the corridor in order to establish the order of priority of the problems. The resultant ranked list of problems is tabulated below:

1. Insufficient supply of express service;
2. Lack of late evening service on the Schenectady end to circulate arriving through riders;
3. Extended running times and schedule adherence problems in the corridor;
4. Overloads on certain existing runs;
5. Delay to through riders because of local use of through buses;
6. Lack of direct service to the Schenectady shopping mall;
7. The allocation of headways to through and local service;
8. The dispatcher's control problem in coordinating through and local services on the Albany end, since the Schenectady-based through services do not originate from the Albany garage;
9. The problems of the zone-fare system and questions about fare level;
10. The lumping of peak headways due to certain interactions of through and local schedules;
11. Poor connection between the retail activities on

Figure 1. Corridor map.

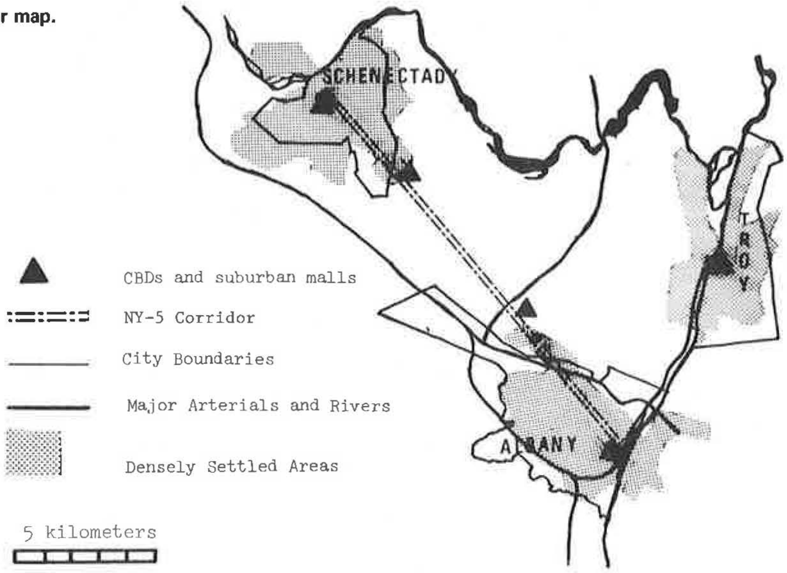


Table 1. Ridership of a composite weekday along the NY-5 corridor.

Route Type	Riders per Weekday by Link: In Albany	In Schenectady	Albany to Schenectady	Suburb Based	Bus Trips per Day	Riders per Trip
Through local	2691	1005	1473	746	106	56
Through express			288	54	6	57
Albany local[a]	2048				126	26
Schenectady local[a]		173			28	23
Total	4739	1178	1761	800	266	32[b]

[a] These local runs include those that may not run on the NY-5 corridor entirely, but only riders between corridor points are included.

[b] Average riders per day = corridor grand total (8478) ÷ total bus trips per day (266).

each side of NY-5, where the malls at the Albany end have arisen;

12. Possible routings of through buses into malls during off-peak periods;
13. Bus hardware problems, including power for hills, seat comfort, and vehicle capacity; and
14. Length of downtown layover times at control points.

The express service problem has been alluded to. The one morning express bus was heavily loaded and good loads on the evening returns indicated the demand for added morning service. This assymetry was probably due to CDTA's hesitation in adding more expresses that, despite good loads, cost more as tripper runs and did not have the productivity of the NY-5 through runs. The evening express buses were added only to relieve an evening loading problem at the Albany end.

The lack of late feeder service at the Schenectady end resulted from earlier decisions, when the system was under private management, to provide no evening local service beyond that required for the shifts at the General Electric plant. This resulted in some constraint on the use of the Albany through services since, in Schnectady, after 6:00 p.m., an Albany-bound commuter who left at 5:00 p.m. would not find a connection available. The remainder of the problems deal with general delay problems on the route and various service adjustments.

ANALYSIS OF SOME OF THE PROBLEMS

A number of detailed recommendations were formulated to address each of the listed problems; however, because of space limitations, only those features of the analysis that will have general interest are discussed here.

Cost Model and the Express Service Problem

From the consumer's viewpoint, the Schenectady-to-Albany express service has an obvious attraction over the local service. The 51-min travel time, end to end, is 27 percent shorter than the travel time for the local service. However, because of the long schedule time, it is not possible to get two useful peak-direction runs in each peak period. The express buses must therefore be run as trippers (short runs that cannot usually be combined into a full shift) and, therefore, accrue extensive operating overhead. This can be seen from the cost model that was applied to CDTA services:

$$\text{Cost per run} = (\$7 + \$17/\text{peak vehicle})/\text{run} + [\$10.69/\text{h (or } \$13.24/\text{overtime h)}]/\text{run} + (\$0.22/\text{km})/\text{run}.$$

Although this cost model is only a good approximation of all the variables associated with a run cost, it does show the cost increments that would be associated with a tripper run in the peak. Note that the overhead value accounts for costs of maintaining the pool of extra drivers and that the hourly overtime rate is not quite 1.5 times the straight rate because certain fringe benefits are not multiplied. The cost of an evening peak express run, with overtime rates applying to trippers, would be as follows:

Cost per evening peak run = \$24 + (\$13.24 × 2.32 h)/run + (\$0.22 × 74 km)/run = \$70.82.

The cost of the morning peak express run is less because the peak vehicle overhead does not apply and because the deadhead is partly allocated to a local run on the Albany end. CDTA has the unusual circumstance of having the heaviest peak in the morning at the Albany end (because of school trips) while the evening peak is the heaviest in Schenectady. Therefore, there is a happy synergy in having the incoming morning expresses to Albany (the expresses do not make a working run in the nondominant direction) help out the peak load of the Albany-based local buses. The morning express run cost is then \$42.96. Given this cost differential, it is even harder to understand the historical circumstance of three evening express runs and only one morning run, but it derives in part from the smaller evening load distribution that, before being remedied by better local and express phasing, was solved by added capacity.

The cost and revenue comparison between an express run and a local through run can then be made as follows:

Run	Cost/Run ($)	Revenue/Run ($)
Morning express	42.96	46.08
Evening express	70.82	40.71
Local (straight run, 6 half trips)	144.07	147.84

This comparison actually makes it harder to see why the evening express buses were favored over the morning bus. Although the morning express bus does show an apparent profit, added express buses would probably not be as heavily loaded and would, in the short run, divert revenues from the local service. Nonetheless, the increases in ridership justified the added express service on a long-term basis. At least one run per peak was also recommended that would run wholly on expressway between Albany and Schenectady. Although it would lose the important suburban ridership, the faster expressway run would allow two trips per peak and thereby reduce costs.

Overload Problems and Larger Buses

The evening peak buses in the corridor were overloaded on some runs that were approaching the 150 percent load criterion. The schedule already contained some expensive trippers (in addition to the evening expresses) to attempt to handle the problem. The problem was most acute for the through buses, which tended to pick up both the local and through traffic, particularly when through- and local-service phasing went awry. In fact, a modest change in through and local scheduling was quite effective in addressing this overload problem. But at the time, some attention was given to the economies that might be provided by larger, articulated buses.

