

A New Index of  
Telephone Service Quality:  
Academic and Regulatory Review

Sanford Berg

Do not quote without permission of the author.  
c. 1993. Columbia Institute for Tele-Information

Columbia Institute for Tele-Information  
Graduate School of Business  
809 Uris Hall  
Columbia University  
New York, New York 10027  
(212) 854-4222

**A New Index of Telephone Service Quality:  
Academic and Regulatory Review**

by Sanford V. Berg\*

**Abstract**

*Historically, state regulatory commissions have used large batteries of pass-fail standards to evaluate the quality of local telephone service. This paper describes the development of an alternative approach to measuring, evaluating, and rewarding service quality. The ideas presented here have evolved through two different processes--academic research and regulatory implementation. These processes have some similarities and differences which complicated our objectives of developing an index which was both analytically sound and policy relevant.*

To be presented at the Columbia Institute for Tele-Information  
Conference on Quality Choices in Network Industries,  
April 23, 1993

**1. Introduction**

The academic ideal, for many scholars, is to do research that is simultaneously theoretically elegant and solves important practical problems in the real world. In this paper, we provide a case history of our own attempt to pursue this ideal. Our work focused on the regulatory measurement and reward of service quality provided by regulated local telephone service companies operating in the state of Florida. We describe the surprisingly nonoverlapping challenges we have faced in (a) maneuvering a relatively straightforward regulatory innovation into regulatory practice, and (b) disseminating the conceptual framework to another community (through publication in scholarly journals read and edited primarily by academics).

This paper describes one effort which has explored the feasibility of a new service quality index. Generalizations from one observation are problematic--but even one data point can shed light on the issues associated with the adoption of rules and procedures which improve resource allocation. Both sound theory and supporting empirical evidence are necessary if approaches to quality are to be strengthened. We offer some low-brow theory and observations on actual regulatory behavior to draw lessons regarding service quality.

---

\*Florida Public Utilities Professor, University of Florida. Helpful comments from John Lynch are acknowledged (without implicating him).

## 2. Overview of Current Regulatory Schemes and the Proposed Index

The importance of service quality has been highlighted by developments in the last decade: divestiture, network interconnection, and technological change. The competitive and complementary service offerings of new entrants raise challenges for incumbent local exchange carriers. Meanwhile, state regulators are still faced with the choice between traditional cost of service regulation and various forms of incentive regulation. Whatever their decision, the role of quality has a higher profile than in the past.

The regulatory process utilizes technical performance features of networks, even while recognizing that consumer satisfaction may depend only indirectly on the engineering measures of service quality. These surrogate evaluations tend to consist of pass-fail standards which were often established decades ago. They have grown by accretion: the most recent NARUC compendium on the subject lists between 90 and 100 separate standards (depending on how one groups some sub-categories).

Quality of service targets are somewhat arbitrary, having arisen from a chaotic process reflecting historical engineering capabilities, political pressures, and administrative happenstance. Consumer valuations of different quality dimensions and corporate recognition of emerging technological opportunities are not likely to be captured by pass-fail standards. In addition, combining information on multiple dimensions into an overall assessment is very difficult for regulators. Information overload could lead to "management by exception." By focusing on the rules that company fails, regulators essentially ignore dimensions on which the company being evaluated has exceeded the standards. Perverse incentives result. Developing an appropriately weighted quality-of-service index is no simple task, but the approach represents a potential improvement over multiple pass/fail quality standards.

### 2.1 Production Possibilities for Pass-Fail Standards

Current reward schemes, in Florida as in other states, compare a company's objective scores,  $Z_1 \dots Z_n$ , on a set of engineering attributes to standards,  $Z_1^* \dots Z_n^*$ , set by the regulatory agency on those attribute dimensions. Below standard performance on any attribute triggers censure, while performance above any standard is not treated as being any better than performance exactly at the standard. This tacitly drives companies to produce quality along each attribute at exactly  $Z_i^*$ , since exceeding  $Z_i^*$  generally takes resources, but is not rewarded in the regulatory system. Beyond this, however, very little could be said about exactly what regulators were rewarding or trying to reward, as they used subjective and intuitive judgment to assess the overall service quality of a firm that had a complex set of above-standard and below-standard scores on the many measured dimensions of quality.

We attempted to make their expert judgment more systematic by modeling regulators' tradeoffs among various dimensions of quality,  $Q = f[(Z_1 - Z_1^*), \dots, (Z_n - Z_n^*)]$ . We discovered that regulators often agreed that overall quality was higher when standard A was exceeded and B was failed (when attribute A was considered relatively more

important) than when A and B were met exactly. Therefore, existing evaluation policies were creating perverse incentives by treating the former company as being in violation and the latter as being in compliance with regulations.

