TRB Paper Manuscript #09-0587

Passenger Environment Survey:
Representing the Customer Perspective in Quality Control

*Alex Lu, Steven Aievoli, Joseph Ackroyd, Chrissie Carlin, and Alla Reddy**
* Corresponding author

Alla Reddy
Senior Director, System Data & Research (SDR)
Operations Planning
New York City Transit
2 Broadway, A.17.92
New York, NY 10004
Tel: (646) 252-5662
Email: *Alla.Reddy@nyct.com*

Word Count: 250 (Abstract) + 5,745 (Text) + 6 × 250 (Figures) = 7,495 Words

ABSTRACT

MTA New York City Transit's Passenger Environment Survey (PES) uses a quantitative and scientific approach to measure the perception of NYCT's 7.3 million daily riders. With 6,388 subway cars, 4,576 buses, and 468 stations in a 321 square miles service area and population of over eight million, quality assurance is truly a colossal undertaking. PES takes a passenger-centric approach by measuring indicators from the customer's perspective.

Since its inception in 1983, PES has evolved to include 68 indicators in four distinct categories measured in four passenger environments: Subway Car, Stations, Bus, and Express Bus. The consistent and well-defined PES standards are clearly understood by operations personnel. Approximately 25 dedicated surveyors, who do not report to operations management, produce semi-annual reports with statistical precision exceeding 95% ± 5%. The data is subject to validation and rigorous quality control by trained statistical analysts.

Central to the PES program is NYCT's genuine willingness to understand the customers' experience, and to represent customers in quality assessment. Internally, PES functions as a performance audit. Monthly reports bring operations management attention to observed deficiencies. Results are considered in promotion and merit decision-making processes. Externally, PES serves as a dispassionate and analytical measure of passenger experience. Reported semi-annually, the scorecard demonstrates long-term progress in continuous improvement. NYCT regularly receives requests nationally and internationally from agencies wishing to model quality programs after PES. This approach demonstrates dedication of NYCT staff in maintaining a friendly and comfortable system, in good repair, of which every New Yorker can be proud.

INTRODUCTION

MTA New York City Transit (NYCT) operates the third largest subway system in the world (by annual ridership), carrying about 5.0 million riders on an average weekday. The subway system extends 835 track miles through four boroughs, covering a service area of 321 square miles and serving a population of 8.3 million people 24-hours, seven days a week. The subway is equipped with 6,388 electric passenger cars stored in 13 yards and two heavy maintenance facilities, travelling a combined total of over 354 million miles a year on 23 routes and three permanent shuttles. The system operates over 468 stations with a total of 5,105 stairways, 906 platforms, 163 elevators, and 174 escalators. On the bus side, the 207 local and 36 express routes serve 12,500 stops and provide almost 55,000 weekday scheduled trips, carrying 2.4 million weekday riders throughout the 1,852 mile route network. The 1.87 billion passenger miles consumed by New Yorkers each year in 926 million discrete trips requires a fleet of 4,576 buses maintained in 19 depots by nearly 17,000 dedicated NYCT Department of Buses (DOB) employees.

The vast labyrinth of NYCT's large and sprawling network makes for quite a daunting quality control task. How does NYCT know how well it's doing with respect to cleanliness and appearance, customer information, state of maintenance, and proper operation of equipment from a passenger perspective? How is a consistent standard applied across the many operating divisions, car barns, and bus depots to monitor crew performance? The Passenger Environment Survey (PES) was first designed in 1983 to address some of these concerns. It has evolved over time to include 68 indicators in four distinct categories measured in four different passenger environments. Nine survey forms and approximately 25 surveyors managed by two supervisors collect more than 2,000 survey assignments per quarter. Three analysts produce semi-annual reports with statistical precision exceeding 95% ± 5% in each indicator. NYCT has such confidence in these measures that the results are considered in the promotion and merit decision-making process. This paper will serve as a primer on the state-of-practice of NYCT's quality processes and discuss potential future improvements and development in PES.

BRIEF HISTORY

PES was introduced in the First Quarter of 1983 and consisted of a handful of very simple indicators designed to quantify how amenable and comfortable the transit environment was to our customers. Although it is difficult to objectively assess the many conditions that relate to passengers' feelings and perceptions, the early indicators provided a systematic framework to assess performance in these areas independently of anecdotal customer comments. In that sense the early PES was a big step forward and an important recognition that NYCT cares deeply about the environment that is presented everyday to our 7.3 million customers (*6*). The early PES included 11 subway indicators, and 21 bus indicators.

Several indicators were added in 1995, expanding the scope of PES. Indicators relating to the station environment such as functioning turnstiles and escalator/elevators were added. A major revision occurred in 2000, when indicators specifically designed to measure the performance of cleaning crews were added. Quarterly reports were scaled back to semi-annual beginning in 2004. NYCT began monitoring the passenger environment aboard express buses in 2007. Additionally, the PES survey methodologies have been used on a variety of "pilot" policy initiatives designed to measure the impact of allocating additional production resources to

cleaning functions. The PES approach has been such a success that NYCT regularly receives requests for details, nationally and internationally, from other transit properties wishing to model quality control programs after NYCT's PES.

Initially, PES data was tabulated manually, and reports were issued quarterly with substantial time lag. Although this enabled managers to act after the fact to improve performance, many months may pass before substantial performance improvement is achieved. With the advent of personal computers, NYCT began keeping PES data and reporting results using spreadsheet applications and databases. Internal reports are now issued much more frequently, allowing management to react to mediocre scores mid-quarter. Data mining capabilities allow special requests from top management to be handled quantitatively.

Today, NYCT routinely produces internal PES monthly reports as a snapshot to assist operating personnel in taking corrective actions. The PES database is being migrated to Transit's enterprise server. The data collection effort is being transitioned from pen and paper to handheld computers. The PES staff provides regular 'Roadshows' in field locations to improve operations managers' understanding of PES ratings and how they might improve PES scores, thereby improving the customer experience. These PES Roadshows have become an important mechanism of information exchange, improving relationships between PES and operating staff, and allows PES staff to gain valuable understanding of operations processes.

MEASUREMENT INDICATORS (WHAT TO MEASURE?)

PES is designed to measure passenger experience quantitatively in four major passenger environments: Subway Car, Subway Station, Local Bus, and Express Bus. In each environment, four aspects of passenger experience are measured: Cleanliness & Appearance, Functioning Equipment, Customer Information, and Operations. These categories were partly devised to align performance indicators with responsible operating departments, to promote a culture of accountability within NYCT.