The articulated vehicles have capacities of 60 seated passengers (depending on interior layout) compared to about 45 seats on the standard coach. However, the articulated vehicles would also have added costs, such as the cost-per-kilometer component for the heavier vehicles, possible driver premiums, deadhead times for bus changes, and garage-modification costs. It was also assumed that added capacity per vehicle would allow the immediate deletion of only some trippers and that the savings for larger vehicles would only accrue over time as loads increased. At the historical 5 percent rate of growth in ridership (although this increased to 10 percent over 1978 figures during the 1979 gasoline crisis period), there would only be about a 16 percent saving in corridor costs after 15 years. Thus, articulated vehicles would not be a panacea, and a capital-operating cost trade-off needs to be carefully studied in their application.

Through and Local Schedule Coordination

Some 53 percent of weekday rides on the corridor occur on the Albany end, where the through and local services overlap. Closed-door operation for the through vehicles had been abandoned in the interests of better headways, and so a critical issue was the distribution of riders between the two services.

The checksheet data showed that the loads between the local and through buses were, in fact, uneven. In the evening period, which was the most critical for loading, the local buses showed only 50-60 percent of the average loads of the through buses. Since the through buses were experiencing standing loads, this represented a serious maldistribution of service usage. Since fully 43 percent of the through bus riders were riding within the local zone, there was clearly room for diversion of riders.

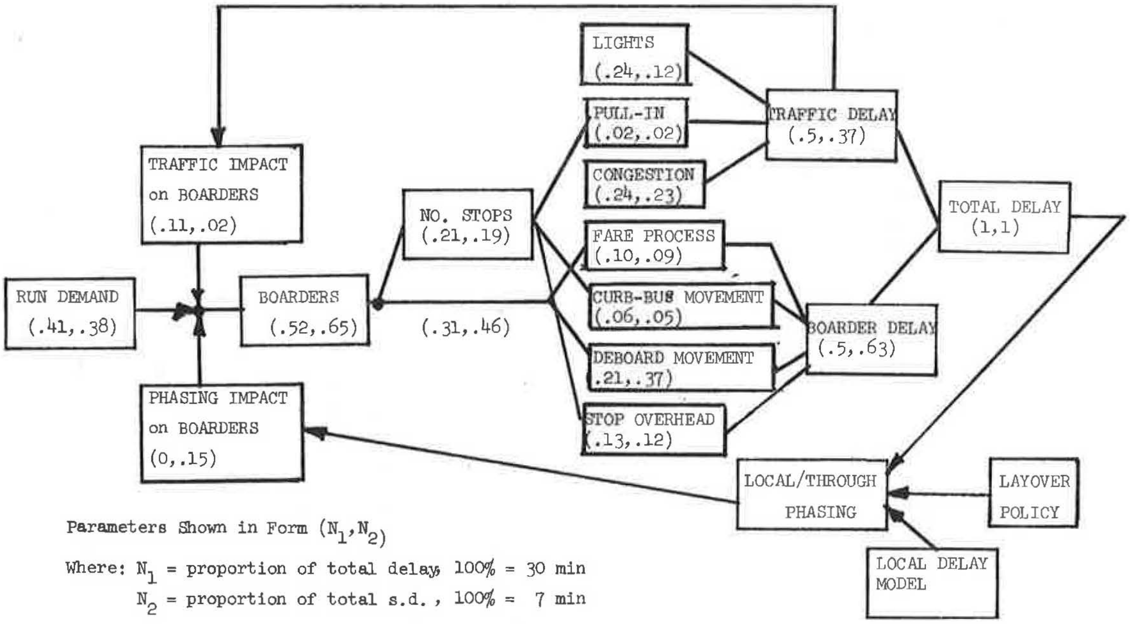
As it happened, a schedule change was introduced into corridor service just prior to the commencement of the study. The effects of this change, motivated primarily by supervisor suggestion, proved to be quite effective and showed the main causes of the load problem. The first problem was that the schedule check points were too tight for the through buses. Adjusting these by allowing some additional scheduling time as well as changing intermediate points provided for better arrivals at control points and, therefore, improved ability to phase the through buses with local buses which, because of their shorter trips, were inherently better in keeping time. However, the main change was in shifting the scheduled time of the local departures slightly. Previously, the through buses had the terminal departure times set for the major work-leave times, such as on the hour and half hour. By preceding the through trips by 5 min, the local buses were probably missing the times when most workers arrived at the stop. By shifting the local leaves up by about 5 min, they were put into the prime position for picking up workers.

Checksheets after the schedule change showed that the loading pattern was much improved. Evening peak loads before the change were 45 on the through runs and 22 on the local runs and changed to 43 and 40, respectively. Of evening peak corridor riders, the local buses previously carried 37 percent and now carried 44 percent. The results show a very simple principle, often overlooked: The absolute schedule timing of two services to be coordinated is at least as important as the relative phasing.

Schedule Delay and Variance

As a long run over heavily used urban arterials, the NY-5 buses showed significant problems in adhering to schedule. CDTA staff debated about whether there were effective operational measures to alleviate this

Figure 2. Causation tree of delay in through buses.



Parameters Shown in Form (N_1,N_2)

Where: N_1 = proportion of total delay, 100% = 30 min

N_2 = proportion of total s.d. , 100% = 7 min

or whether the primary factors were beyond control. Part of this debate was whether or not schedule time changes were sufficient to solve the problem.

Analysis of the schedule-adherence data showed that the average lateness for various trips was between 2 and 6 min but that the standard deviation of these means were from one to two times the mean values. This indicated that, in fact, the variance was large and lateness could not be solved by timetable adjustments. In fact, when the schedules were moved back just prior to the study, a number of early arrivals began to show up. A reasonable rule is that average lateness should be about the standard deviation in arrival deviation.

In order to attack the basic causes of delay and variance, measurements were then made of separate bus delay components, and from them the variances were modeled. The result is Figure 2, which shows the causal relations of the various components. Based on this, estimates could be made of the effects of various strategies for delay reduction on running time and schedule adherence. Two solution packages were postulated: one that took a Draconian approach to solving traffic and other problems and a more realistic package. It was estimated that even the most severe, and probably infeasible, strategy would delete 67 percent of delay on the through services, but this was only 27 percent of total travel time. The more realistic, but still ambitious, package would delete about 32 percent of delay. It was decided that, although some of these improvements were worthwhile, they were nonetheless long-term projects (e.g., traffic intersection improvements) that were not wholly under CDTA auspices.

A contention existed that there might be a vicious circle involving schedule-adherence problems, phasing of local and through buses, overloading of the through buses, and more delay. If this were the case, then small delay improvements would have amplified effects on corridor performance. However, examination of correlations between schedule delay and loadings did not reveal convincingly that this effect was at work, which was somewhat surprising. It appeared rather that, despite the supposed regularity of trip makers in when they appear at the bus stop, the basic variance in demand at stops is what is responsible for loading variance on individual trips.

CONCLUSION

The approach in this study was to handle many site-specific problems by use of modest analytical techniques. Yet the study showed that many common wisdoms derive from an operations-research approach. Wider dissemination of case-study analyses derived from transit authority staff work and consultant studies might be a useful service.

Publication of this paper sponsored by Committee on Bus Transit Systems.