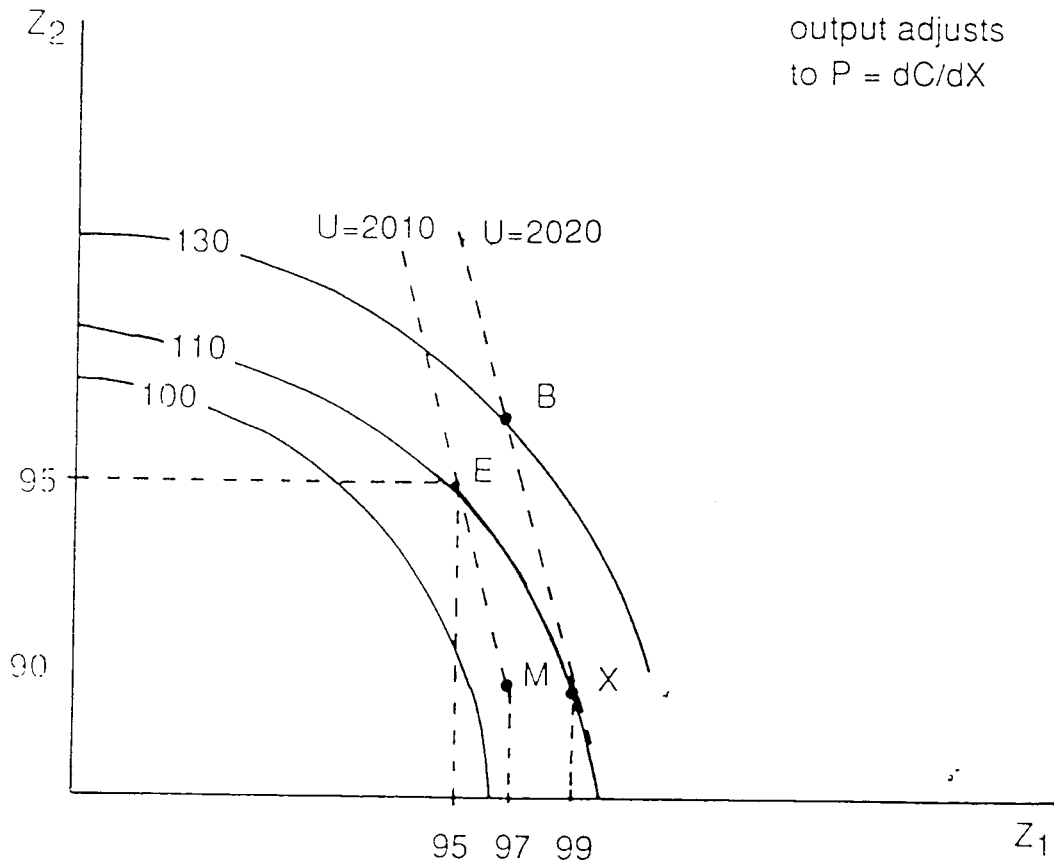
In our proposed alternative index, we define  $\hat{Q}^*$  to be the predicted level of overall quality associated with meeting all  $n$  quality standards exactly. We then propose that any combination of  $Z_1 \dots Z_n$  that leads to  $Q \geq \hat{Q}^*$  should be treated as meeting the standard. Under this regime, companies would offer whatever combination of substandard and superstandard  $Z_1 \dots Z_n$  that allows them to achieve  $\hat{Q}^*$  in the most efficient way given that company's cost structure. For each 1% point above a standard, the quality index rises--depending on the weight given to that particular standard. Similarly, shortfalls result in reductions in the index.

For simplicity, consider two dimensions of service quality monitored by regulators: dial tone response ( $Z_1$ ) and call completions ( $Z_2$ ). Like other commissions, the FPSC has standards for each of these. Florida requires that 95% of all calls receive a dial tone within 3 seconds. The intra-office call completion standard requires the successful completion of 95% of all calls to numbers with the same first three digits as the call. A welfare-maximizing regulator will induce the firm to equate the marginal benefits from service quality improvements with the marginal costs. This condition for optimality is depicted in Figure 1 (adapted from Berg and Lynch, 1992). Three production possibility frontiers are shown--those which are further out require that additional resources be devoted to the production of quality: \$100, \$110, and \$130, respectively. For a given level of real resources, improvements in one quality dimension involve a deterioration in the other. The PPF represents engineering and resource constraints.