Simply measuring the passenger experience is not sufficient for certain indicators. Indicators measured during service cannot assess cleaning crew performance. In a city of 8.0 million people a clean and spotless subway car does not remain that way for very long. During the 2000 revision of PES, special surveys of key indicators were added before service specifically to monitor cleaning crew performance. Thus, PES also serves as an internal audit tool. While cleaning supervisors and managers are responsible for day-to-day quality assurance, PES provides random sample checks prior to entering service.

In total, PES measures 68 discrete indicators: 15 for Subway Cars, 17 for Subway Stations, 24 for Local Bus, and 12 for Express Bus. Table 1 summarizes all PES indicators currently measured.

Table 1
Passenger Environment Survey (PES) Indicators Currently Measured by NYCT

	Environment			
Category	**Subway Station**	**Subway Car**	**Local Bus**	**Express Bus**
Cleanliness and Appearance	• **Pre-AM Peak:** Cleanliness Litter Graffiti • **Post AM-Peak:** Cleanliness Litter	• **At Terminal:** Cleanliness Litter • **In Service:** Cleanliness Litter • **Graffiti:** Interior Exterior Window • Cracked Windows	• **Before Entering Service:** Cleanliness Litter Exterior Dirt • **While in Service:** Cleanliness Litter Exterior Dirt • Panels & Bumpers • Cracked Windows • **Graffiti:** Interior Exterior	
Customer Information	• Delay Announcements • Correct Maps • Passenger Information Center (PIC) • System Map Availability	• Public Address Announcements • System Maps • Signage Correct	• Bus Announcements • **Route Signs:** Front Side Rear • Bus Map • Priority Seating Stickers	(N/A)
Functioning Equipment	• Annunciator • Escalators/Elevators • Telephones • Booth Microphone • Trash Receptacles • Turnstiles	• Door Panels • Lighting • Climate Control	• Climate Control • Operative Rear Door • Kneeling Feature • Wheelchair Lift • Operative Windows	• Reclining Seats • Reading Lights
Operations	• Proper Uniform • Displaying Badges	• Proper Uniform	• Proper Uniform • Displaying Badges • Board / Discharge Passengers	(N/A)

Passenger Environment Survey (PES) Standards

Quantitative measurements require clear and well-defined standards. The majority of PES standards were developed from the same quality control criteria used by operating departments to manage their employees. Quality control criteria are found in various agency documents. Some are codified in rules and regulations, union contracts, standard operating procedures, and other official documents; others appear in training materials, field manuals, and management procedures distributed to field staff. For example, New York City Transit's Rule Book (*1*) Rule 10(f) requires:

> "Employees required to wear uniforms must at all times when on duty wear the prescribed uniform and badge. The uniform must be kept neat and in good repair…"

The corresponding PES standard explicitly references the *Uniform Standard* issued by the Division of Rapid Transit Operations (RTO), which contains more detailed description of what constitutes proper uniform. Train announcements, while covered in the Rule Book under Rule 9.01(z) Parts 1 & 2, are further clarified by a field manual known as the Service Delivery Blue Book (*2*):

> "Here is an example of how the routine cycle of announcements works: This is A Street. Transfer is available to the X Train. This is a Bronx-bound Y Train. The next stop is B Street. Stand clear of closing doors, please."

The corresponding PES standard states:

> **Subway Cars Public Address Announcements:** This indicator measures the % of correct announcements heard out of the total number of potential announcements expected. All announcement types are assessed per station stop: Next Station (while enroute or arriving at a station); Transfer Options (if applicable); Route Designation (Letter/Number); Route Destination (Borough/Terminal); Next Station (while standing in or when leaving a station); "Stand Clear of Closing Doors".

Use of actual operating standards is important: it lends PES a degree of official credibility, and improves acceptance by management and the public alike (*5*). Use of a different criteria to monitor performance would be unfair to both employees and management. Additionally, use of operating standards ensures by default all field employees would be familiar with PES indicators, and thus clearly understand the yardstick by which their performance is measured.

SURVEY METHODOLOGY (HOW TO MEASURE IT?)

Because of PES's audit and merit assessment functions, NYCT must ensure data collection forces have no vested interest in survey results. Biases associated with using operating staff for data collection have been observed in many organizations. Difference between traffic checker ridership counts, conductor reported passenger counts, and ridership levels implied by fare receipts often differ in commuter railroads and other transit organizations. To provide a neutral data collection workforce, a cadre (group) of specially trained surveyors is used. Surveyors report directly to a performance measurement unit in central administration, minimizing the influence of operating departments.

The surveyors undergo four weeks of regular training upon induction to Transit. PES offers additionally a rigorous two-week training program, including classroom work and field demonstrations covering all aspects of PES standards and survey procedure. Once initiated into the PES program, surveyors are not permitted to switch work assignments until the end of the 'pick', a four-times-a-year event when surveyors may choose their work assignments. Typical promotional paths for surveyors include Station Cleaner and other operating titles, but surveyors are hired directly from outside through the usual HR process. Therefore, they start at Transit with minimal pre-conceptions about operations. These procedures together ensure that PES truly represents the perspective of customers – seeing the system as they see it.

The PES cadre has position for 25 part-time (5~6 hours per tour of duty) surveyors, plus two surveyor supervisors. This represents a substantial resource commitment from NYCT solely to monitor performance. To offset these costs, Transit also utilizes the PES cadre for collecting data relating to National Transit Database (NTD) FTA Section 15 and ADA Federal compliance surveys. These surveys require a strictly random sample that results in scattered observations in both time and space, thus a large surveyor cadre must be 'on-hand' to cover all possible combinations of times and locations.

Grouping Indicators into Surveys

It is not practical to measure all 68 discrete indicators at the same time. Conversely, having 68 different survey forms is impractical. Indicators are grouped into surveys based on location and time requirements for measurements. For instance, the In-Service Subway survey includes measurement of on-board temperature, and examination of subway car for cleanliness, litter, graffiti, functioning doors, and map conditions. These are generally independent variables; surveying them together does not introduce undesirable sample bias. Table 2 shows all PES surveys currently measured by NYCT, assigned time periods, and the observation quota required to reach design levels of statistical significance.

