

# Service Quality Certification in Brussels, Belgium

## A Quality Process with Teeth

Jean-Claude Liekendaël, Peter G. Furth, and Theo H. J. Muller

European public transport operators, first in Paris and then in Brussels, Belgium, have developed a framework, a standard, and certification procedures for the management of service quality. Based on Averous's service quality cycle, this framework incorporates both customer satisfaction and performance evaluation with service standards and service indicators that are customer oriented yet objectively measurable. The development of this quality movement is documented, and its implementation in Brussels is used as a case study. There, the management contract provides financial incentives based on the percentage of passengers served by lines whose service quality has been certified, and the organization chart includes several managers responsible for service quality. The process of defining service targets is described with a framework that emphasizes extreme values and specification of both a base level and a level of unacceptability that demands immediate action. Special attention is given to two methods of performance measurement: automatic processing of archived automatic vehicle location data and "mystery shopper" surveys. Improvements, organizational changes, and new programs developed in response to the quality process are described.

A new approach to managing service quality in public transport has taken hold in Europe: public transport agencies are submitting themselves to an external program of certifying individual lines for service quality and are earning substantial bonuses for meeting the service quality incentives negotiated with their funding agencies. Two of the leaders of this movement are the public transport agencies serving Brussels Belgium, and Paris, the Société des Transports Intercommunaux de Bruxelles (STIB) and the Régie Autonome des Transports Parisiens (RATP). This paper describes the development of industry standards and certification programs for public transport service quality in Europe and how it has been applied in Brussels as a case study.

The United States has likewise witnessed a growing interest in service quality and customer satisfaction in public transport (1-2). TCRP Report 88 (3) notes that the European Union's QUATTRO research project (4), completed in 1998, proposed a framework for quality systems in public transport operations. The European expe-

rience since then, which includes adoption of a European standard, development of a certification standard, use of contract incentives, and extensive implementation in both large and small public transport agencies, offers valuable lessons for public transport agencies worldwide.

### DEVELOPMENT OF SERVICE QUALITY PROCESS STANDARDS

In the mid-1990s, pressures to open the public transport market to competitive contracting led to a recognition of the need for service quality standards that could be formally incorporated in service contracts. (Ironically, the quality movement has since been driven by publicly owned service providers.) The QUATTRO project adopted a framework for service quality management that had been developed by Bernard Averous, RATP's director of quality (5), and had been formalized in 1997 as a French experimental standard, XP-X-50-805. This framework is based on a service quality loop, shown in Figure 1. The left side of the loop emphasizes the customer's perspective: customer desires and perceptions, whose comparison determines customer satisfaction. The right side of the loop emphasizes the service provider's perspective: quality targets or service standards (which are to be based on customer desires) and measures of actual performance, whose comparison determines conformity.

When the QUATTRO project ended in 1998, its work was continued by a working group of the European Committee for Standardization (CEN), whose members include the national standards bodies of most European countries. This working group struggled with the idea of establishing standards for service quality. One idea was to use ISO 9000 standards, whose focus is on the quality of the production process, not the output. This has been described as assuming that if the hens are well cared for, they will lay good eggs. However, in transportation service, the link between production processes and output can be weak and poorly understood; so a focus on output from the customer's perspective was preferred. (As an example, for escalator availability, a common measure of mechanical reliability is down time. It seems straightforward that with small down time, escalator availability to customers should be high. However, customer surveys at STIB showed that despite little down time, passengers still complained about escalator unavailability. The source of the discrepancy—discovered only because the customer perspective suggested something was wrong—was routine maintenance, which was often performed during busy travel hours and not considered by the technical staff to be down time.) In addition, the ISO standards of 1994 (unlike those of 2000) were developed mainly for the relationship between supplier and manufacturer and did not give an important role to the customer.

J.-C. Liekendaël, Délégué Général à la Qualité, Société des Transports Intercommunaux de Bruxelles, Avenue de la Toison d'Or 15, Brussels, 1050, Belgium. P. G. Furth, Department of Civil and Environmental Engineering, Northeastern University, Room 400 SN, 360 Huntington Avenue, Boston, MA 02115. T. H. J. Muller, Faculty of Civil Engineering and Geosciences, Transportation and Planning Section, Delft University of Technology, Stevinweg 1, Delft, 2628 CN, Netherlands.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 1955, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 88-95.