## 2.2 Relative Valuations of Service Characteristics

Relative valuations for the two dimensions of service quality are also shown in Figure 1. In this example, the subjective trade-offs by customers are reflected in the preference mapping characterized by  $U=2010$  and  $U=2020$ . That is, for any given level of satisfaction (for example,  $U=2010$ ), if one dimension deteriorates (say  $Z_2$ --call completions--falls from 95% to 90%), then  $Z_1$  must increase if customers are not to be made worse off. Here,  $Z_1$  (dial tone response) must rise from 95% to 97% if the customers are to remain on  $U=2010$ .

In this example, points E and M represent the same level of satisfaction, met by different combinations of service qualities. Point E would not be a welfare maximizing point since point M is valued equally by consumers and costs less to achieve. At point E, the subjective marginal rate of substitution between  $Z_1$  and  $Z_2$  does not equal the marginal rate of transformation (as reflected in the slope of the production possibility frontier). The proposed approach would drop the pass-fail standards of 95%, 95%, and give the telephone company flexibility in selecting least cost ways to achieve a given level of performance. In the example, point X could be achieved for the same resource cost as point E--but benefits would now be  $U=2020$ .



**Figure 1**  
**Relative Valuations**  
**of Quality Attributes**

To illustrate how firms could be presented with a regulatory objective function--and allowed to trade off high cost (low valued) quality dimensions for low cost (highly valued) quality dimensions. In the simple example, if (95,95) yielded an "acceptable" overall level of quality, one scoring function which would signal the telco to modify its quality mix would be  $Q = Z_2 + (5/2)Z_1$ , and the minimum quality "score" is  $Q = 332.5$ . The firm would go to point M (97, 90).

One issue is whether the minimum quality "score" is appropriate. In the simple example, if  $Q = 337.5$ , the firm will be driven to point X, instead of point M. If point M corresponds to \$105, and an additional \$5 lets the firm attain point X, then the outlay is worth it if X is valued \$5 or more than the quality bundle at M. We did not try to attack the incremental cost issue, but focused on replacing the (95,95) standard with a minimum quality score of 332.5, with the weights as described above.<sup>1</sup>

This is an extremely simple idea that can be easily shown to represent an improvement over the current regulatory regime--so long as there is agreement as to the weights to be given the various standards (Noam, 1991). Yet the legal, political, and institutional roadblocks to its full adoption and appropriate implementation have been numerous. At the same time, we have published and presented this work in several academic venues. However, the primary theoretical piece is still battling to emerge from a four-year review process at a prestigious academic journal that strives to merge theory and practice.

### 3. Scientific Review and the Creation of Knowledge

Publication lags, like regulatory lags, arise from the existence of numerous check-points where stringent review criteria are applied. Academic gate-keepers provide critical reviews of analyses--giving readers some confidence that the published article represents a contribution to the literature. External reviewers can help researchers focus their efforts and remedy potential flaws in analyses. The review process screens potential contributions, asking whether the submission contributes new and creative insights regarding the issue at hand. A second, and perhaps more fundamental question is whether the issue under consideration is actually important.

Theoretical constructs, empirical tests, and historical evaluation provide the three legs upon which a policy science stands. Given the gains from specialization and division of labor, researchers will tend to tackle issues from one of the three perspectives, although good analysis using any one of the three modes cannot ignore the other two. For example,

---

<sup>1</sup>Note that the scoring function presented earlier is merely a transformation of the constraint equation described here:

$$Q_a = W_0 + W_1(Z_1 - Z_1^*) + W_2(Z_2 - Z_2^*)$$

$$Q_a - (W_0 - W_1Z_1^* - W_2Z_2^*) = W_1Z_1 + W_2Z_2$$

$$Q = W_1Z_1 + W_2Z_2$$

good theory recognizes the historical setting which establishes the institutional context for theoretical analysis. In addition, theory often depends on empirical observations (in the form of stylized facts) to provide bounds on key parameters or determine the signs of particular relationships. Reviewers of potential contributions know the economic paradigm from which models are derived and quantitative tests conducted. Deviations from the widely-accepted neoclassical economic framework face a hurdle in the review process. Since our work does not draw heavily upon the paradigm of the rational, evaluative, maximizing consumer, we have had to justify the framework to skeptical academic reviewers.

Our concern here is with the evaluation of telecommunications service quality. The literature on product (or service) quality is voluminous. Since the theory is summarized elsewhere (Berg and Lynch, 1992), it will not be surveyed here. Suffice it to note that the models are elegant, often insightful, and difficult to test. Quality outcomes under competition, monopoly, and regulation depend on a host of factors, including incremental costs, incremental benefits, and average benefits associated with quality changes. Furthermore, the introduction of multiple dimensions of quality greatly complicates the analysis.