Table 2
Current PES Surveys with Sample Quotas

No.	**Survey Name**	Survey Code	Unit (Strata)	Quarterly Quota	Approx. Equiv. Clusters	# of Units	Unit Level of Significance	Effective Time Period	Duration (Hours)
1	In-Service Subway (Roving)	CR	Route	120	5	23	90% ± 6%	0600-2200 Weekdays	2
2	At-Terminal Subway	NR	Route	200	3	23	Spotcheck	0600-2200 Weekdays	2
3	Subway Announcement (Roving)	RA	Route	120	N/A	23	90% ± 6%	0600-2200 Weekdays	1
4	Early AM Bus	ND	Depot	120	2	19	Spotcheck	0400-0800 Weekdays	3~4
5	PM Bus	CD	Depot	120	4	19	Spotcheck	1700-2200 M, T, W	3~4
6	In-Service Bus	ST	Depot	120	9	19	90% ± 6%	0700-1900 Weekdays	3
7	Bus Announcement (Roving)	BA	Depot	N/A	7	19	Variable	0700-1900 Weekdays	2
8	Express Depot	ED	Depot	120	3	2	90% ± 6%	0500-0800 Weekdays	3
9	Express Bus Terminal	EXST	Depot	120	9	2	90% ± 6%	1500-1800 Weekdays	3
10	Station Announcement	SA	Program	343	N/A	1	Variable	0600-2200 Weekdays	1
11	Daytime Station	CS	Station District	85~108	16	5	90% ± 8%	0800-2200 Weekdays	0.5
12	Early AM Station	NS	Station District	85~108	16	5	90% ± 8%	0200-0800 Weekdays	0.5

Note: Spotchecks are not representative at the quarterly level. Spotchecks aim to achieve a small number of checks per quarter, but collect extensive data during those checks. Quarterly performance results represent a composite of shifts on duty when checks were carried out.

Surveys that have an active operating component (e.g. In-Service Subway) are generally restricted to weekdays during daytime or evening. Those focusing on long-term state of maintenance (e.g. System Maps) are conducted at any time. Surveys designed to monitor performance of one specific maintenance crew must be conducted near the end of that crew's shift (e.g. Early AM Stations). Some surveys require cooperation from operating departments, thereby restricting timeslots to when personnel are available to assist the surveyor (e.g. PM Bus).

Designing Data Collection Forms

Graphic 1 shows a typical completed survey form. The In-Service Subway survey requires surveyor to board a train, recording actual boarding time and location (09:44, East Tremont Avenue). As train arrives, the surveyor scans it for exterior graffiti. While traveling between stations, surveyor notes the car number and surveys interior car environment. The form demonstrates important features in this type of quality control survey:

1. **Quantitative Measurements are Taken Whenever Possible:** Surveyors are required to count or measure quantities. Number of doors not operating is counted. Interior temperature is measured with a digital thermometer. Quantitative data enables analysis to be conducted after the fact, and reduce the risk of surveyor errors in judgment.

2. **Auditing Information is Gathered Simultaneously:** By collecting information such as car numbers, boarding and alighting stations and times, an analyst can later establish whether the surveyor was actually present at the correct location. The fleet numbers can be cross-checked with block register information independently recorded by operations personnel in signal towers. This is particularly critical when invariably a bad score causes operating departments to challenge the validity of one surveyor's work. If discrepancies are found, results are discarded and surveyor asked to explain how mismatching car numbers appeared in their data. Additionally, spotchecks by surveyor supervision helps to maintain data quality.

3. **Failing Scores are Clearly Documented:** Whenever a failing score is recorded, the surveyor records the reason on the form. This demonstrates the surveyor understands PES standards and their correct application. The data is also used to analyze root causes of failing scores. The requirement to record specifics (e.g. color of observed graffiti) allows for post-facto auditing of surveyor activity. Analysts regularly review the forms to gauge surveyors' comprehension level and act proactively to prevent recurring errors, by reinstructing surveyors as necessary.

4. **Qualitative Measures Require Clearly Defined Rating Standards:** Occasionally, quantitative measure is difficult or impossible to collect. An ordinal 'rating' standard is then used with clear definitions. Examples are shown in Graphic 1. Bottles and cans, and food matters trigger an automatic failure rating.

Graphic 1
Typical Completed PES Survey Form, with Definitions

PASSENGER ENVIRONMENT SURVEY — IN SERVICE RAPID — AS OF 2/2007

JOB # 6 ROUTE 6 PAGE 1 OF 3

DATE January 17, 2008 DAY: THURSDAY MORNING

SURVEYOR WEATHER Cold

BOARDING TIME 0900 944 BOARDING STATION Grand Central - 42 St - East Transept

ALIGHTING TIME 1100 1022 ALIGHTING STATION Brooklyn BRIDGE

		1	2	3	4	5	6	7	8	9	10	11
CAR #		7226	7227	7228	7229	7230	7555	7554	7553	7552	7551	
EXTERIOR GRAFFITI	1,2	1	1	1	1	1	1	1	1	1	1	
# DOOR PANELS OUT	0-8	0	0	0	1	0	0	0	0	0	0	
LIGHTS	0,1,2	1	1	1	1	1	1	0	1	1	1	
# LIGHT BULBS OUT		0	0	0	0	0	2	0	0	2	3	
TEMPERATURE #1		66.7	66.3	65.3	67.0	68.4	67.6	67.5	66.3	65.7	67.6	
SYSTEM MAP #1		1196	1190	1190	1190	1190	N/MAP	1190	1190	1190	N/MAP	
SYSTEM MAP #2		1190	1190	1190	1190	1190	1190	1190	1190	N/MAP	1196	
ROLLSIGN	1,2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	
INTERIOR GRAFFITI	1,2	1	1	1	1	1	1	1	1	1	1	
TEMPERATURE #2		65.1	66.3	65.5	66.3	65.0	66.6	66.5	66.1	67.7	67.5	
LITTER	1,2,3,4	2	1	1	1	1	1	1	1	1	1	
GRAFFITIED WINDOW	1,2	1	1	1	1	1	1	1	1	1	1	
SCRATCHED WINDOW	1,2	1	2	2	2	2	2	2	2	2	2	
CRACKED WINDOW	1,2	1	1	1	1	1	1	1	1	1	1	
DIRTY FLOOR/SEATS	1,2,3,4	1	1	1	1	1	1	1	1	1	1	
UNIFORM	0,1,2	0	0	0	0	0	1	0	0	0	0	

Car #1 Tissue, ~~Snack~~ Paper Bag, Empty Coffee cup

Car #2

Car #3

Car #4 R2 Door Panel

Car #5

Car #6

Car #7

Car #8

Car #9

Car #10

Car #11

NOTE: BOTH "**LITTER**" AND "**DIRTY FLOOR / SEATS**" RECEIVE A RATING OF "**4**" FOR BOTTLE(S), CAN(S), LARGE QUANTITIES OF FOOD PRODUCTS OR BIOLOGICALS.