Abridgment

Note on Bus Route Extensions

Daniel K. Boyle

This paper investigates the circumstances under which extensions of bus routes can be feasible. Ten recent route extensions in Albany and Rochester, New York, are examined with respect to ridership generated, length and frequency of service, type and size of the new population served, and additional operating cost. The extensions included extensions to new residential and industrial sites, reverse commute services, and services to major employment sites. A simple revenue/cost ratio for bus route extensions is used to compare the results. The paper concludes that route extensions that are most likely to be successful (a) are short, (b) serve a dense area of concentrated employment or residences, and (c) do not increase main-route headway.

Decisions on proposals for route extensions are generally made on an ad hoc basis because no criteria have been developed to guide the decision-making process. A few transit operators have set standards, which involve measures such as passengers per vehicle kilometer or revenue/cost ratio, that must be met within a certain period of time if a route extension is to be made permanent (1). Some areas have used relatively sophisticated measures (such as transit access time) to serve as a basis for service decisions. Others use analogies to similar areas where transit service already exists to judge whether a route extension is justified. Often, the results are less than satisfactory; route extensions are frequently abandoned because the ridership does not materialize.

Ideally, a cost/benefit model could be constructed and used to determine the relative merit of any route-extension proposal. This report is a first step in that direction. Four route extensions in the Albany, New York, area (Capital District) and six extensions in Rochester, New York, are examined to see how different areas respond to improved access to public transit. This study gives a preliminary indication of which types of extensions are most efficient in terms of benefit/cost analysis. It also provides a basis for development of a route-extension model.

DATA

Of the 10 route extensions, 4 served new residential developments and 2 routes were extended to hospitals, 1 to an industrial park, and 1 to a commercial area. The remaining 2 extensions were taken as efficiency measures to make routing patterns or turnaround at the end of the route more convenient.

Seasonally adjusted ridership data were readily available for all routes. Three routes had headway changes associated with the route extensions. In these cases, appropriate service elasticities (2) were applied to determine the ridership change due to change in service along the original route. This change was subtracted from the change in total route ridership to arrive at the ridership change associated with the route extension.

Data on households, population, employment, and land area were collected by traffic analysis zone (TAZ). Most extensions involved only one zone, but in certain cases data for two or more zones were needed. Table 1 presents ridership and demographic information for each route extension.

The heavy ridership loss on route 80 demonstrates how outside factors can overshadow minor changes in ridership that result from route extensions. An extension to a hospital in Rochester had no effect on ridership, but a significant increase in ridership resulted from a similar extension in Albany (although the Albany extension served an established residential neighborhood as well as a hospital). A significant industrial employment area shows the same ridership response as a minor commercial employment area. Ridership increases in residential areas do not match up well with population or population density.

The size of the analysis zones may explain the lack of consistent trends. Although preferable to census tracts, these zones still encompass a larger area than is actually served by the extensions. The extensions are targeted for specific developments; a measure of population and employment on a smaller scale would be more conducive to analysis of the relationship between transit ridership and various demographic figures.

Information concerning income, number of automobileless households and the like can, in all probability, aid in explaining different responses to route increases in otherwise demographically similar areas. At the current time, the socioeconomic data base in the Capital District and Rochester is incomplete, but further research should yield fruitful results.

The data obtained for route extensions are sufficient to determine revenue and cost changes. A revenue/cost ratio (R/C) can be calculated as follows:

$$\Delta R/\Delta C = (\Delta \text{ ridership} \times \text{ average fare})/(\Delta \text{ vehicle kilometers} \times \text{ average operating cost per vehicle kilometer}) \qquad (1)$$

Changes in both ridership and vehicle kilometers were calculated for an average weekday. In calculating revenue, $0.40 was used as the average fare in the Capital District (where $0.40 is the base fare) and Rochester (average of peak and off-peak fares). For park-and-ride express routes, an average fare was estimated based on the fare structure.

Average operating costs per vehicle distance traveled were obtained for the Capital District and Rochester areas from the latest transit operating assistance report (3). In Rochester, average operating cost was $1.13/vehicle-km ($1.81/vehicle mile); the Capital District figure was $0.86/vehicle-km ($1.38/vehicle mile). These averages overestimate the actual cost of operation in the route-extension area because most extensions are in outlying areas where average speed is higher; therefore, the operating cost per vehicle kilometer is lower than the systemwide average. However, the degree of inaccuracy introduced by the use of average operating cost per vehicle kilometer is slight and is outweighed by the ease of calculation.

REVENUE/COST ANALYSIS

The results of the R/C calculations are presented in Table 2. Of the 10 extensions, 4 showed a $\Delta R/\Delta C$ ratio greater than 1.00, which indicates that these not only paid for themselves but showed a profit. Two of the four profitable route extensions occurred on park-and-ride routes and were targeted for employee concentrations. The PR 2 extension served a reverse-commutation demand by bringing workers to a surburban office location from Rochester. The PR 1 and 2 extensions provided service within a major industrial park, thus eliminating a long walk for transit patrons.

Table 1. Route extensions, ridership, and demographic data.

Route Number	Length of Extension (km)	Δ Average Weekday Ridership	Population in TAZ	Household Density in TAZ (households/km^2)	Employment Density in TAZ (employees/km^2)	Purpose of Extension
Capital District						
80	1.1	-599	1 900	277	37	Convenience
82	0.8	2	2 010	423	91	Residential
84	3.2	105[a]	10 077	1184	2103	Hospital and residential
89	1.6	142[a]	4 965	314	353	Residential
Rochester						
21	1.6	40	831	62	65	Residential
PR2	1.1	40	732	59	74	Commercial
3	1.1	0	5 916	519	348	Hospital
4	0.6	2[a]	4 353	692	153	Convenience
PR 1 + 2	4.2	40	4 704	647	6983	Industrial
RIT	0.4	50	1 005	49	178	Residential

Note: 1 km = 0.62 mile; 1 km^2 = 0.39 $mile^2$.

[a] Adjusted to take headway changes into account.

Table 2. Revenue/cost ratios for route extensions.

Route Number	Δ Ridership	Δ Revenue ($)	Δ Vehicle Kilometers	Δ Operating Cost ($)	$\Delta R/\Delta C$
Albany					
80	-599	-239.60	59.4	51.20	-4.68
82	2	0.80	19.2	16.56	0.05
84	105	42.00	57.6	49.68	0.85
84[a]	31	12.40	-70.0	-53.00	-0.23
89	142	56.80	20.8	17.94	3.17
Rochester					
21	40	16.00	35.2	39.82	0.40
PR2	40	26.00	2.3	2.57	10.12
3	0	0.00	8.6	9.70	0.00
4	2	0.80	68.6	77.65	0.01
PR 1 + 2	40	40.00	8.4	9.45	4.23
RIT	50	20.00	4.2	4.71	4.25

Notes: 1 km = 0.62 mile.
All data measured for an average weekday.

[a] Taking into account effects of associated headway changes on existing portion of route.

The RIT and route 89 extensions served residential areas; it is hypothesized that certain socioeconomic variables in the extension area can account for the success of these extensions. Local conditions might also account for different responses to extensions. For example, the route 89 extension improved transit access in an area that had a significant concentration of public housing and no sidewalks. Previously, it had been very difficult to walk the distance to the bus stop; the extension brought service to a large pool of likely transit users. Knowledge of such local conditions is both useful and necessary in judging the relative merits of a specific extension.