Nevertheless, the neoclassical paradigm suggests that informed consumers will evaluate alternative service offerings and select consumption bundles based on their preferences. Thus, the ultimate judge that matters is the rational customer--not some regulatory surrogate. We agree. However, policy analysis cannot abstract from the institutional context: the past matters. Regulatory reviews are designed to screen for a different set of problems than those which might concern scientists. Our work has tried to be responsive to both review processes. Indeed, our studies have benefitted from discussions with academics and regulators actively participating in both processes.

#### **4. Regulatory Review and Procedural Fairness**

The regulatory review process is grounded in procedural fairness. Disruptive transitions are costly to buyers and sellers alike--so administrative delays can sometimes serve as mechanisms for smoothing out the impacts of changes in demand or technologies. Review lags ensure that proposed changes are understood by all those affected by new regulations. In addition, administrative procedures are designed to provide opportunities for complaints to be heard. Hearings and informal workshops serve as forums in which stakeholders present their concerns.

The heavy role of legal, accounting, and engineering expertise at Public Utility Commissions suggests that economics by itself provides an inadequate foundation for regulatory decisions. Emphasis on legal precedent gives continuity to the rate-making process--forcing some consistency in the face of emerging problems. Protection is afforded both the regulated firm and its customers. From the standpoint of service quality, fairness towards customers requires that data be verifiable. If reported data are fabrications or the result of improper manipulation of data collection procedures, the firm loses credibility. Credibility is essential if the regulatory process is to be accepted by consumers.

Similarly, the use of accounting data constrains much of the debate to the consideration of actual outlays--rather than hypotheticals. Although future test years are utilized in many jurisdictions, the focus is still on accounting rather than economic costs. Economic opportunity costs may be considered, but accounting data and cost allocation procedures based on historical developments dominate rate cases so long as there are no alternative suppliers. In addition, engineering data are particularly relevant for considering quality of service issues, since these are objective and subject to review.

The regulatory concern is that regulated (or partially regulated) firms may choose the wrong level and mix of quality--recognizing that quality is multidimensional. But what does "wrong" mean? Too little quality? Too much quality? An inappropriate mix of quality components? Inappropriate pricing of quality components?

The framework developed here is appropriate for encouraging the right mix of quality characteristics. By itself, it does not address the overall level of quality. However, the current battery of pass-fail standards addresses neither the mix nor level of quality. By building upon current data collection efforts, maintains continuity. In addition, the proposed framework allows regulators to formalize what they mean by quality. The procedure for soliciting weights has its limitations. However, the weights can be refined. With the adoption of a quality index, managers can make network investment and operations trade-offs based on precise weights applied to a specified set of characteristics. This represents an improvement in the process, since the current pass-fail targets are subject to uneven regulatory application. Such a process can be quite inefficient; it also potentially unfair.

Thus, from the standpoint of procedural fairness, the proposed quality index is quite promising. As with any new instrument, however, telecommunications firms will be hesitant to accept new rules. For example, two problems with the movement to price caps are the determination of the starting price level and the calculation of the productivity adjustment. Both would have to be determined in advance before a telco would support the replacement of rate of return on rate base regulation with price caps. Similarly, before firms will accept a new quality index, they will want to know how it is to be used in the regulatory process. If it is seen as another tool for bludgeoning the firm, the index will not receive their support. Strong opposition by regulated firms reduces the likelihood of adoption.

Intervenors will also be skeptical of any departures from current conventions. For example, the Public Counsel's Office will primarily be representing residential customers. The weights appropriate for these customers may not be the same as those for high volume commercial and industrial demanders. Since the dimensions of quality tend to be collective consumption goods (quality available to one is available to all), consumer advocates may not want to give the firm the discretion involved in meeting an overall quality index constraint. Rather, they may prefer to focus on items of particular concern to their constituency.