- **Exterior Graffiti:** Presence of graffiti on the carbody (excluding windows) results in a failing score of '2', otherwise a passing score of '1'.
- **Door Panels Out:** Modern cars are equipped with eight doors (in four doorways) per side. The surveyor counts the number of doors failing to open at station stops.
- **Lights:** While underground, lights correctly lit are rated '1' (Pass); lights incorrectly powered off are rated '2' (Fail). While above ground, the lights are always rated '0' (Not Applicable) regardless of its condition.
- **Light Bulbs Out:** Count of fluorescent tubes not operating correctly.
- **Temperature #1/#2:** The interior ambient temperature is measured with a digital thermometer (temperature probe). #1 is measured at one end of the car, and #2 at the opposite end. NYCT is currently in the process of testing infra-red surface temperature guns as an alternative. Subway cars have ceiling sticker targets indicating where temperature guns should be aimed, to ensure uniformity in measurements. To achieve a passing score, cars must have an average temperature between 58˚F and 78˚F.
- **System Map #1/#2:** Last four digits of the posted System Map's commodity number is recorded. Analytical staff uses it to determine if the map is outdated. To pass, all old maps must be replaced within six months of the date of issuance of an updated map (with a new commodity number). Failure if map is torn or defaced.
- **Rollsign:** Older cars are equipped with curtain signs that show the train's destination, which sometimes rip or are rolled incorrectly. As these have been upgraded to electronic message signs, NYCT will be developing a standard for monitoring the electronic signs.
- **Interior Graffiti:** Presence of pigmented graffiti (excluding windows) results in a failing score.
- **Litter:** Surveyor uses PES standard to determine rating:

Rating	Result	Definition	Examples
1	Pass	Basically litter free	No litter.
2	Pass	Small amounts of scattered dry litter	Metrocards, tissue, cigarette butts, food wrappers (no food), lottery tickets, empty cup, plastic bags, newspaper.
3	Fail	Noticeable assortment of dry litter	Larger quantities of the above, and batteries.
4	Fail	Heavy litter Opened or spilled food Malodorous conditions Hazardous conditions	Extreme amount of assorted trash, bottles, cans, chicken bones, half eaten burger, banana skin, broken glass, biological waste, vermin.

- **Graffitied Window:** Presence of graffiti (excluding scratchitti or acid-etching) results in a failing score.
- **Scratched Window:** Presence of scratchitti results in a failing score. Not currently reported.
- **Cracked Window:** Presence of cracks in any window pane results in a failing score.
- **Dirty Floor/Seats:** Surveyor rates cleanliness condition on a scale of '1' through '4' using a defined standard (similar to litter).
- **Uniform:** Surveyor rates conductor's uniform based on operating standards.

SAMPLING ISSUES

Statistically valid measurements require even and representative samples. Biases due to sampling must be minimized, to ensure validity in PES's audit and merit assessment functions. At the same time, sample size must be minimized for cost reasons. Striking a balance between a statistically valid sample and minimum surveyor cost is not simple. Historical data analyses suggest PES scores are correlated with time-of-day, location, and service route, even though indicators themselves are independent. PES scores are also subject to clustering effects: data collected sequentially on the same day may be adversely affected by common factors such as a school condition, weather conditions, or special events. A pure random sample is an inefficient method of collecting such data, but observations made sequentially cannot be treated as independent observations.

Generally, a good statistical sample should fulfill three criteria:

1. Sample error margins should be appropriate for the intended use of the statistic.
2. Sample should be representative of the underlying population.
3. Data collection method should not violate assumptions (e.g. assumption of independent observation) used when designing the sample.

Acceptable Error Margins

To determine sample size required for an acceptable error margin, the binomial sampling formula is typically used:

$$\text{Sample Size } (n) \quad = \quad Z^2 \times \frac{p \times (1-p)}{e^2}$$

where

p = Assumed Rate of Detection (i.e. receiving a failing score)
e = Acceptable Standard Error

PES indicators typically have an average passing rate of 80%. To collect statistics at the route level having ± 6% error at 90% confidence level, 120 observations are required. To report data quarterly, 120 observations per route are required every quarter.

The error margins are important. Questions often arise particularly when passing rates fall. Adverse changes in PES statistics may be due to random fluctuation resulting from sampling, or true performance degradation. When actual performance changes result in a lower sample mean score, it is termed a 'statistically significant change'.

Technically, to rule out score reduction due to sampling error, a *t*-statistic (comparing two sample means) should be calculated using raw data (*3*). However, analysis has shown a high degree of correlation between significant PES score changes (defined by the *t*-test) and score changes exceeding sample error margin. Rigorous significance testing is not used in day-to-day management reporting, due to the difficulty of explaining the mathematics to the public and to field personnel alike. Nonetheless, an active effort is made to educate the data consumer that a score change within the error margin does *not* reflect a significant change in performance.

However, a score that is trending such that the cumulative changes summed over several reporting periods exceed the error margin generally *is* a significant change. For all practical purposes at NYCT, when score improvements above the error margin are achieved, it is considered a cause for celebration and commendation.

Ensuring a Representative Sample

A representative sample weighs each sub-class of the population equally. Applying that principle to PES requires analysts to determine *a priori* what factors might affect PES scores, and stratify the sample based on buckets of these variables. However, PES scores are affected by many factors. With only 120 observations per reporting period and many scheduling constraints, a true stratified sample is not possible. NYCT uses a heuristic (rule-of-thumb) to minimize the impacts of skews introduced by uneven sampling:

- Surveys are evenly distributed throughout the whole reporting period, e.g. nine required Bus Terminal surveys are equally distributed throughout a quarter, with three surveys assigned each month.
- Where surveys are conducted at multiple locations, or roving, assigned locations and directions are evenly distributed.
- Surveyors are scheduled such that they will not repeat any depot, terminal, or route within a quarter.
- For surveys with a long allowable timespan (e.g. 0800-2200), samples are evenly distributed between morning, afternoon, and evening.

Minimizing Clustering Effects

A representative sample should not rely on many observations gathered sequentially, in close temporal or spatial proximity. Without knowing much about correlation between sequential observations and factors driving them, it can be difficult to know to what extent clustering is a problem. Past experience with PES suggests that approximately 10 clusters per unit per reporting period generates results that would not be excessively skewed by a ‘bad’ cluster.

However, data collection is subject to practical constraints. Gathering many observations in one cluster is efficient because travel time per observation is minimized. Practically speaking, PES aims for at least 5 clusters per reporting unit, except for depot-based surveys where a bare minimum of 2-3 clusters are tolerated. Indeed, PES historical data sometimes show large score swings in depot-based surveys, most likely the result of a single bad cluster that may account for up to one-third of overall score. Error margins are not reported for depot-based surveys; they are thought of as quality assurance ‘spotchecks’ where the surveyor arrives without prior notification and expects to always find very good results.