A general R/C model of the form of Equation 1 is suggested for use in evaluating route extensions. In cases where extensions have been put into effect, use of this model is straightforward. For potential extensions, a method must be developed for estimating changes in ridership. This might be of the form:

$$R = p \times f(a) \qquad (2)$$

where p is the pool of potential transit users (e.g., number of residents in a new housing development or number of workers in an industrial park) and f(a) is an attraction function dependent on socioeconomic data and quality of service. Different f(a)'s could be developed for different land use areas. With such functions, the feasibility of a route extension in a given area would depend on the size of the potential transit pool, relevant socioeconomic data, the length of the extension, and the quality of service offered.

The model allows for calculation of R/C ratios; however, criteria for judging the success of a route extension is subjective. A profit criterion would stipulate that the route extension be taken if $\Delta R/\Delta C$ is greater than or equal to 1.0. Alternatively, an equal subsidy criterion might suggest that the route extension be taken if $\Delta R/\Delta C$ for the extension is greater than or equal to the R/C ratio for the existing route or if $\Delta R/\Delta C$ for the extension is greater than or equal to the R/C ratio for the entire transit system.

SUMMARY AND FUTURE DIRECTION

It is clear that the success of route extensions depends heavily on the character of the area to which the extension is made. Land use, population, population density, income, and number of automobileless households are some of the variables hypothesized to be salient in determining the response to a route extension.

The census tract or the traffic analysis zone is too large to be used as the geographic base for the collection of demographic data. Since most extensions are targeted for a specific residential development or employment concentration, demographic data are needed on an approximately small scale.

The form of the R/C model indicates that, in similar land use areas, the success of a route extension depends directly on the size of the pool of potential transit users and inversely on the length of the extension. Obviously, a short extension to an area that has a large residential or employment population is most efficient in terms of the R/C ratio.

In cases where headway on a given route must be increased due to an extension of the route, the decline in level of service along the original portion of the route serves as a counterbalance to the new service on the route extension. The quality of service may decline along with the quantity of service as the original portion of the route becomes more crowded. For route extensions that have corresponding headway increases, the riders on the existing route are in effect subsidizing the extension through a decline in service on the original portion of the route.

Four of the 10 route extensions showed an R/C ratio greater than 1.0. An examination of each of these successful extensions highlights various factors discussed previously.

The PR 2 extension had the highest $\Delta R/\Delta C$. This extension was made to a commercial area in order to serve reverse-commutation trips from Rochester. This extension of an express park-and-ride line was the only 1 of the 10 extensions to serve a commercial area. It has been suggested that high-quality transit service at a high price is most likely to be self-supporting.

The PR 1 and PR 2 extension into Kodak Park also proved to be profitable despite the fact that it was the longest of the 10 extensions. The previous comment concerning high-quality service is also applicable here. Extensions to areas of significant employee concentration appear to be most promising in terms of R/C ratio.

The RIT extension to a residential area was the shortest of the 10 extensions. This demonstrates the importance of the length of the route extensions.

The route 89 extension brought service within easy reach of public-housing residents, many of whom are captive transit riders. Local factors also contributed to the positive ridership response to this extension.

In conclusion, size of population, type of land use, quality of service, and length of extension are four major factors in the determination of the success of route extensions. Areas that have a significant concentration of employees seem most likely to support profitable extensions. Special local conditions can also influence ridership changes connected with route extensions. A general R/C model can be used to evaluate route extensions, and the criteria used to judge extensions can be left to the discretion of local operators. The problem of increased headways associated with route extensions resulting in a decline in service on the original portion of the route must be taken into account when it arises. Finally, conventional units of data collection (such as census tracts or TAZs) are too large for the purpose of evaluating route extensions.

Directions for further research in the area of route extensions are clear. Collection of data on a small scale commensurate with the area actually served by an extension and explicit correlation of these data with changes in ridership are the immediate next steps to be taken. The development and testing of attraction functions for different types of land use follow these steps. A general predictive model of the effects on transit ridership of route extensions can then be constructed. This paper has suggested the basics for such a model and has provided preliminary findings concerning the most salient factors in determining the outcome of a proposed route extension.

REFERENCES

1. J. P. Attanucci, L. Jaeger, and J. Becker. Bus Service Evaluation Procedure: A Review. Urban Mass Transportation Administration, Interim Rept., UMTA-MA-09-7001-79-1, March 1979.
2. D. L. Weiss. The Effects on Automobile Energy Consumption of Transit Fare and Service Changes. New York State Department of Transportation, Albany, memo, Feb. 16, 1978.
3. Public Transportation Operating Assistance Programs in New York State: 1978 Annual Report, Technical Supplement. Planning Division, New York State Department of Transportation, Albany, March 1979.

Publication of this paper sponsored by Committee on Bus Transit Systems.

Hierarchical Procedures for Determining Vehicle and Crew Requirements for Mass Transit Systems

Lawrence D. Bodin and Robert B. Dial

This paper presents procedures for determining vehicle and crew requirements for mass transit systems. Some of these procedures are very fast computationally but only give lower bounds, upper bounds, or estimates of resource requirements. Other procedures are slower computationally but give actual crew and vehicle schedules. Depending on the type of analysis being performed (long-range planning, short-range planning, or operational planning), all of these procedures play a useful role in the design and analysis of proposed mass transit systems. The paper has two sections: (a) the first discusses techniques for determining vehicle requirements and (b) the second discusses techniques for determining crew requirements. Within each section are a set of procedures that range from the very simple to the complex, along with comments on their usefulness and shortcomings.

The design of mass transit systems occurs in various planning scenarios: long-range planning (5-20 years in the future), short-range planning (1-5 years in the future), and operational planning (less than 1 year in the future). The long-range planning analyst does not need (and cannot afford) the same information on crew and vehicle requirements as the operational planner. Whereas the operational planner needs actual feasible crew and vehicle schedules, the long-term planner may only need an estimate or lower and upper bounds on total crew and vehicle requirements for the analysis. Thus, the long-range transit planner should use fast crude estimation procedures to help evaluate a proposed transit system, since he or she may consider scores of alternative transit systems in attempting to find the optimal system.

In this paper, hierarchical procedures for determining crew and vehicle requirements are given. Some procedures require only manual calculations and furnish inexpensive (albeit crude) estimates. Others consume a significant amount of computer time and give more accuracy and detail. As will be seen, if the planner requires a more exact or more detailed vehicle or crew schedule, a higher cost must be absorbed in terms of computer time and human effort.

Table 1. Timetable 1.

	Time		Location	
Trip	Start	End	Start	End
1	7:03	8:13	l_1	l_2
2	7:18	8:28	l_1	l_2
3	7:35	8:45	l_1	l_2
4	7:48	8:58	l_1	l_2
5	8:04	9:14	l_1	l_2
6	8:18	9:28	l_1	l_2
7	8:35	9:45	l_1	l_2
8	8:48	9:48	l_1	l_2
9	9:04	10:14	l_1	l_2
10	9:18	10:28	l_1	l_2
11	9:35	10:45	l_1	l_2
12	9:48	10:58	l_1	l_2
13	10:03	11:13	l_1	l_2

The results in this paper evolved out of the design and implementation of program UCOST (1) for the Urban Mass Transportation Administration (UMTA). Many of the procedures discussed here have been implemented or will be implemented within the various procedures contained in the Urban Transportation Planning System (UTPS) (2) or within future computer-based transportation planning systems to be implemented and distributed by UMTA. A more detailed description of these procedures can be found in Bodin and Dial (3).