Given these observations regarding the focus on continuity and aversion to change, stakeholders will delay adoption of new instruments by regulatory agencies until they have



**Figure 2**  
**Regulatory and Academic Reviews**

1986

FPSC/PURC Discussions

1987

Methodology Developed

FPSC Data Collected

1988

Draft Report

Telco Data Analysis

Draft Article

1989

PURC Seminar  
"Final" Report

FPSC/PURC Meeting

Revised

Submitted/Conferences

1990

FPSC Hearings on Service  
Interruptions

Data Collection

Rule Making

Revised Report

Reviews Received

1991

Public Counsel Files

PURC Testimony

Recommendation Deferred

Revised

Resubmitted

1992

FPSC Staff Accepts Index

Reviews Received

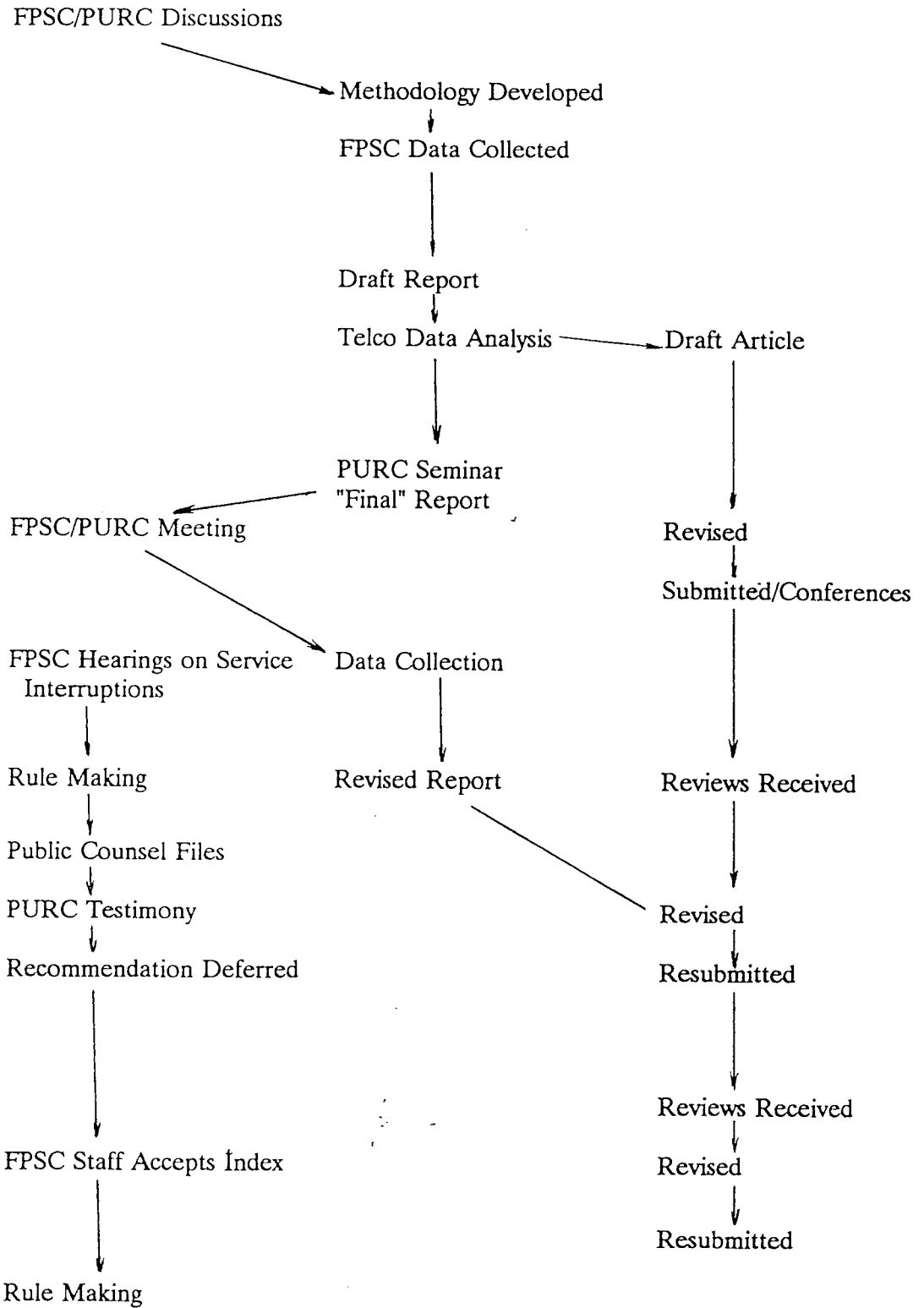
Revised

1993

Rule Making

Resubmitted

1994



made thorough checks on implications for performance. Regulators, utility managers, and consumer advocates will all need to be comfortable with the new approach.