SCHEDULING ISSUES

Each PES survey has a unique set of requirements and constraints, and no survey occupies an entire shift. Even though constraints are well defined, flexibility in scheduling is substantial. Adding to complexity is the different surveys to be covered each day, depending on previously missed or incorrectly completed assignments, and surveys remaining to be assigned. The resulting work-piece matching problem is fairly unique, unlike typical run-cutting and scheduling:

1. Fixed-shift surveyors must start at a pre-defined time, but variable-shift surveyors may start at different times.
2. Each part-time surveyor receives 5~6 hours' worth of work. Each surveyor must have tasks assigned during the entire shift, including travel.
3. Surveyors may be assigned overtime if work and travel exceeds their shift, but their permission must be sought.
4. Observation count and sampling evenness constraints must be met for all surveys.

Piece-matching problems can generally be solved in real-time using a linear program, but the objective function is not clear. Minimizing surveyor-hours is desirable, but the minimum surveyor-hour solution may compromise sample evenness, or threaten sample completion within the reporting period. Safeguarding sample evenness and utilizing available surveyor hours to 'work ahead' both involve human judgments, therefore PES scheduling is currently a manual process.

Recurring surveys by subway line and bus depot turns out to be a substantial scheduling exercise. In a typical quarter, PES assigns 1,300 PES surveys, 180 Federally-mandated Section 15 surveys, hundreds of new Staten Island surveys, and an additional 260 pilot surveys for NYCT Policy Initiatives. More than 2,000 surveys per quarter provide work for about 25 surveyors and require the back-end support of three to four analysts.

The Scheduling Process

Surveyor schedules are issued weekly, about one week ahead. Vacation and overtime requests, submitted two weeks in advance, are inputs to scheduling process. The analyst begins by scheduling Federally-mandated Section 15 surveys. Section 15 surveys occur at random locations and times. PES schedulers must ensure surveyors are available to cover this work – if necessary by changing variable-shift start-times to provide sufficient surveyors at times requested.

Next, the scheduler checks an inventory of returned surveys. Sample quota must be met in each quarter. Generally, the analyst selects surveys that show a lower percent-complete. By always choosing to schedule these surveys, the sample should be naturally distributed evenly.

The surveys (1~3 hours in duration) are linked together to form a shift. This process is repeated until all available surveyors have work for their tour of duty. Work schedules are entered into a scheduling spreadsheet. A program generates requisite data collection forms, collated in order, for all surveyors daily.

PES surveys are generally linked with a Federal compliance survey or another PES survey. The linking process assigns surveys in geographic and temporal proximity sequentially to the same surveyor. Therefore, surveys near random Section 15 locations have greater likelihood of being picked. Later in the reporting period, as more surveys are complete, the bucket of work remaining is reduced, resulting in somewhat higher travel times and less efficient surveyor assignments.

Pre-Blocking Approach

The scheduling process consumes substantial analyst time, and efforts were made to simplify it. The approach was to permanently link surveys into blocks representing exactly one shift, to improve efficiency. Linking is done *a priori* for the entire reporting period, instead of week-by-week. Permanent pre-blocking has several advantages:

1. Permanently assigned work pieces could be distributed more efficiently than ad-hoc assignments linked to the Section 15 sample. For instance, "In Service Subway" surveys are conducted on subway lines while travelling. These are permanently coupled with "Daytime Station" surveys, which invariably require substantial subway travel to reach assigned stations.
2. Blocks are re-cycled each quarter. The weekly scheduling chore reduces to assigning surveyors to blocks, rather than assigning custom survey sets worked out each week.
3. Surveys are more easily spread out evenly. Clusters are pre-assigned at roughly even intervals and are geographically evenly distributed.

The pre-blocking approach does not account for dropped partial blocks due to surveyor error. Partial blocks still require manual linking to other dropped blocks and Section 15 assignments, which are random and non-repeating. Nonetheless, this approach resulted in reduced scheduling workload.

ANALYZING PES DATA

Data analysis is a very important part of any data collection exercise. Analysis draws conclusions and answers initial questions posed by the survey designers. Incorrect data analysis can lead to erroneous conclusions just as poorly designed surveys or badly collected data can.

Basic premise for the PES survey design has each observation of each indicator result in a 'pass' or 'fail' score, except where quantitative data is collected. Data collection forms reflect this philosophy, where a value of either '1' (Pass) or '2' (Fail) is required for each indicator. This is a good approach for two reasons: (1) simplified instruction of surveyors on making correct and valid observations; (2) simplified analysis methodology, where the reported passing rate is basically passing observations divided by total observations.

Where quantitative data is collected, or where criteria for making a pass/fail judgment are too complex for reliable instruction to surveyors, they are asked to record raw measured values. NYCT surveyor supervisors often use the mantra, 'write what you see', when giving instructions.

Data Entry and Validation Processes

Surveys returned from the field are entered into database for analysis. The database is currently housed in Microsoft Access 2002, with main data tables having planned migration to NYCT's Oracle Database Enterprise Server. Data access shall be via ODBC link over Ethernet local area network. As data is entered, it is subject to validation rules implemented by either the analyst or the database itself:

1. When severe oversampling has occurred in an assignment, every other observation is entered. If 40 observations are expected, and surveyor returns 76 observations, only 38 observations (every other, not the first half) are entered. This prevents particularly prolific surveyors from overwhelming the sample quota.
2. Database checks for depot-bus route consistency. The validation prevents misallocation of buses to depots from affecting depot-based scores.
3. Duplicated bus or subway car numbers surveyed on the same day are rejected. Bus depots are large; some surveyors survey hundreds of buses in one survey. The validation prevents surveyor error in repeatedly checking the same bus from affecting scores. Occasionally, duplication occurs on rail lines with short car cycles. The latter of duplicate observations is discarded.
4. Analyst checks if ratings are substantiated by surveyor comments as each record is entered. Occasionally, surveyors do not apply PES standards correctly (e.g. a bottle marked as '3' rather than required '4'.) Obvious errors are corrected where possible.

When all data is entered for the quarter, first the sample quota fulfillment is checked. Next, database queries are used to check for unusual patterns of activity. If one surveyor's scores show significantly different trends compared to others, surveyor supervision is informed and field efficiency test is conducted. Reinstruction on PES standards is given if errors are detected and data is discarded.