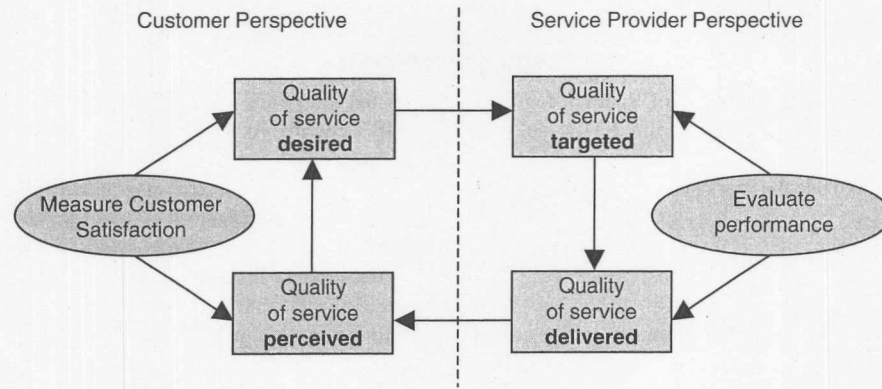


FIGURE 1 Service quality loop.

Another idea was to measure service quality directly with customer satisfaction surveys, as had been successfully done in other service industries such as tourism and as had been explored for public transport in the United States (1). However, the working group found this approach ineffective for managing quality for several reasons. First, unlike Club Med customers, most public transport users' perceptions are cumulative; therefore, satisfaction ratings respond slowly to changes in service quality. Second, customer satisfaction can be influenced by external issues such as fare increases or political scandal. Third, customer satisfaction surveys tend to be a blunt tool for identifying needed improvements. For example, the RATP found that while customers expressed dissatisfaction with station cleanliness, ratings in this category didn't go up after RATP started cleaning stations more; what ultimately improved the rating was adding more lighting. Later investigations showed that to customers, station cleanliness included lighting, odor, condition of installed equipment, and the absence of people considered undesirable (5). Customer satisfaction surveys also tend to be global in scope; this makes it hard to know where improvement is needed.

Ultimately, the chosen approach was RATP's quality loop, which forms the basis of the service quality standard adopted by CEN provisionally in 2000, and finally in 2002, as standard EN13816 ("Transportation—Logistics and Services—Public Passenger Transport: Service Quality Definition, Targeting and Measurement." Standard EN13816. European Committee for Standardization, 2002). While the framers of EN13816 gave customer satisfaction surveys a role on the left side of the quality loop, they put the greater emphasis on the right side, the side of the service provider. Accounting for technical and financial constraints as well as customer desires, an agency chooses quality targets, organizes itself to meet those targets, and measures its performance against those objective standards. This side of the loop lends itself better to management. Service standards can be defined in a way to reveal just what kind of improvement is needed. Well-defined indicators will respond immediately to a change in service quality. The feedback cycle between measurement and improvement can be very short; at STIB, some measures offer both daily and monthly feedback leading to improvements.

U.S. transit agencies also have a long history of using service standards (6). However, a distinction should be made between standards used in service design, such as geographical coverage or stop spacing, and standards used to judge the quality of service delivered. The latter is well developed for only a few criteria, notably on-time performance, where performance versus a standard is sometimes used as a financial incentive.

Completing the quality loop comes with the recognition that perceived quality may not be the same as measured quality. The quality process therefore involves customer satisfaction surveys to assess perceived quality, identify gaps between measured and perceived quality, and address them both by improving the way quality is defined and measured and by taking actions to bring customer perceptions more into harmony with actual quality. The loop is actually a spiral, cycling around every one to three years as customer expectations evolve and drive changes in service targets. The cycle is designed to "condemn" its user to continual improvement, for as service in both the transport agency and in other industries improves, customer expectations rise, leading to better refined and higher quality targets and creating a virtuous circle.

Starting in 2000, several public transport agencies interested in applying the CEN standard formed the club CyQual, which meets twice a year to share best practices. Regular participants include the transit agencies of Paris; Brussels; Berlin, Germany; Prague, Czech Republic; Geneva, Switzerland; and Madrid, Spain. A recent addition is Société de Transport de Laval (STL) in suburban Montreal, which hosted CyQual's most recent meeting.

## SERVICE QUALITY CERTIFICATION IN AND BEYOND FRANCE

In France, legislation in 1995 led to the development of a lively industry in quality certification. For French public transport agencies, a natural next step after developing a quality process was to seek certification. To meet that need, AFNOR Certification, an affiliate of the French National Standards Association, developed regulations NF281, covering all modes of public passenger transport, and NF286, with further detail for five modes of urban public transport (Service de Transport Urbain de Voyageurs. Standard NF 286. Paris: AFNOR Certification, 2002). First published in 1998, they were revised in 2002 to match the CEN standard. While much of the NF regulations simply relate to documenting compliance with EN13816, they also add specific quality requirements and standards, as explained in the case study.