## **5. Case Study of a New Quality Index**

The rationale behind the proposed quality index, Q, has already been described. Lessons can be learned by reviewing the parallel evolution of the conceptual framework and associated regulatory rules. Let us begin at the beginning: the investigation was initiated in late 1986, when FPSC staff approached the Public Utility Research Center (PURC) about exploring ways to evaluate quality of service provided by local exchange companies. Based at the University of Florida, PURC attempts to bridge principles and practice--so the issue clearly fit into its purview.

### **5.1 Initial Data Collection**

Academic institutions are not consulting groups who can switch resources rapidly from one activity to another. Two marketing professors had ideas about how a more comprehensive indicator of quality might be developed. Interactions with FPSC staff yielded what we thought was a good understanding the pass-fail standards applied to telcos. Teaching responsibilities and other research commitments meant that much of the initial investigation occurred in the summer of 1987. Figure 2 outlines the stages of regulatory and academic reviews associated with the proposed framework.

The basic methodology involved a survey in which experts made comparisons of quality bundles. The weighting scheme was developed by having experts from the FPSC rate different hypothetical company profiles of performance on the rules within nine Rule Clusters. Each profile was rated on a scale from 1 (worst possible performance) to 10 (best possible performance). Similar comparisons were made across rules, so that a comprehensive score could be assigned to a telco based on its observed performance on the 38 dimensions. The entire procedure is a form of conjoint analysis called "Hierarchical Conjoint Analysis (Louviere, 1984). This approach is suited to capturing the trade-offs experts make in overall evaluations of objects that can differ on a very large number of attributes that can be logically grouped into subsets of related attributes.

In January 1988, a report was ready for the FPSC (Buzas and Lynch, 1988). The methodology for determining weights to be given the various dimensions was outlined in some detail, and associated statistical tests presented. It provided illustrative calculations for hierarchical conjoint analysis--so the derivation of individual weights could be described. The study analyzed how agreement and disagreement among survey participants could be identified. The report included a discussion of each item on the questionnaire, defined technical terminology, and showed which of the 38 dimensions had greatest weights--based on responses provided by FPSC staff. The formula derived showed the weight of a one percentage point change on each of the 38 dimensions.

The three member research team prepared an academic working paper in July 1998. This paper was the first project output focusing on both the methodological and policy issues associated with the new index. It provided illustrative calculations for hierarchical conjoint analysis--so the derivation of particular weights could be described. The implications of agreement and disagreement among survey participants were also outlined. The report contained examples of techniques for determining the reliability of estimates.

Finally, limitations to the study were identified. In particular, the separate role of costs was discussed. Also, since the initial trade-offs were made by FPSC employees, we noted that these experts might have a different perspective than either consumers or companies. Third, the trade-offs were made without reference to specific locations or clientele. It is plausible that the PSC might want to reward compliance on some rules more heavily or lightly in certain geographic areas. For example, compliance on public telephone dimensions might be more important in rural areas than in urban areas because the phones are further apart in the former.

We needed to determine whether experts at telephone companies would give the same weights to the various rules. We thought that telco representatives might be more aware of the relative benefits of meeting the different rules. Alternatively, different firms face different mixes of customers (due to demographics, per capita income, degree of urbanization). Such factors could mean that customer valuations for different components of quality differed across firms--making a single weighting scheme inappropriate. Despite the remaining issues, we believed that the results were highly suggestive: progress was being made in this very complicated area. The research team viewed the methodology as offering a way to introduce greater rigor and content into the quality evaluation process. The project turned to the issue of telco weights.

The data collection effort moved forward in earnest, as we sought cooperation from regulated firms. Some were willing to devote personnel to the effort, but others were concerned about implementation issues. A PURC Seminar was held in February 1989. Berg, Buzas, and Lynch described the rationale and methodology behind the comprehensive index. Results from the FPSC sample (12 employees) and two companies were presented. Alan Taylor, Chief of the Bureau of Service Evaluation, represented the FPSC. Also in attendance were representatives from the major Florida telcos. These representatives expressed a concern as to how the proposed index might be used: to compare firms at a single point of time? to evaluate trends over time for a single firm? to compare firms? The different service territories and degree of network modernization influence the starting point for each firm, raising a concern for fairness. Cross-firm comparisons might not recognize potentially different technological opportunities. Executives tended to see the index as another stick that would be used for hitting them at the next rate case.

Most of the formal presentation focused on basic methodological points:

- (1) Different dimensions had different weights.

- (2) Only a few dimensions really mattered a great deal.
- (3) There was a very high correlation (Pearson correlation coefficient of better than .90) between weights obtained from the two companies and from the commission.