Monthly Diagnostic Report to Operating Divisions

PES surveys have a duality of functions: first and most importantly, as an internal and external 'scorecard' for field managers' performance. It provides an independent audit and a measure of merit and effectiveness. Secondly, it is a data-mining tool to assist field managers in finding 'problem areas' within their district of responsibility. At NYCT, district operations managers are responsible for as many as 5 depots and 1,000 buses, or up to 5 lines, 150 rush-hour train-sets, and more than 1,000 daily train starts. While responsibility for service delivery rests squarely upon operations managers, PES results serve as early indicators of developing 'hot spots' due to difficulties in certain areas or aspects of service delivery. The most frequently requested item at PES Roadshows is a monthly performance barometer that gives operations management a snapshot of how well the team has been doing.

Table 4

SUBWAY - Monthly Report

% None & Light

Cleanliness Conditions - by Line

Measured "At Terminal" with cleaners present

Results thru May 31, 2008

	Quarterly Results													Monthly Results		
Year	2005				2006				2007				2008	2nd Qtr. results thru…		
Quarter	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	April	May	
1	77%	97%	99%	97%	99%	98%	99%	100%	100%	100%	100%	100%	99%	95%	97%	
2	98%	100%	81%	83%	96%	97%	100%	100%	92%	96%	100%	99%	100%	100%	99%	
3	88%	99%	94%	100%	97%	92%	100%	81%	100%	92%	96%	99%	97%	84%	92%	
7	96%	76%	71%	94%	95%	94%	89%	96%	98%	96%	100%	97%	98%	96%	98%	
District 1	**91%**	**94%**	**89%**	**95%**	**97%**	**96%**	**97%**	**93%**	**98%**	**96%**	**99%**	**99%**	**99%**	**94%**	**96%**	
4	99%	93%	92%	98%	98%	97%	95%	94%	100%	100%	94%	96%	100%	98%	99%	
5	100%	99%	98%	98%	86%	100%	97%	98%	99%	96%	98%	99%	100%	100%	100%	
6	97%	97%	97%	100%	97%	99%	99%	92%	100%	94%	99%	95%	97%	100%	100%	
District 2	**98%**	**96%**	**95%**	**99%**	**94%**	**99%**	**97%**	**95%**	**100%**	**97%**	**97%**	**96%**	**99%**	**99%**	**100%**	
B	97%	97%	97%	98%	100%	100%	100%	98%	100%	98%	100%	96%	96%	100%	99%	
D	98%	77%	88%	91%	97%	90%	89%	100%	95%	98%	96%	96%	88%	81%	91%	
N	93%	87%	100%	99%	94%	97%	100%	95%	93%	89%	97%	77%	95%	99%	99%	
Q	93%	98%	95%	87%	97%	97%	100%	100%	99%	89%	98%	99%	99%	100%	99%	
W	98%	98%	100%	100%	99%	92%	99%	99%	97%	98%	99%	99%	98%	98%	95%	
District 3	**96%**	**91%**	**95%**	**95%**	**98%**	**95%**	**98%**	**99%**	**97%**	**95%**	**98%**	**94%**	**96%**	**96%**	**97%**	
A	96%	94%	100%	100%	84%	93%	99%	98%	99%	100%	99%	100%	98%	99%	96%	
C	93%	94%	97%	97%	96%	94%	98%	97%	91%	96%	99%	70%	98%	95%	97%	
J Z	98%	92%	99%	99%	99%	93%	99%	91%	98%	100%	99%	96%	91%	51%	73%	
M	100%	99%	100%	99%	97%	97%	100%	95%	85%	95%	95%	95%	95%	95%	94%	
L	96%	98%	100%	89%	99%	99%	99%	100%	98%	98%	96%	98%	99%	99%	99%	
District 4	**96%**	**96%**	**99%**	**96%**	**94%**	**96%**	**99%**	**98%**	**93%**	**97%**	**97%**	**90%**	**96%**	**88%**	**92%**	
E	95%	97%	100%	91%	99%	92%	99%	95%	92%	91%	93%	88%	86%	88%	94%	
F	90%	82%	75%	84%	98%	84%	94%	87%	84%	89%	97%	84%	80%	100%	99%	
G	93%	96%	88%	100%	91%	95%	81%	88%	97%	74%	96%	94%	91%	96%	93%	
R	93%	96%	99%	99%	98%	98%	93%	95%	95%	91%	96%	89%	99%	100%	98%	
V	92%	99%	74%	99%	100%	83%	88%	81%	88%	84%	94%	89%	85%	96%	89%	
District 5	**92%**	**94%**	**85%**	**94%**	**97%**	**90%**	**91%**	**90%**	**91%**	**86%**	**95%**	**89%**	**88%**	**96%**	**94%**	
Systemwide	**95%**	**94%**	**93%**	**96%**	**96%**	**96%**	**96%**	**96%**	**96%**	**95%**	**97%**	**94%**	**96%**	**94%**	**95%**	

Drawing statistical conclusions based on a monthly sample (n=40) is practically impossible by line or depot. Monthly indicators are helpful on a district level (groups of between 3~5 lines or 4~6 depots). It gives early indication of areas where district managers can focus their trouble-shooting to maintain scores. Frequently, a low monthly score can be attributed to a specific problem at one location that is resolvable once management attention is directed to it. For all of these reasons, NYCT makes cumulative monthly averages for selected PES indicators available to field management monthly.

Early AM Depot or At Terminal Litter and Cleanliness scores are of particular interest to operations managers. While a clean subway car might not stay very clean in service for very long, no excuses can be made for buses or cars leaving depot or terminal (where cleaners are

situated) dirty. Table 4 shows the internal Subway PES Monthly Report distributed to line managers and district superintendents. Quarterly results for last three years are also shown, to give historical context.

NYCT does not conduct routine analysis to examine correlatory or causal factors for PES scores. Based on internal monthly reports, which are grouped by car class, car barn, and line, managerial accountability and any vehicle or location specific factors are apparent. The data would not support quarterly analysis at a more granular level. Analysis of long term trends, and correlation with input factors such as headcount and person-hours, are conducted as special projects upon senior management request.

Semi-Annual Report to the Public

The main function of PES is to provide a public benchmark for NYCT's performance. PES provides accountability for operations management and indicates areas requiring improvement. To this end, PES results are reported to NYCT Committee semi-annually as part of Committee Agenda. The NYCT Committee comprises of State, City, and Local political appointees serving as NYCT's Board of Directors.

Graphic 2 shows a sample page from an official semi-annual PES report (*4*). The report publishes all indicators at the systemwide level, with very high confidence levels (95% with error margins of ± 2% or less). For indicators of particular interest to the public (i.e. Station Litter and Cleanliness), results are published by district with somewhat larger error margins.