One important feature of the French National (NF) regulations is that certification is by line. Quality is measured at the line level, so that some of a public transport agency's lines may be certified, and others not. When an aspect of service involves a resource shared by several lines, such as telephone information or the attitude of bus drivers, measurements are taken at the appropriate level of aggregation such as a garage or systemwide.

The French public transport agencies of Paris, Lyon, Toulouse, and several smaller cities have had one or more lines certified; this allows them to use the well-known mark NF Service. Outside France, the Madrid Metro was certified on the basis of Spanish regulations modeled after the NF regulations. In Belgium, STIB, lacking a national certification procedure, has had lines certified by AFNOR Certification, entitling them to use the mark shown in Figure 2.

## STIB BACKGROUND

The Brussels Capital Region consists of 19 communes with a combined population of 1.1 million and employment of 386,000. Since 1991, the region has contracted with STIB to manage and operate public transport. STIB has 3 metro, 17 tram, and 45 bus lines, served by a fleet of 90 railcars, 292 trams, and 571 buses. Annual ridership has grown by 20% over the past 4 years to 240 million.

Quality has a formal role in STIB's organization chart. A deputy general manager for quality has responsibility for quality programs throughout the STIB. The metro, tram, and bus divisions each include two quality managers reporting to the division director: one is responsible for service quality, and the other is responsible for quality in the maintenance and repair shops.

Management contracts between STIB and the region have given increasing emphasis to quality. The 1991 to 1995 contract required STIB to have its maintenance and repair shops certified under ISO 9000. The 1996 to 2000 contract required STIB to measure performance against service quality standards; however, those standards had little practical effect because no penalties or rewards were tied to performance.

By 2000, the European process of developing standards for public transport service quality had advanced to the point that STIB sought to have them included in its next management contract. As a result, the 2001 to 2005 management contract offers a bonus equal to roughly 0.7% of the annual budget (about 1.4% of the annual subsidy) for meeting quality targets. Targets are expressed as percentage of passengers served by certified lines, and each year's target is expressed as a range. For 2002, the target was 10% to 20%; in subsequent years targets were 35% to 50%, 65% to 80%, and finally 90% to 100% for 2005. Each year STIB gets the full bonus if the upper end of the range is achieved, no bonus if the lower end is not achieved, and a prorated amount for intermediate levels of achievement.

The progression in these certification targets suggested an obvious sequence. First, STIB focused its efforts on its three metro lines, which together serve about 45% of STIB's passengers and already enjoyed the best service quality within the network. Next, efforts extended to those tram and bus lines least affected by traffic congestion; last came the challenge of lines suffering from poor reliability.

In fact, STIB met upper bound targets in 2002, 2003, and 2004. For 2005, it is clear that STIB will achieve its lower bound; however, the upper bound of 100% remains out of reach because of bus and tram lines with severe traffic congestion problems beyond STIB's control. As discussed later, even this inability to achieve the ultimate quality certification target is a valuable part of the process.

In Paris, the RATP also negotiated to have quality targets in its 2000 to 2003 and 2004 to 2007 management contracts, which include both incentives and penalties. Unlike STIB, RATP's quality targets are not directly related to certification; instead, they are based on specific performance measures—the same measures needed for certification, except calculated systemwide (by mode) rather than by line. While RATP has no direct financial incentive to have individual lines certified, it has still chosen to pursue certification of individual bus and metro lines as a tool for achieving systemwide quality targets.

## CUSTOMER INPUT

The quality loop requires surveys to determine customer satisfaction and customer expectations, conducted annually at STIB under the name "customer barometer." Results and analysis are published on line. Special attention is paid to attributes with a low satisfaction rating and a high importance rating.

Another source of customer input, incorporated in NF regulations and required by French law governing quality certification, is a tripartite committee with representatives of customers (which in Brussels are drawn from existing citizen groups), the service provider (STIB), and the funding agency (the region). NF regulations give this committee the responsibility for approving quality targets, methods of achieving them, and methods of measurement. To keep the burden manageable, STIB divides the committee into four subcommittees, one each for metro, tram, bus, and the ticket sales agencies, which meet once or twice a year. One STIB official and one region official serve on all the subcommittees and help ensure uniform treatment of common issues.

## SELECTING SERVICE STANDARDS AND TARGETS

In the CEN framework, objectives are established only through defining performance measures (indicators) and targets. Each performance measure must be given three targets. The first, a threshold of the performance measure, is the service standard or the base service level. The second is the target level of achievement, usually expressed as a percentage of passengers receiving the base level of service. The third is a description of what constitutes an unacceptable situation. In Table 1 are some examples of measures and targets specified by the NF standard.