All of the procedures described for the determination of vehicle requirements and line-by-line analysis for estimating crew requirements have been used in Dade County, Florida, for the design of the bus system that is to feed the proposed urban rail system. Those procedures allowed for the myriad of possible feeder bus systems to be reduced to a few by finding reliable capital and operating cost estimates. Some of the procedures have not been used in the field as yet: The histogram approach is included as part of program UCOST and the interactive procedures are currently under development. The RUCUS system (4) has been modified by several organizations and has been used with varying degrees of success in several cities.

BASIC STRUCTURE OF A TRANSIT SYSTEM

A transit system can be depicted by a set of transit lines that presents data for each line in one of two ways. The first way gives a timetable (headway sheet) for the system that shows, for each trip in the timetable, its line number, start time, end time (including layover), start location, and end location. This is the kind of data RUCUS (4, 5) requires as input. Preparation of the data in order to depict the transit system in this detailed manner is expensive.

The second way gives the length of time to cover any trip on the line, the time between runs on the line (called the headway for the line), and the start and end locations for each trip on the line. To take into account variable traffic patterns and demands for service, both the time to cover a trip and the headway for the line can be a function of time of day. The second way costs less to prepare but does not specify a timetable directly.

In long- and short-range planning, an actual timetable may not be necessary in order to perform the desired analysis. Moreover, Bodin and Rosenfield (1) showed that the determination of a well-designed timetable (in terms of passenger transfer times) from the line data specified in the second way is a challenging computational exercise. However, a daily timetable (which may not be well designed) can be quickly generated in the following manner. The first run of each line in a time period can be assumed to begin at the start time of the time period. Then, the other runs for the line in the time period are found by increasing the start and end times of the previous run for the line by the headway. The timetable generated in this manner may be unsatisfactory for operational planning since the lines are not synchronized, but this timetable may be adequate for long-range planning and for some short-range planning exercises.

The two ways of depicting a transit line can be illustrated as follows. A timetable for a line in a period is given in Table 1 (this will be referred to as timetable 1 in the remainder of this paper). The headway between adjacent trips in a timetable need not be the same; therefore, the start and end times for each trip in the timetable must be specified. Since many transit systems have several thousand trips, the preparation of the data (unless the headway for a line is constant) can be a significant undertaking.

In this paper we attempt to demonstrate what a planner can discern about crew and vehicle requirements when only headway information for each line is available. Furthermore, we attempt to show what additional information can be determined about crew and vehicle requirements when an actual timetable of trips is available. Finally, we assume that the layover time is a requirement of the system and is included in either the start or end time of the trip if a timetable is given or as part of the time to cover the trip if a timetable is not specified.

DETERMINATION OF VEHICLE REQUIREMENTS

Procedures for determining vehicle requirements range from a simple procedure that can be done manually to a complex optimization procedure that requires a computer. Although these procedures are not the only approaches available for determining vehicle requirements for transit systems, they illustrate a hierarchical approach to this problem and demonstrate the additional information gained by using a more complete (i.e., costly) approach.

Maximum Number of Vehicles: Line-by-Line Approach

The vehicle requirements for each line in a proposed transit system are estimated as follows:

$$\text{Vehicles for line i in time period} = \lceil \text{Time to cover a trip} \div \text{headway of line i in time period} \rceil \quad (1)$$

where $\lceil x \rceil$ is the smallest integer greater than or equal to x. The number of vehicles to service the entire transit system in a time period is the sum over all the lines in the system of the number of vehicles needed to service each line as found in Equation 1. Thus, if the time to cover a trip is 70 min and the headway is 15 min, then an estimate of the number of vehicles needed to service the line in the time period is $\lceil 70/15 \rceil = 5$.

This quick procedure is useful for quick determination of a maximum number of vehicles (i.e., a capital requirements analysis). As such, this analysis need only be performed over the peak time periods. The vehicles required are the maximum of the vehicle requirements needed in each of these time periods.

If a vehicle is to service trips in both directions of a two-way line (or a trip on one line followed by a trip on the second line), then the time to cover a trip on the line is equal to the time to service a trip in each direction plus the time that the vehicle needs for turning around at each end of the line. For one-way lines, the time to cover a trip on the line is equal to the time to traverse the line in one direction plus the time to deadhead back to the beginning of the line plus the turnaround times.

Computation of the deadheading time for all pairs on terminal points can be an expensive enterprise. Therefore, an estimate of deadheading time (as a linear function of distance) may be appropriate for this procedure.

This procedure can determine feasible vehicle schedules if a vehicle is restricted to servicing only one line and the input data satisfy the requirements listed above. If the requirements are satisfied but the vehicle is allowed to deadhead between ends of the lines (i.e., service more than one line), then the above procedure is an upper bound on the estimation of vehicle requirements. If the deadheading and turnaround times are not known, then the procedure gives an estimate of vehicle requirements, but not necessarily an upper bound. In this situation, the procedure may underestimate vehicle requirements.

If the assumptions above are satisfied but deadheading between lines is allowed, the resulting upper-bound estimate of vehicle requirements may be a considerable overestimate of actual vehicle requirements. This procedure can be performed manually.

Example

Let line 1 have a headway of 15 min and a duration of 70 min and let line 2 have a headway of 15 min and a duration of 50 min. Assume that the end location of line 1 is the start location of line 2 and vice versa. Furthermore, assume that it takes 22 min to deadhead from the end locations of each line to its beginning location and assume that the turnaround time at the ends of the line is 4 min. The following estimate of vehicle requirements can be made.

Vehicles for lines 1 and 2 together $= \lceil (70 + 4 + 50 + 4) \div 15 \rceil = 9$

Note that, in the case of a trip that has the same start and end locations, the deadhead time does not enter the computations.

Lower Bound on Vehicle Requirements: Histogram Approach

In the histogram approach, a timetable must be specified. In this timetable, the start and end times for each trip are known. Let a 1440-strata (= 60 min x 24 h) histogram be specified where stratum i corresponds to the ith minute of the day. If trip j starts at time k and ends at time e, then a vehicle is required for the k, $k + 1$, $k + 2$, ..., $e - 1$ minutes of the day. In this case, 1 is added to the values of strata k, $k + 1$, $k + 2$, ..., $e - 1$ in the histogram. The above procedure is repeated for all trips in the timetable. Let m_i be the number of vehicles required in the ith stratum (i.e., the ith minute of the day) and let $M = \max_i (m_i)$. Then M is a lower-bound estimate of the number of vehicles required.

M is a lower-bound estimate because M denotes the maximum number of vehicles required by the timetable but fails to consider any deadheading or dead time that may require additional vehicles in an operational schedule. When actually scheduling vehicles, it may be necessary to deadhead a vehicle over the stratum that designates the peak number of vehicles. Hence, this procedure gives a lower bound.

If the planner only wants to estimate the vehicle requirements, then this analysis need only be performed over the peak periods.

We have found that, in many cases, M is a surprisingly accurate estimate of actual vehicle requirements.

If all lines operate over the entire time period (i.e., no special trips needed over a small portion of the time period), then the results of this procedure are (for the most part) independent of the timetable used. In this case, a timetable may not be required to use this approach and only line data are used as specification of the transit system.

Example

Part of the histogram for timetable 1 is given below. The histogram oscillates between 4 and 5 until 10:12, when it begins to damp out. The peak number of vehicles estimated is 5.