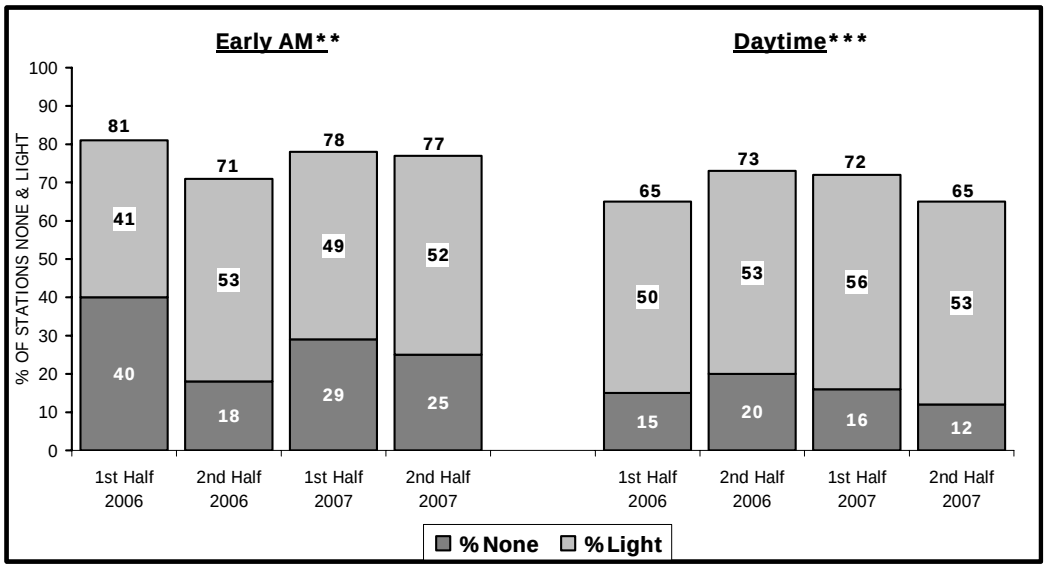
NYCT is particularly careful in its public reporting not to imply a degree of significance greater than the measurements can support. "Increases or decreases of less than 3% are statistically unchanged" is clearly noted on the page shown. Fluctuations of score within error margin may be due to sampling error or actual changes in performance. The PES sampling plan allows detection of performance changes with 95% confidence only when it exceeds 3%.

Definitions of PES indicators are clearly shown, to prevent misunderstandings or misinterpretations. In NYCT's public correspondence about PES and indicator results, the survey methodology is always clearly and concisely explained. The open survey standards prevent invalid comparisons with other independent surveys conducted by advocacy groups. Other surveys use entirely different methodologies to measure passenger experience and system performance.

Graphic 2

Passenger Environment Survey

Litter Conditions in Stations* (without Trackbed)



* Includes mezzanine, passageway, stairway and platform components only, not trackbed.
** Measured before heavy passenger utilization (pre-AM Peak).
*** Measured after heavy passenger utilization (post AM Peak).

Definition

Litter Conditions in Stations (Presence of Litter)

None- basically litter free;
Light- scattered dry litter;
Moderate- noticeable assortment of dry litter;
Heavy- heavy litter; any opened or spilled food, or hazardous conditions (bottles, cans).

2007 Annual Goals: (% none & light) Early AM: N/A Daytime: N/A

Semi-Annual Results

	Early AM				Daytime			
	None	Light	Mod.	Heavy	None	Light	Mod.	Heavy
2nd Half 2007	25%	52%	21%	2%	12%	53%	32%	3%
1st Half 2007	29%	49%	20%	2%	16%	56%	25%	3%
2nd Half 2006	18%	53%	25%	4%	20%	53%	24%	3%
1st Half 2006	40%	41%	16%	3%	15%	50%	31%	4%

Discussion of Results: an increase/decrease of less than 3% is statistically unchanged.

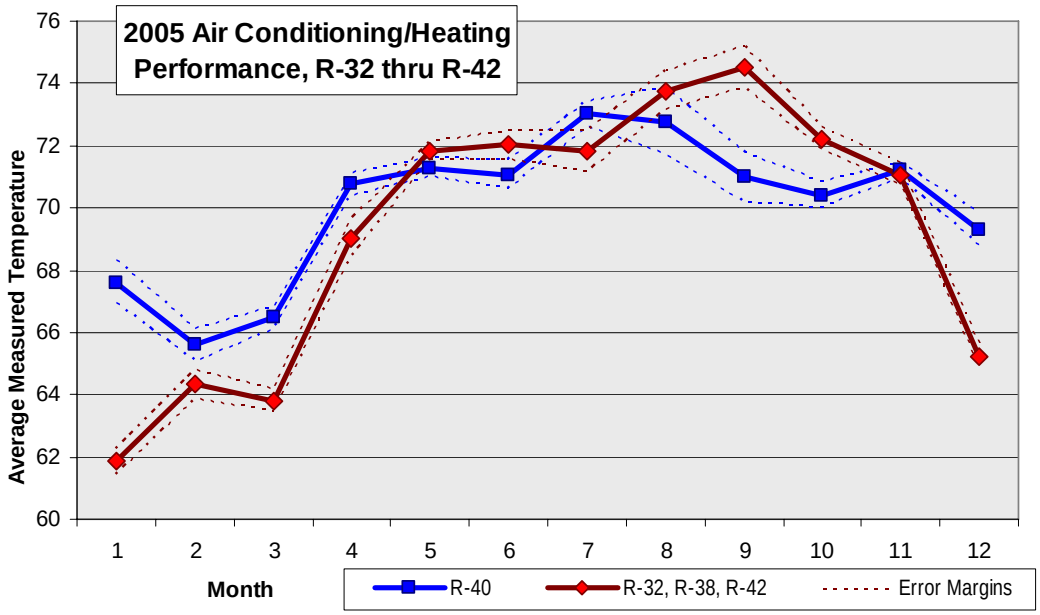
2nd Half 2007 vs. 2nd Half 2006: "Early AM" showed an increase of (+6%) while "Daytime" results decreased (-8%). As a result, a pilot program to address station cleanliness has been initiated.

Data Mining

The PES surveys result in a rich and varied dataset about NYCT's performance history. While potential combinations of analyses are endless, internal or external requests sometimes generate interesting questions easily answered using the PES dataset. Operations Planning analysts use standard database and spreadsheet applications to perform ad-hoc analysis and observe trends in the data as required.