A hypothetical example was given, and the method of calculation presented. To illustrate the usefulness of a comprehensive index, data for two hypothetical companies were presented, emphasizing the difficulty of making pass-fail comparisons. The politics of regulation were not given much attention, although attendees were probably more worried about implementation issues than statistical refinements.

Within weeks, the Final Report on Telephone Service Quality was sent to the FPSC. The academics thought their jobs were done. They moved on to submitting "Regulatory Management and Evaluation of Telephone Service Quality" to *Management Science*. In addition, a review paper on service quality was presented at the Telecommunications Policy Research Conference and the Southern Economic Association meetings. This study was directed to a mix of academics and technically-trained policy-makers. It reviewed the literature and described our work on the quality index consolidating the 38 characteristics. After being rejected by a theoretical journal, a revised version received favorable reviews at *Telecommunications Policy*, and it was published in early 1992 (Berg and Lynch, 1992).

## 5.2 Adapting to Reviews

While still waiting for initial academic reviews of the literature review spin-off, we obtained formal reactions from the FPSC on our "Final" Report. Gene Ferguson, a FPSC engineer, identified a number of deficiencies in the proposed Evaluation Weighting System:

I would suggest that when the PSC experts are again chosen, they select those of us who understand the rules, procedures, essentials of traffic switching, maintenance and maintenance terms, network structure and network switching, business office and repair facility operation, and the effects of any deficiencies on the subscriber, either directly or indirectly and to the rate-payer.

It seems clear the Bureau of Service Evaluation felt that it was not adequately represented in the initial survey! Furthermore, Ferguson was concerned with the wording of questions and the number of quality dimensions omitted from the initial survey. Lack of initial input from FPSC technical engineers clearly put off the staff. Yet we thought we were going through correct channels and had no inkling that key "stakeholders" (in this case, commission engineering staff) had not been utilized--either in developing the questions or taking the survey.

For example, the term "NR" reported in PSC evaluation reports are often interpreted by the company as "No Rule". However, the "NR" means that No specific percentage

Requirement was spelled out in the rule. Often 100 percent compliance is required (when the word "all" is implied). In other cases, the FPSC selects "realistic" targets. The initial study did not have these standards

Also, when staff applied the weighting scheme to a recent evaluation in which a company had failed to meet pass-fail standards (requirements) in 19 areas, the index score was "Above Average". While the calculation involved a misapplication of the methodology, it raised red flags regarding the implications of switching to a single index. In the case of "same day restoral" (omitted in the initial survey), the FPSC requires 80 percent. Since this is perceived as an important service standard, they wanted this item included. A number of other omitted items were identified. Ferguson wanted some weight given to these items.

A subsequent meeting with FPSC staff in Tallahassee cleared up a number of misunderstandings (on both sides). Follow-up telephone conversations between Buzas and Ferguson helped each understand the others' concerns. The result was a memo by Buzas and Lynch, "Issues in the Implementation of the Weighted Index of Telephone Service Quality," which addressed data collection, aggregation and evaluation. They made a number of points:

*1. Is the index a gross or net measure of relative benefits?*

Since service quality is conceptually distinct from cost of service, the focus here was on the benefit side. The trade-off against cost could be handled separately from the relative valuation of quality dimensions. The concern had been expressed that the adoption of a comprehensive index could lead to gold-plating. The key point was that for many dimensions, improvements above the standard yields incremental benefits which are greater than the associated incremental costs.<sup>2</sup> Furthermore, the incentives to provide excessive levels of quality were no different than at present. "Just meeting" every standard might be more costly than reaching another set of performance targets that yield an identical comprehensive score. If the weights are "correct," just meeting each standard is "gold-plating" in the sense that costs are too high for achieving this overall level of quality. Current FPSC mechanisms for evaluating network modernization programs and other prudence tests would apply to the cost side. At least with a single index, the task of the commission (and firms) would be simplified. The PSC can focus separately on evaluating the additional costs associated with achieving higher quality scores.

*2. Should a firm pass overall, when failing several pass-fail standards?*

Between 1985-88, companies failed to reach a standard 191 times. Of these, 162 (or 84.8%) were on the 21 (of 38) rules with a standard of 100%. On average, a company failed 13.6 rules of which 11.6 were on rules with standards of 100%. This suggests that the complexity of evaluating performance by 38 (or more) pass-fail standards is problematic

---

<sup>2</sup>Given the difficulties in sorting out potential cost complementarities, the quality dimensions might be considered in related bundles.

when degree of "substandard" performance is not captured in the summary index. The relative importance of the failures warrants attention.