Time Interval	No. of Vehicles
7:03-7:17	1
7:18-7:34	2
7:35-7:47	3
7:48-8:03	4
8:04-8:12	5
8:13-8:17	4
8:18-8:27	5
8:28-8:34	4

The line-by-line analysis, which generally gives an upper bound on vehicle requirements, may not be accurate, but the results can be found without having to use a computer program. The histogram approach generally gives a lower bound and is accurate, but it needs a simple computer program to derive the desired estimate. For capital cost estimation, both procedures can be used in a sketch-planning mode, and the histogram approach can be used in a short-range planning mode. The concurrent scheduler (described next) should be used in a short-range planning mode if both a capital cost analysis and operating cost estimate are needed.

Feasible Vehicle Schedules: Concurrent Scheduler

The concurrent scheduler is a straightforward heuristic that creates a feasible vehicle schedule (set of blocks) for a given timetable. The trips for all lines are merged together and are sorted by starting time from a specified beginning time of day. Although which beginning time of day to select for a 24-h timetable is not obvious, we have found that the results for the concurrent scheduler and the Dilworth chain decomposition procedure are not greatly affected by this beginning time of day, as long as the beginning time of the day is in an off-peak time period.

The concurrent scheduler operates as follows:

1. Orders the trips in the timetable by time of day; call this list of trips the sorted list;
2. Assigns trip 1 in the sorted list to vehicle 1 (i.e., block 1);
3. Assumes that the first k trips in the sorted list have formed m partial vehicle schedules (blocks). Then, it is possible to assign trip $k + 1$ in the sorted list to partial vehicle schedule n, $n = 1, 2, \ldots, m$ if (a) $E(n)$ = start time for trip $k + 1$ - end time for partial vehicle schedule $n \geq$ some minimum time as specified by the planner and (b) $E(n)$ + safety factor $\geq$ time to deadhead from the end location of partial vehicle schedule n to the start location of trip $k + 1$;
4. If trip $k + 1$ can be assigned to more than one partial vehicle schedule, then the scheduler assigns the trip to the partial vehicle schedule that minimizes $E(n)$, $n = 1, 2, \ldots, m$ or to the first partial vehicle schedule found that satisfies the above conditions;
5. If trip $k + 1$ cannot be assigned to any partial vehicle schedule, then it creates a new partial vehicle

schedule m + 1 beginning with trip k + 1; and

6. Repeats steps 2-5 for all trips in the sorted list.

The above procedure gives vehicle schedules that are feasible but not necessarily optimal. An example is presented in Bodin and Dial (3) that illustrates this point. The Dilworth procedure (6) discussed in the next section determines a minimum number of vehicle schedules for a given timetable.

The procedure is very fast computationally because it has to pass only once through the sorted list of trips.

If only capital requirements are needed, then the concurrent scheduler need only be applied to the trips in each peak period and the maximum selected as the peak requirements.

Example

Let timetable 1 be specified in Table 1 (where l_1 is the start location for each trip in timetable 1 and l_2 is the end location). Furthermore, let the timetable for line 2 (called timetable 2) be specified in Table 2. The deadhead times are as follows: $d(l_1, l_2) = d(l_2, l_1) = 22$, $d(l_1, l_1) = d(l_2, l_2) = 0$. The turnaround time at the end of each trip is 4 min. The sorted timetable and the vehicle assignments of each trip, by using the concurrent scheduler, are given in Table 3. The number of vehicles required by the solution to the concurrent scheduler is 10. The estimated number of vehicles as found in the line-to-line analysis is 9. This number can be attained by the concurrent scheduler if the start and end time for each trip in timetable 2 is increased by 4 min. Therefore, trip 1 for timetable 2 would be the following:

Time		Location	
Start	End	Start	End
7:04	7:54	l_2	l_1

Note that the estimated number of vehicles that use the line-to-line analysis was made independent of the timetable used, whereas the results from the concurrent scheduler were based on a timetable.

Table 2. Timetable 2.

	Time		Location	
Trip	Start	End	Start	End
1	7:00	7:50	l_2	l_1
2	7:15	8:05	l_2	l_1
3	7:30	8:20	l_2	l_1
4	7:45	8:35	l_2	l_1
5	8:00	8:50	l_2	l_1
6	8:15	9:05	l_2	l_1
7	8:30	9:20	l_2	l_1
8	8:45	9:35	l_2	l_1
9	9:00	9:50	l_2	l_1
10	9:15	10:05	l_2	l_1
11	9:30	10:20	l_2	l_1
12	9:45	10:35	l_2	l_1
13	10:00	10:50	l_2	l_1

Optimal Number of Vehicles

To derive feasible vehicle schedules (blocks) that minimize the number of vehicles needed, a procedure such as the Dilworth chain decomposition (6) must be used. The Dilworth chain decomposition finds the minimum number of chains needed to cover all the nodes of an acyclical directed network. Each chain corresponds to a vehicle schedule. The nodes in this network are the trips from the timetable, and the arc from node i to node j implies that it is feasible (vis-à-vis the conditions in step 3 of the concurrent scheduler) to service trip i and then trip j on a vehicle schedule. A description of the implementation of the Dilworth procedure for transit scheduling can be found in Bodin and Rosenfield (1).

The Dilworth procedure does not minimize deadhead requirements, and the solution from the concurrent scheduler (for the entire day) is a good starting solution to the Dilworth procedure. To our knowledge, there is no procedure available that can simultaneously minimize both vehicle requirements and deadhead distance. The vehicle scheduling procedure in the RUCUS computer system minimizes deadhead distance but not vehicle requirements, is much slower computationally, and only handles much smaller problems. The network in the RUCUS procedure is the same as the network in the Dilworth procedure except for the costs on the arcs of the network.

The minimization of vehicle requirements and then deadhead distance (given vehicle requirements) requires a two-step procedure. The first step performs the Dilworth procedure to minimize vehicle requirements. The second step uses the RUCUS vehicle scheduling procedure while fixing the number of vehicles to be allowed (as found in the Dilworth procedure). This is accomplished by fixing the lower and upper bounds on flow on the branch from the supersink to the supersource equal to the Dilworth solution and using the Dilworth solution as the starting solution from this minimum cost-flow problem.

To derive a solution that trades off between number of vehicles used and total deadheading requires that the RUCUS BLOCKS model be modified as follows. A relative weight is chosen to be associated with the number of vehicles; this weight reflects the value of a vehicle with respect to a deadheading unit (i.e., distance or time). This weight is used as the cost on the arc from the supersink to the supersource. The solution to the minimum cost network problem would then be the one that trades off vehicles with savings in deadheading. The cost of using this model would be essentially equal to that of using the present RUCUS model.

The three models described above give different answers to the same problem based on the objective the planner wishes to use. The planner must decide whether it is worth the investment in computer time to run either of the latter two models (the two-step model or the com-

Table 3. Sorted timetable.