As the examples of Table 1 show, certification regulations specify performance measures and targets for some criteria, while for others their definition is up to the service provider, with the approval of the tripartite committee.

In the CEN framework, all measures of quality are to be from the customer's perspective. In terms of headway regularity, for example, operational measures such as the headway coefficient of variation or percentage of trips with long headways are not acceptable. Customers do not observe headways; they observe how long they wait for the next bus or train. The regularity criterion shown in Table 1 is therefore evaluated using passenger waiting time as a measure.



FIGURE 2 Mark of certification for service quality.

TABLE 1 Example Service Measure Targets

Family of Criteria	Base Standard	Attainment Target	Unacceptable Situation
Punctuality for bus lines with long headway	Choose (a) or (b): If D = scheduled departure time, a bus departs (a) between D and D+3 min, or (b) between D and D+5 min	For standard (a), the base standard is experienced by 90% of passengers; for standard (b), by 80% of passengers.	<ul style="list-style-type: none"> <li>• A passenger waits more than 20 min after the posted time.</li> <li>• A bus departs early.</li> </ul>
Waiting time for metro	Maximum acceptable waiting time, by period of the day, is to be specified.	At least 98% of the passengers have a conforming waiting time.	Unacceptable levels of waiting are to be specified by period of the day.
Bus cleanliness	The traveler has a bus that is clean, neat, and free of undesirable odors. These notions shall be defined by the service provider and approved by the tripartite committee.	80% of passenger trips conform.	<ul style="list-style-type: none"> <li>• A passenger is injured or his clothes damaged due to poor equipment condition.</li> <li>• A bus operates with graffiti more than 3 working days after notification.</li> <li>• Other unacceptable situations are to be defined.</li> </ul>

For the most part, NF regulations specify that quality be measured in terms of extreme, not mean, values of performance measures, for example, 98th percentile waiting time, or 80th percentile load. Service quality tends to be a matter of avoiding extremes; that is, high levels of crowding, long waiting times, dirty vehicles.

The need for quality targets to be based on something measurable presents a special challenge for softer aspects of quality such as driver attitude and station cleanliness. The CEN standard suggests using the "mystery shopper" method used in other service industries such as retail: surveyors behaving like customers make randomly selected trips, rate components of cleanliness, driver behavior, and so forth. Both STIB and RATP have adopted this means of measurement. While the cost of such a program is not small, once an agency has committed to it, the marginal cost of adding new quality measures can be very small. More detail about STIB's mystery rider program is given in a later section.

AFNOR Certification's framework admits three classes of criteria: obligatory, specific, and complementary. For bus, as an example, there are 10 families of obligatory criteria, dealing with

- Information at bus stops,
- Information on the outside and inside of the vehicle,
- Information about routes and schedules,
- Attitude of the driver,
- Regularity and punctuality,
- Availability of onboard equipment (e.g., means of calling for a stop),
- Vehicle reliability,
- Vehicle cleanliness,
- Passenger load, and
- Driving comfort.

NF286 also lists several areas for which establishment of specific criteria should be considered. For buses, they are

- Customer information at a distance (i.e., apart from stops and vehicles),
- Telephone service,
- Information about service interruptions,
- Cleanliness of bus stops,
- Fraud (nonpayment) control, and
- Response to written complaints.

STIB has established specific criteria for the first four of these areas.

Complementary criteria are optional and are selected by an agency in its own pursuit of excellence, perhaps as pilot measures that may in later years be integrated into specific or obligatory criteria. STIB uses two complementary criteria: one dealing with its system of real-time information on next bus arrivals and one dealing with information given to passengers during nonprogrammed service interruptions.

Like obligatory criteria, specific criteria must be defined in terms of a measure of performance, a base level, an achievement target, and a description of what constitutes an unacceptable situation. For a line to be certified, every obligatory criteria and two-thirds of the specific criteria must have attained their specified level of achievement. For complementary criteria, no level of achievement need be specified; the main issue there is to avoid unacceptable situations and to show good faith effort toward improvement.

### Standards of Unacceptability

In keeping with the notion that quality is a matter of avoiding extremes, every criterion includes definition of an unacceptable situation and a method for responding to unacceptable situations. For service providers under contract, the response may be financial penalties. Another approach, adopted aggressively in Laval, is customer redress. STL offers its customers a free ticket if a bus departs from their boarding stop either early or more than 5 minutes late, if a bus driver was not courteous or drove too abruptly, or if a passenger was inconvenienced by incorrect published information about routes and schedules. If a bus seat or bus interior wall dirties a passenger's clothes, STL will reimburse the customer's cleaning cost (7). A third approach, adopted by RATP, is giving a public account of failures. In commuter rail stations, a monthly report card of on-time performance, including an explanation of every unacceptably long delay, is posted.