As a hypothetical example, Graphic 3 shows analysis of in-cabin temperature (reflecting customer's perception of air-conditioning and heating performance) on NYCT's older generation B-Division cars. Compared with cars of approximately the same vintage, the Raymond Loewy styled R-40 'Slant' cars (soon to be retired) achieved a warmer average cabin temperature in the winter months of December through April, and cooler cabin temperatures in the late summer months of August through October.

Graphic 3



While heating and air-conditioning performance of specific subway car fleets is of limited interest, this analysis demonstrates the power of quantitative data collected primarily for a quality assurance and performance measurement program. When combined with appropriate research and analysis, the dataset can answer many questions and inform management decision-making quantitatively and scientifically. Graphic 3 is just one of many examples of possible ad-hoc analyses. The applications of PES data is only limited by the user's imagination and error margins inherent in any sample data.

FUTURE DEVELOPMENT

Future development plan for the PES quality control program currently lies in three major areas: conversion to paperless data collection, expansion to Staten Island Railway (SIR) and other MTA agencies, and growth in pilot monitoring programs as new initiatives designed to address customer experience deficiencies are added.

Paperless Data Collection

The most time-consuming aspect of PES data collection is manual data entry and validation processes needed once paper forms return from the field. Manual keying introduces undesirable lag-time from data collection to reporting, giving operations managers relatively little lead time to correct deficiencies identified by PES on an ongoing basis.

Since 2006, NYCT had been experimenting with paperless data collection strategies for PES using handheld computers known as Personal Digital Assistants (PDAs). Initial efforts to migrate data collection to PDAs were limited by software constraints. The 2006 state-of-art PDA software did not support the level of customization required for complex PES surveys.

Today, Transit was able to fit barebones PES data collection and validation capabilities into a small application on Compaq iPAQ PDAs running Oracle Lite with a Visual Studio Compact Framework front-end. Importantly, Oracle Lite supports direct synching with an Oracle Database Server, allowing direct data download from PDA to database automatically whenever PDAs are docked. Electronic data collection for Bus PES surveys are now in testing, and development of applications to support Subway and Stations surveys are ongoing.

PES Staten Island Program

NYCT was directed to expand PES to cover SIR. In consultation with SIR management, PES staff is developing performance standards, designing data collection forms and scheduling tools for monitoring customer experience on SIR.

PES Policy Initiative Monitoring Programs

PES survey methodologies and standards are well-established. Surveys provide quantitative and reproducible results. Because of its scientific and dispassionate nature, top management has tapped the PES infrastructure to monitor "pilot" policy initiatives designed to improve system performance by allocating additional production resources to certain maintenance functions. These independent monitoring programs are particularly important as they serve to inform governmental funding entities that additional production resources are being used to provide tangible and measurable benefit to the riding public.

Monitoring programs take an intensive approach. A special random sample is generated to monitor a specific line or defined geographic area. The same fidelity as typically achieved over one quarter is achieved bi-weekly. The data is collected for both 'before' and 'after' conditions, with results reported at 95% ± 5% level every two weeks. With this heroic data collection effort, results trending up clearly and unambiguously demonstrates that addition of resources improved passenger experience significantly. Bi-weekly reporting cycle allows management decisions to

be made on an ongoing basis; problematic areas receive almost immediate feedback on performance in the policy initiative.

NYCT is currently running three concurrent initiatives: Station Cleanliness Program on the Broadway, Lexington, Flushing and Canarsie Lines; Subway Car Cleanliness Program on the Flushing and Canarsie Lines; and Pilot In-Service Subway Survey, to measure combined impacts of initiatives on passenger perception. As additional resources become available, policy initiatives could be added to address weakest performing areas as identified by the PES.

CONCLUDING OBSERVATIONS

New York City Transit's Passenger Environment Survey (PES) is a quantitative and scientific approach to measuring passenger perception of NYCT's services. The PES standards are consistent, well-defined, and clearly understood by operations personnel. The indicators are measured from a passenger perspective by a group of dedicated and impartial surveyors, and subject to rigorous data quality control by trained statistical analysts. Internally, PES functions as a performance audit for field operations management. Externally, PES serves as a dispassionate and analytical 'scorecard' measure of passenger experience.

First designed in 1983, PES has evolved over time to include 68 indicators in four distinct categories measured in four different passenger environments. Nine survey forms and 25 surveyors produce semi-annual reports with statistical precision exceeding 95% ± 5%. NYCT has such confidence in these measures that PES results are considered in the promotion and merit decision-making process. The rich and extensive dataset generated by these periodic sample measurements are used throughout NYCT to support quantitative management decision making.

Central to the whole PES program is a genuine willingness on Transit's part to understand the environment that over seven million daily customers experience. NYCT is approaching the challenge of quality control from the customer's perspective. The NYCT quality processes are driven by what the customers see. We reaffirm that commitment to accountability and transparency by reporting publicly the standards to which we hold ourselves, and publish our scorecards periodically. The PES program is both a management tool to promote internal accountability, and a valuable customer-relations tool to demonstrate the dedication of NYCT staff in maintaining a friendly and comfortable system, in good repair, of which every New Yorker can be proud.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the assistance of following individuals during the preparation of this paper: Tewfik Berri, Senior PES Analyst; Iliberth Popovits, PES PDA Administrator; Louis Balfan, Policy Initiative Analyst; Minh Tran, PES Scheduling Algorithm Developer; Autherine Jiles and Anthony Lewis, Senior PES Surveyor Supervisors; M.D. Hasan, PES PDA Developer. Responsibility for errors or omissions remains with the authors.

REFERENCES

(1) Metropolitan Transportation Authority. *Rules & Regulations Governing Employees of MTA New York City Transit, Manhattan and Bronx Surface Transit Operating Authority, and South Brooklyn Railway*. New York, N.Y. (2007).

(2) Lombardi, M.A. *Service Delivery Blue Book, Department of Subways, MTA New York City Transit*. New York, N.Y. (2002).

(3) Cranshaw, J. and Chambers, J. *A Concise Course in A-Level Statistics with Worked Examples*, Chapter 9, Significance Testing. Stanley Thornes, Cheltenham, Gloucester, England. (1984).

(4) New York City Transit. *Standard Follow-Up Reports: Service Quality Indicators*, in New York City Transit Committee Meeting Agenda. New York, N.Y. (2008).

(5) Ryus, P. *TCRP Research Results Digest, Number 56*. Summary of TCRP Report 88: A Guidebook for Developing a Transit Performance-Measurement System. Washington D.C. (2003).

(6) Kittelson & Associates Inc. *TCRP G-6 Interim Report:* A Guidebook for Developing a Transit Performance Measurement System. Washington D.C. (2002).