Time		Location		
Start	End	Start	End	Vehicle Assignment
7:00	7:50	l_2	l_1	1
7:03	8:13	l_1	l_2	2
7:15	8:05	l_2	l_1	3
7:18	8:28	l_1	l_2	4
7:30	8:20	l_2	l_1	5
7:35	8:45	l_1	l_2	6
7:45	8:35	l_2	l_1	7
7:48	8:58	l_1	l_2	8
8:00	8:50	l_2	l_1	9
8:04	9:14	l_1	l_2	1
8:15	9:05	l_2	l_1	10
8:18	9:28	l_1	l_2	3
8:30	9:20	l_2	l_1	2
8:35	9:45	l_1	l_2	5
8:45	9:35	l_2	l_1	4
8:48	9:58	l_1	l_2	7
9:00	9:50	l_2	l_1	6
9:04	10:14	l_1	l_2	9
9:15	10:05	l_2	l_1	8
9:18	10:28	l_1	l_2	10
9:30	10:20	l_2	l_1	1
9:35	10:45	l_1	l_2	2
9:45	10:30	l_2	l_1	3
9:48	10:58	l_1	l_2	4
10:00	10:50	l_2	l_1	5
10:03	11:14	l_1	l_2	6

bined RUCUS model) rather than the Dilworth procedure in order to determine the trade-off between vehicle requirements and deadhead distance.

DETERMINATION OF CREW REQUIREMENTS

In this section, we present procedures for determining crew requirements for a proposed transit system. The crew requirements problem is more complex than the vehicle requirements problem because a vehicle can operate the entire day without a break, but a crew has specific work rules that restrict the total amount of work that can be done during the day. Furthermore, the cost of a crew depends on the type of shift worked, the length of the shift, the time of day worked, overtime, and so forth.

In a simplified model, there are three basic crew workdays: full-time shifts, split shifts, and tripper shifts. A full-time shift is a complete workday for a crew with one embedded short break for lunch. A split shift is a complete workday for a crew with an embedded longer break of several hours that splits up the workday. A tripper shift is a part of a workday and has no scheduled breaks. Since transit systems generally have a morning and evening peak surrounded by lesser requirements during the off-peak hours, tripper shifts and split shifts usually exist to service the peak periods, and full-time shifts are scheduled to handle the nonpeak demands in both the peak and off-peak periods. The cost of a crew is a function of the type of shift, the time of day that the shift works (generally associated with the starting time of the shift), and the length of the shift (i.e., overtime).

The crew scheduling problem can be thought of as a very large set-covering or set-partitioning problem. A description of the set-covering and set-partitioning formulations of this problem can be found in Bodin and Dial (3).

The RUCUS implementation and the set-partitioning and set-covering formulations of the crew scheduling problem represent one-shot batch-optimization procedures for solving this problem. In a batch procedure, all parameters that guide the solution process are set prior to the computer run itself. The computer program then finds a solution to the problem based on the parameters set and the data. We believe that a batch-optimization algorithmic procedure for solving the crew scheduling problem is computationally prohibitive in most cases. Therefore, the development of heuristic procedures or man-machine interactive procedures for solving this problem appears necessary. The heuristic procedures were discussed in detail at a meeting on operator scheduling (Workshop on Automated Techniques for Scheduling of Vehicle Operators for Urban Public Transportation Services, April 27-29, 1975). Many heuristics exist for solving the crew scheduling problem, including a particularly effective one that adapts the RUCUS system (4, 5). Because of the diverse nature of these heuristics, they will not be discussed in any detail in this paper.

It is possible, however, to develop procedures for simply estimating or bounding total crew requirements that do not depend on costly crew scheduling heuristics. Such estimates are invaluable for cost-estimation purposes and provide targets at which schedulers who use run cutting can aim. Procedures that can play a central role in the planning and operation of transit systems are discussed below.

Estimation of Crew Requirements Without a Timetable: Line-by-Line Analysis

In time period p, let L(i,p) be the duration of a trip on line i, including layover in minutes, and let n(i,p) be the number of trips on line i in the period. Then CR(i), which is the estimated number of crews who work a full shift on trip i, is found as follows:

$$Cr(i) = \left\lceil \sum_p L(i,p)\, n(i,p)/E \right\rceil \quad (2)$$

where E is the number of minutes in an effective workday for a full shift. E is discussed below. The number of crews who work a full shift (DR) is found as follows:

$$DR = \sum_i CR(i) \quad (3)$$

This approach can be performed manually.

If a line is to operate over the entire duration of period p, the duration of period p is D(p) minutes, and the headway of line i is H(i) minutes, then

$$n(i,p) = \lceil D(p)/H(i) \rceil \quad (4)$$

where $\lceil x \rceil$ is the smallest integer greater than x. Thus, if x is an integer $\lceil x \rceil = x + 1$.

E, the number of minutes in an effective workday, needs some clarification. Let T be the duration of the workshift for a full-time crew. Let each crew spend, on the average, t minutes in nonrevenue activities such as deadheading to and from the garage, lunch break, or time between runs. Then E = T - t is the number of minutes in a day that a crew spends on revenue activities (i.e., actually serving passengers). The revenue activities are the runs specified in the line schedule or the timetable.

It is difficult to estimate E without having actual crew schedules (i.e., it is often difficult to discover the time to and from the garages or the time between trips). To find E requires the use of previous experience with the transit system. As a rule of thumb, we have found that an estimate of E between 6 and 6.5 h provides reasonable estimates of crew size for a traditional bus operation.

Given a more reliable estimate of E and L(i,p), we can better estimate CR. Thus, any preliminary scheduling that can be performed is useful. For example, let line 1 go from node A to node B and let line 2 go from node B to node A and let both lines have the same headway. If the following crew schedule is to be run: line 1, line 2, line 1, and so on, then L(i,p) can be redefined as the time to complete the round trip and start out on the next available trip on line 1. In this case, dead-time information is embedded within the computation of required work time. Thus, in the expression E = T - t, t equals the time to deadhead to and from the garage to the specified start and end points of the lines plus the time for lunch. Since this definition of t gets rid of much of the variability attributable to crew scheduling, experience has shown that this estimate of E gives a more reliable estimate of n(i,p) than the estimate of E described previously.

In many cases, transit planning is performed one time period at a time. To discover an estimate of operating cost for a time period, we need the estimated equivalent number of crews who work a full shift in time period p, which we call CRP(p). CRP(p) is found as follows:

$$CRP(p) = \left\lceil \sum_i L(i, p)\, n(i, p)/E \right\rceil \qquad (5)$$

Since these time periods can be short in duration, no attempt is made to break down CRP(p) into a line-by-line analysis.

Let G be the cost per day of a full-time crew. Then GTOT [or GTOT(p)], the estimated operating cost attributable to the crews (or the estimated operating cost attributable to the crews in time period p), is given by

$$GTOT = G*CR \;[\text{and } GTOT(p) = G*CRP(p)] \qquad (6)$$

GTOT is a simple estimation procedure; it disregards pay differentials as a function of shift type and time of day, but it can give reasonable answers to crew size requirements with a minimal investment in data preparation and computer implementation.

Example

Suppose that timetables 1 and 2 are to operate from 7:00 a.m. to 7:00 p.m. The headways on both lines are 15 min, each trip for timetable 1 has a duration of 70 min, and each trip for timetable 2 has a duration of 50 min. Assume that it takes 30 min to get from the garage to either l_1 or l_2 and each driver is to get a 45-min lunch hour. If both lines can be serviced by the same crews, then the length of a trip = 70 + 4 + 50 + 4 = 128 min. Therefore, $CR(1 + 2, 1) = \lceil(128)(49)/375\rceil = 15$. If the cost of a crew is \$50/day, then the crew cost estimate in this example is (\$50) [CR (1 + 2, 1)].