Unacceptable situations can also be dealt with by established operating procedures. For example, if a STIB metro station is found to be unacceptably dirty, the station manager and the superior are to be notified and must take corrective action within 2 days.

By having both a base level and a level of unacceptability, the CEN framework establishes for each criterion three categories of performance. This appears to be a good balance between the desire to

distinguish multiple levels of performance and the desire to keep things simple and actionable. If performance is good, no action is needed; if unacceptable, immediate action is needed; if in between, it will inform annual (or shorter term) action plans within the quality improvement cycle.

### Publishing the Service Commitment

As required by the CEN framework, STIB publishes its service commitments (8), along with a less technical customer charter. It also runs two public information campaigns per year to heighten awareness of a selected aspect of the service commitment.

While STIB adopted some standards from NF286 with little or no change, others required extensive local development. For example, for the criterion bus driver attitude NF286 specifies three behaviors to be checked (stopping close to the sidewalk, looking at customers as they board and buy tickets, and selling tickets with appropriate detail). On the basis of input from customers, STIB added five more items: responding to customer inquiries, using the radio, and checking before closing doors, departing a stop, and skipping a stop.

### Integration with Enterprise Performance Measurement System

There is always a danger that if a quality process is poorly integrated into the overall management system, it will have little effect on the organization. At STIB, which has a data-driven strategic management process with 31 enterprise-level indicators, the strategic management and quality systems are integrated in two ways. First, several of the strategic indicators are exactly the same measures used for quality certification (e.g., percentage of passengers finding escalators working), except aggregated over all lines. Second, one strategic management indicator is the percentage of customers served on certified lines. This indicator effectively incorporates all the quality system's criteria into the strategic management process. As mentioned earlier, this indicator determines STIB's quality bonus.

### CERTIFICATION MECHANICS

Each year STIB documents its entire quality process, showing compliance with the service quality loop. It describes the annual action plan that resulted from analysis of conformity and customer satisfaction measurements and accounts for each action's implementation. Documentation requirements are stricter for new lines submitted for certification than for lines already certified.

AFNOR Certification professional staff reviews the submitted documentation, audits documentation and procedures in 1-to-2-week visits, and makes the ruling as to whether each line submitted for certification qualifies.

### MEASURING SERVICE RELIABILITY

Measuring performance from the customer viewpoint can be a challenge. This section and the next describe two interesting aspects of STIB's measurement program.

At STIB, service reliability is defined in terms of regularity for metro and for surface lines whose headway is 6 minutes or less and in terms of punctuality for surface lines with headways over 6 minutes. Its measurement system uses archived automatic vehicle location (AVL) records.

For punctuality, NF286 offers a choice of two standards: 80% of passengers served in an on-time window of 0 to 3 minutes late, or 90% of passengers in an on-time window of 0 to 5 minutes late. For bus service, STIB uses the former.

Of course, direct punctuality measurements are made for vehicles, not passengers. The translation to how many passengers were served in an on-time window has two complications. One is straightforward: the punctuality of every vehicle at every station is weighted by passenger demand by station and time period. The other is trickier. Suppose a bus route has 10-min headways, and every bus is running 12 min late. No bus is operating in its on-time window, yet all passengers, except those coming to meet the first bus, will find a bus within their on-time window. Therefore, the computation task is not to match observed departures to their scheduled departure times, but to match scheduled departure times to the next (chronologically) observed departure. Because STIB's analysis software still uses the former matching, its measurement of performance is actually worse than what passengers observe.

For regularity, as mentioned before, the performance measure is passenger waiting time: What fraction of passengers waited longer than a standard  $s$ ? If scheduled headway equals  $h_{\text{schedule}}$  and service is perfectly regular, no passenger should have to wait longer than  $h_{\text{schedule}}$ . Assuming that customers expect a small level of irregularity, the standard for high frequency surface lines is  $s = h_{\text{schedule}} + 2$ ; for metro,  $s$  is set by the tripartite committee at a value usually 1 to 2 min greater than  $h_{\text{schedule}}$ .

Assuming uniform passenger arrivals during each headway, the fraction of passengers waiting longer than a value  $s$  can be determined from headway lengths, which are, in turn, calculated from AVL records. If a headway's length  $h$  is less than  $s$ , nobody arriving during that interval waits longer than  $s$ ; otherwise, the fraction  $(h - s)/h$  of passengers wait longer than  $s$ . For example, if  $s = 6$  min, then if  $h = 5$  min, nobody's wait exceeds the standard, while if  $h = 8$  minutes, 25% of that trip's passengers (those who arrived in the first two minutes of the interval) waited more than 6 minutes.