The line-by-line analysis gives an estimate of crew requirements (assuming that each crew is full time) that may not be accurate, but the results can be found without a computer program. As such, it should be used in a long-range planning environment. The histogram procedure described in the next section gives a more accurate estimate of crew requirements, uses to some extent differing shift types in building its model, but requires a computer.

Histogram Procedure

The procedure in the previous section gives an estimate of crew size requirements assuming that crews only work a full-time shift. The procedure in this section gives the following:

1. Estimate of the number of crews needed by shift type (full-time shift, split shift, and tripper);
2. Estimate of the number of crews needed by time of day; and
3. Estimate of the total crew costs, taking into account pay differentials.

This procedure does not give actual crew schedules.

Input to this procedure is the set of aggregated trips or vehicle schedules. Each aggregated trip represents a collection of trips or blocks that must be serviced by a crew and vehicle. The concurrent scheduler or the Dilworth procedure can be used to create the set of aggregated trips. The aggregated trips are essential to avoid double counting the required number of crews and overestimating the number of crews required to service the transit system.

The first step in this procedure is to form a histogram (called the demand histogram) of the number of crews required to cover all the trips of the day. Input is the set of aggregated trips or blocks. To construct this histogram, the time of day is broken down into time intervals, where it is assumed that any blocks that fall into any part of these small time intervals require a crew to service them for the entire time interval. For example, a block that starts at 7:09 generates a crew requirement from 7:00 to 7:10 if a 10-min time interval is used in defining the histogram. Experimental evidence indicates that a 10-min time interval derives accurate estimates of crew size.

The crew estimation is based on allowable shift segments that the planner specifies. A shift segment consists of a consecutive number of work hours that a crew is to work. A shift segment specification is designated by the crew cost (including pay differential and overtime), first permissible time of day when crews can report to work on this shift segment, and last time of day when crews can report to work on this shift segment. The shift segment specifications form the alternatives on which the crew estimation is to be based. Split shifts are combinations of two shorter shift segments. Thus, if a crew is to work a split shift from 9:00 a.m. to 1:00 p.m. and from 5:00 p.m. to 9:00 p.m., then the crew estimation component assumes that the crew works two shift segments, each of 4-h length. One shift segment is from 9:00 a.m. to 1:00 p.m., and the second shift is from 5:00 p.m. to 9:00 p.m.

From the demand histogram and the shift segment specifications, a network is created. The out-of-kilter algorithm (6) is then employed to determine the number of crews of each type of shift segment that are needed to cover the demand histogram. A detailed description of this procedure along with an example that illustrates the procedure can be found in Bodin and Rosenfield (1).

This procedure gives an estimate of the number of crews required by time of day and an estimate of the number of crews to be assigned by time of day. Hence, the difference is how many crews are present but not assigned to a particular activity. The times of day when this difference is positive may be times of day when runs can be added to the timetable without requiring an additional crew. Hence, these runs can be serviced at no additional crew cost. This characteristic of the solution to this problem is useful in attempting to design a transit system.

The solution of the minimum cost-flow problem requires no more computer time than the generation of the network. Therefore, this procedure can derive an estimate of crew requirements in the same amount of computer time [about 10 s of central processing unit (CPU) time on an IBM 370/168] independent of the line schedule used.

A slightly more difficult problem is to allow the planner to place bounds on the number or percentage of crews of various types. This problem cannot be solved with a network-flow algorithm but can be solved with a moderate-sized linear problem (150 rows, 1500 columns). This latter model is a planned improvement to the UCOST software.

Optimal Crew Scheduling

We do not feel, given the current state of computing machinery, that the optimal crew scheduling problem, in general, can be solved by use of a batch algorithm. We do feel, however, that the problem would be solvable by using a man-machine interactive procedure. Such a procedure would begin with a feasible schedule composed of two subschedules. One subschedule would consist of those fixed crew schedules that cannot be altered and the other would consist of free schedules that could be changed. Also, certain partial schedules could be fixed as if they were a run and then joined as a block on a full-day schedule. The procedure would then improve the given schedule by manipulating the set of free schedules.

The procedure would then write the results of this analysis to a structured data base. The planner could then query the data base to find instances of crew schedules that were unacceptable. The planner would then alter the fixed and free subschedules to reflect complaints and reexecute the algorithm.

Mathematically this problem can be formulated analagous to the Dilworth chain decomposition procedure except that the length of each chain is restricted to be within certain bounds (to reflect length of shift requirements). The fixing of schedules corresponds to the forcing of a flow of one over certain branches in the network. The prohibition of the joining of two runs on a crew schedule corresponds to the fixing of a flow of zero on the appropriate branch. The Dilworth chain decomposition with length of chain restriction can then be solved over this smaller network.

Output from this analysis would then be a set of tripper shifts or half-day schedules. A heuristic or exact 1-match procedure (7) can then be employed to join these half shifts into full-day work shifts. The 1-match procedure can be designed to take into account secondary considerations, such as the allowable percentages of tripper shifts, fixed number of full-time shifts, and allowable overtime. If the solution of the 1-match procedure is not acceptable to the planner, then the half-day schedules or tripper-shift schedules can be changed to reflect the planner's complaints and the 1-match solution used. We are currently carrying out a project to design and implement the above procedure.

ADDITIONAL COMMENTS

We have presented a variety of manual and computer-based procedures for estimating, bounding, or determining exact vehicle and crew requirements for transit systems. We have also attempted to illustrate the strengths and weaknesses of these procedures, their assumptions, and their possible utility. A problem that planners encounter in analyzing their problems is (a) they expect too much from some models or (b) they overbuild and complicate their model in attempting to get their desired results. In the first case, their results are superficial and incomplete; in the second case, their results are extremely costly to derive. We have demonstrated that, if the goals of the planner are modest, simple procedures will derive the required answers. However, if a more detailed result is needed, a much more complex and costly model needs to be constructed. Therefore, the planner has to decide whether the more detailed result is needed and whether he or she is willing to pay the price (in terms of data collection, computer programming, and computer time) to find the results. Finally, we hope that the results in this paper demonstrate that estimates (and not necessarily bounds) of the solution are of use in a planning environment.

ACKNOWLEDGMENT

Research into the results presented in this paper was sponsored by an UMTA contract with the University of Maryland.

REFERENCES

1. L. Bodin and D. Rosenfield. Estimation of the Operating Cost of Mass Transit Systems. College of Urban and Policy Sciences, State Univ. of New York, Stony Brook, Rept. WAHCUPS-UMTA-1-76, Sept. 1976.
2. UTPS Reference Manual. Urban Mass Transportation Administration, 1975.
3. L. Bodin and R. Dial. Determining Crew and Vehicle Requirements for Mass Transit Systems. College of Business and Management, Univ. of Maryland, College Park, 1979.
4. Vehicle Scheduling and Driver Run Cutting (RUCUS) Package User Documentation. Mitre Corporation, McLean, VA, MTR-6678, May 1974.
5. MTA Run Cutting Requirements Analysis. De Leuw, Cather, and Company, Washington, DC, RUCUS Technical Rept., No. 1, Nov. 1976.
6. L. R. Ford, Jr., and D. R. Fulkerson. Flows in Networks. Princeton University Press, Princeton, NJ, 1962.
7. N. Christofides. Graph Theory: An Algorithmic Approach. Academic Press, New York, 1975.

Publication of this paper sponsored by Committee on Bus Transit Systems.