Each month STIB's AVL software calculates this fraction for each departure and aggregates it over all departures from a given station in each half-hour period, with the enterprise database being populated with the fraction of passengers meeting the waiting time (or punctuality, as appropriate) standard for every station per day per half hour. Within the enterprise database, results can be aggregated to the line or system level, weighting by boarding rates for each station per half-hour.

For bus lines, NF286 specifies that 90% of passengers should wait less than the standard level, while for metro the specified level of achievement is 98%. Clearly, this reflects the idea that customers expect a metro to be more reliable than a bus. Of course, the same effect could be achieved by the waiting time standard being varied rather than the level of achievement. Elsewhere (9) it is argued that because of the important role that 95th percentile waiting time plays in passenger travel budgets, it may be better to define waiting time standards for both modes using target achievement levels of 95%.

## MYSTERY RIDER PROGRAM

To measure qualitative aspects of service quality, STIB uses mystery rider surveys. In this program, surveyors make trips selected at random, behave like a customer, and rate various aspects of service following a checklist. Mystery riders are not inspectors. For example, they don't search in stations for litter or broken equipment; they wait as a passenger would wait and observe what a waiting passenger might observe. As part of their trip, they seek information in advance, purchase a ticket (or observe other passengers purchasing tickets), ask (or observe others asking) a driver or station attendant for information, and so forth.

Mystery riders rate items that are components of various criteria. For some criteria such as posted information, the rating for each item is binary (e.g., is a system map posted?). When aggregating over binary items, the composite measure conforms only if every component item conforms.

Other criteria such as station cleanliness and comfort of the drive are composed of multiple items that, being more subjective, are rated on a scale of 0 to 20. For these criteria, the composite rating is a weighted average of the component ratings and is judged to be in conformance if the composite rating is at least 80% (16 out of 20) and no unacceptable situation was detected.

When STIB first began certification measurements, weights used to generate composite ratings were selected by STIB staff. Later, special purpose customer surveys were conducted to update those weights to reflect the importance customers give to various items. The survey showed considerable divergence from what the STIB staff had assumed. For example, for the criterion vehicle cleanliness, metro customers cared less about graffiti, and more about the floor being free of litter and the seat being clean. When the new weights were first applied, conformance levels dropped immediately—because STIB, following its assumptions about what was important to customers, had been giving a lot of attention to removing graffiti, not to removing litter from vehicles. In response, STIB took action by having its cleaning subcontractor add a team at each metro terminal to give the cars a quick sweep at every layover; that quickly brought the metro lines back into conformity.

A challenge is to make mystery rider measurements as consistent and objective as possible. Surveyors are trained to understand what rating should be applied for various situations, including both classroom training (using photos) and field training under the supervision of experienced surveyors. To preserve objectivity, the program is contracted out to two market research companies.

During the course of a mystery ride, surveyors fill in a station checklist for every station visited and a vehicle checklist for every vehicle used. A special checklist is used on trips with a programmed service perturbation to check on the quality of information given to passengers about the service change. The rating sheets are scanned, the data are stored in a database, and composite scores are calculated. If any criterion is not in conformance, or if an unacceptable situation was detected, a report is sent automatically to the relevant managers within 2 days of the mystery ride. The report includes the full detail of the rating sheet, plus any comments the surveyor may have written. This rapid feedback helps motivate managers to correct situations needing it immediately and to keep operations focused on service quality.

Mystery rides are conducted year-round, at a rate of 1950 per year on bus, 750 on tram, and 900 on metro. Sampling is done in a way

to assure representation on every line. Within a line, trip origins, destinations, and times of day are selected to reflect actual customer travel patterns.

Each month, mystery ride data are uploaded to STIB's enterprise database, which managers can query to generate service quality reports for any desired group of criteria, lines, and date range. The data can be viewed at any level of aggregation, including drilling down to the original rating sheets. With the results integrated into the enterprise database, it is easy to weight results with passenger demand by line, station, and time of day; to create reports combining measures based on mystery rides and measures based on AVL data; and to generate reports for line-level certification, systemwide performance, or at whatever level appropriate for a before and after study.

## IMPACT OF THE QUALITY PROGRAM

STIB's quality program has created a noticeable mind change within the agency. In interviews with STIB staff in a variety of departments, one frequently hears that one or another action has to be taken to meet quality goals. Thanks to the decentralized certification process and the measurement and feedback systems implemented, the motivation of financial incentive felt by upper management has been successfully passed down to middle-level managers and to employees.

### Changes to Organizational Structure

Mobilizing to achieve service quality certification involved some changes to STIB's operational structure. Where there had been a single operating division responsible for all three modes, STIB created separate operating divisions for metro, tram, and bus to focus responsibility better for performance. With a service quality manager assigned for each mode, the new structure tripled the number of service quality managers in operations.

Control center operations were likewise decentralized. Before, service regulators focused mostly on metro operations; with only three bus dispatchers, little attention was paid to bus unless there was a major problem. The bus division is now responsible for its own operational control. It has a new control center with 12 regulator employees, who not only respond to major problems but also take smaller actions to improve regularity and punctuality. Likewise, responsibility for posting information about service interruptions was decentralized over the three modal divisions.

The metro division was reorganized by appointing a supervisor for each line and station managers for groups of two to four stations. One now sees station managers sometimes pick up litter or empty a trash can themselves—something previously not imaginable.

Decentralized responsibility is reinforced by the rapid feedback from the mystery rider program, as managers must be able to show that they respond to every detection of an unacceptable situation.

### Adjusting Scheduled Running Times

Normally, the terms of STIB's management contracts discourage making upward adjustments to scheduled running or recovery time because payment is based on vehicle kilometers, not vehicle hours. With actual running times and running-time variability increasing

because of traffic congestion and a new policy requiring passengers to board through the front door, punctuality was poor on many bus routes. Previously, this poor punctuality might have been tolerated. However, the need to meet punctuality standards to win certification forced a complete review of bus running times. The adjustments required an additional 46 buses in peak hours.

The new importance of the punctuality standards also prompted some new thinking about how scheduled running times are chosen. Until now, schedule makers used mean running time by segment, with ad hoc increments to try to deal with reliability problems. Recovery times were based on simple rules of thumb. In 2005, STIB's chief engineer for operations studies began experimenting with a new statistical approach that, like service quality standards, is based on extreme rather than mean values. It uses three target parameters, best explained in an example. If the target parameters are 5 min, 80%, 95%, a trip's one-way running time will be scheduled so that, assuming buses begin on time, 80% will be able to reach the end of the line no more than 5 minutes late, and half-cycle time (running time plus recovery time) will set such that 95% of the buses can start their next trip on time. It is still too early to evaluate this approach.

### New Programs

Because service quality certification became an enterprise objective, managers at every level were asked what they could do to help achieve it. At the same time, the bonus STIB got for meeting quality standards permitted managers to apply for funding for new programs. As a result, several new programs were initiated to improve service quality. One already mentioned is the all-day train sweeping at metro terminals. Three more examples follow.

A customer representative on the tripartite committee requested that every station have ticket validators relocated to an elevation of 80 cm for the convenience of passengers in wheelchairs. This desire was adopted as a quality target and added to the mystery rider evaluation grid. That immediately put all metro stations out of conformity; this motivated line and station managers to see to it that their validators were quickly repositioned. In less than a year all stations came into compliance. A similar result was recently reported for Prague's public transport agency, where the percentage of passengers served in stations meeting accessibility standards rose from 30% to 87% in a year following adoption of the standard by the quality program (10).

While bus stops are owned by the Brussels communes, their appearance and upkeep affect STIB's quality ratings. Therefore, STIB created a database for tracking complaints and work orders at all its bus stops and positioned itself between customers and the responsible party (either the commune or the private contractor responsible for shelters) to ensure that problems at bus stops are quickly repaired.

On several bus and tram lines, peak hour traffic often blocks intersections, creating enormous delays to both buses and general traffic. To improve bus service punctuality, STIB took the initiative of contracting with several Brussels communes to station police officers at key intersections to enforce the law against entering a blocked intersection. STIB pays 70% of the cost.

### Focus for Congestion Relief Programs

As STIB nears the end of its 2001 to 2005 management contract, it is clear that despite its best efforts, some bus and tram lines cannot be certified for quality because of severe traffic congestion. The

quality program helps put a focus on those nonconforming lines. STIB hopes that the region's goal of achieving 100% certification will put pressure on the agencies responsible for roadways to develop congestion relief programs (enforcement, exclusive lanes, better traffic signal management, traffic restrictions) where needed.

In 2005, STIB began an annual program of compiling a report of hot spots (points noires) where buses and trams suffer substantial traffic delays. It shows how speeds at about 50 locations drastically decline during peak and sometimes midday hours and estimates the operating cost savings that could be achieved by congestion relief. This report also emphasizes the customer benefits in riding time and punctuality that would accrue from congestion relief.

### CONCLUSION: LESSONS LEARNED

From the Brussels experience, five lessons stand out that may be helpful to others:

1. The importance of a financial incentive. Having a financial incentive in the management contract mobilizes the entire agency staff, because it makes achieving service quality something that matters to the entire enterprise and gives everyone a sense of responsibility for its success.
2. The value of external certification. External certification lends legitimacy to the quality process, both within the agency and to the funding government authority. It provides a structured approach to achieving service quality by specifying targets and requiring procedures for measurement, reporting, and responding. The fraction of customers benefiting from certified service becomes a simple, comprehensive, and powerful measure of quality achievement that is well suited to financial incentives and strategic management.
3. The need for reliable measurement. Measurement has a magical power to improve quality by creating awareness and providing positive feedback for improvements. The simple fact that things were being observed and recorded resulted in many problems disappearing when the certification program was first launched. However, it is vital that the measurement system be seen as fair and accurate or employees will dismiss its results. STIB and RATP invest a considerable sum annually in mystery rider measurements, and STIB invested a lot in developing its archived AVL data processing and database systems. Both agencies are participating in drafting a new European standard (EN15140) for measurement in support of public transport service quality.
4. The power of a service-quality focus to reorient an organization as customer-centered. Defining service quality standards using the customer's perspective focuses the entire agency on serving customers, bringing a uniting and energizing sense of mission that transforms the organization. In a field long dominated by tension between labor and management, a focus on service quality introduces a third party above both, to whom both labor and management feel an obligation.
5. The possibility of extending the service quality framework to authorities responsible for roads and traffic signals. Because travel speed depends mostly on roadway and traffic conditions beyond the control of the service provider, it is not a criterion in STIB's quality program. Nevertheless, speed improvements are essential for making public transport more competitive. Likewise, only roadway authorities can make many of the improvements needed for service reliability. Such improvements should be guided and evaluated by customer-oriented service quality measures. Engaging the roadway authorities

in a customer service quality process is an exciting frontier that STIB is currently pushing.

## ACKNOWLEDGMENTS

This research was funded by the Brussels Capital Region under the program Research in Brussels. Thanks also to Pierre Weil, executive officer for service quality at RATP, for helpful information.

## REFERENCES

1. MORPACE International, Inc., and Cambridge Systematics, Inc. *TCRP Report 47: A Handbook for Measuring Customer Satisfaction and Service Quality*. TRB, National Research Council, Washington, D.C., 1999.
2. Kittleson and Associates, Inc.; KFH Group, Inc.; Parsons Brinckerhoff Quade and Douglass, Inc.; and K. Hunter-Zaworski. *TCRP Report 100: Transit Capacity and Quality of Service Manual*, 2nd ed. Transportation Research Board of the National Academies, Washington, D.C., 2003.
3. Kittleson and Associates, Inc.; Urbitran, Inc.; LKC Consulting Services, Inc.; MORPACE International, Inc.; Queensland University of Technology; and Y. Nakanishi. *TCRP Report 88: A Guidebook for Developing a Transit Performance-Measurement System*. Transportation Research Board of the National Academies, Washington, D.C., 2003.
4. Organisation-Gestion-Marketing, S.A. *QUATTRO Project Final Report*. European Commission Transport Research Program, 1998. [www2.eur.nl/quattro/uk/q\\_report\\_00.html](http://www2.eur.nl/quattro/uk/q_report_00.html). Accessed June 27, 2005.
5. Averous, D., and B. Averous. *Mesurer et Manager la Qualité de Service: la Méthode CYQ*, 2d ed. INSEP Consulting Éditions, Paris, 2004.
6. Benn, Howard P. *Bus Route Evaluation Standards*. Synthesis of Transit Practice 10, TCRP, National Academy Press, Washington, D.C., 1995.
7. Société de Transport de Laval. Our Quality Commitment. [www.stl.laval.qc.ca/gq/garantie-en.shtm](http://www.stl.laval.qc.ca/gq/garantie-en.shtm). Accessed June 28, 2005.
8. Société des Transports Intercommunaux de Bruxelles. Service Référence Metro, Service Référence Tram, and Service Référence Bus. [www.stib.irisnet.be/FR/12090F.htm](http://www.stib.irisnet.be/FR/12090F.htm). Accessed June 28, 2005.
9. Furth, P. G., and T. H. J. Muller. Service Reliability and Hidden Waiting Time: Insights from Automatic Vehicle Location Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1955*, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 79–87.
10. Dozek, Z. Qualité et Management. Presented at the CyQual Colloquium, Laval, QC, Canada, May 2005. [www.stl.laval.qc.ca/colloque/presentations.shtm](http://www.stl.laval.qc.ca/colloque/presentations.shtm). Accessed June 28, 2005.

---

*The Transit Capacity and Quality of Service Committee sponsored publication of this paper.*