The fact that a company can fail numerous rules yet receive a "passing" grade is an integral part of the proposed scheme. Setting the numerical passing score does not necessitate a "degradation in overall quality". Higher overall scores *could* be required (although such ratcheting up ought to be justified in terms of low incremental costs or high incremental benefits). For example, the lowest passing score achieved by any company might be established as the lower bound, or the average passing score could be taken to be the standard. A minimum acceptable performance rule would be established for each rule, while overall performance could be gauged on the basis of a higher level than if all such rules were met exactly.

*3. Do the weights reflect a narrow constituency?*

We became sensitized to the likelihood that the FPSC had multiple constituencies to protect. Tourists and non-subscribers rely heavily on Public Telephones. Residential subscribers are less dependence on Public Telephones, and so would weight less heavily those dimensions associated with Public Telephone performance. Similarly, businesses which are using digital transmissions would place a premium on other aspects of quality.

It would be possible to develop separate indices for each constituency. Multiple scores could be reported and evaluated, or these could be weighted to obtain an index of aggregate performance. We did not see such complications at presenting problems. At least the proposed methodology forces decision-makers at commissions and companies to acknowledge that different dimensions of quality performance have different weights.

*4. How does the exclusion of important rules from the weighting formula affect its usefulness?*

As it turned out, omitted rules tended to be those without set standards for performance. The initial FPSC liaison was asked about this issue prior to developing the initial survey. The liaison was later sent the questionnaire for prior approval. Unfortunately, that person was in the process of leaving the commission. Lesson: when trying to interact with regulatory agencies, be sure you have the "right" contact point. Academics might not fully appreciate the need to have an on-site advocate for the methodology (or conceptual framework) under consideration. The exclusion of important rules hurt the index's credibility.

*5. Did the right experts complete the survey?*

Since the initial liaison did not distribute the survey instrument to key technical personnel with knowledge of the standards, the commission weights were brought into question. However, the weights obtained from the telephone company panel of experts were statistically the same as obtained from the FPSC. It was clear, however, that a new survey had to be developed and given to additional FPSC staff as well as to cooperating telcos.

6. *How are the formulae to be calculated?*

As with any new methodology, a clear understanding of its components was essential for its successful application. In the initial FPSC application, incorrect weights were given to several rules. Also, there was a misunderstanding regarding calculation of the overall index. These points could be easily addressed by preparing clearer instructions, including background information and illustrative calculations.

7. *Could the quality index be rescaled?*

Instead of a score of 1 to 10, a 0 to 100 scale was requested. This was easily handled by a linear transformation, converting old scores to new scores.

In summary, the exchange of memos and meetings increased the research team's awareness of the administrative processes used to evaluate quality. We tried to address concerns, while directing attention to the advantages to have a comprehensive quality index.

The official FPSC response was positive. J. Alan Taylor, Chief Bureau of Service Evaluation, noted that he found

. . . the analysis to be excellent and insightful a to the problems facing regulators who base their quality of service decisions on simple pass/fail rule criteria. Certainly the regulated industry has long been apprehensive of a service quality measurement regime which focuses primarily on failures, without giving some consideration to more economical improvements in overall performance levels. PURC's weighting system therefore appears to be an appropriate way to assure that general quality of service levels remain high, particularly as we move into an era of regulatory flexibility through an incentive approach to governmental oversight.

The FPSC decided to fund a new survey and the development of new weights. Lynch and Buzas began the new study.

### **5.3 New Data and New Developments**

As the new data collection effort proceeded in early 1990, reviews on the comprehensive academic paper were received. Major revisions were required. The most important criticism was that we should have asked end users (consumers) what their tradeoffs were rather than asking expert regulators and managers to give their views of the tradeoff that would be in consumers' interests. We had described theoretical reasons to expect that customers of regulated monopolies would not have established tradeoffs--because they have no occasion to choose--but reviewers were unconvinced. We argued that only if they have expertise are consumers likely to have established tradeoffs.

By late fall 1990, Lynch and Buzas had a draft of the new Report, including a LOTUS 1-2-3 spreadsheet that allowed the FPSC to easily calculate a company's overall quality score using the weights uncovered by the project. The Executive Summary was for practitioners. Technical supporting material was included so that an independent expert could revise the weighted index.

In December 1990, FPSC initiated rule-making. The rule-making process turned out to be long and tortuous. Ultimately, Lynch testified on the methodology in mid-1991. However, a glitch appeared which further delayed things. The index got caught in regulatory cross-fire. In February 1991, the Public Counsel's Office filed complaints that a telco had falsified quality of service reports. The company was alleged to have falsified records on out-of-service repairs (a pass-fail quality standard). There are two related standards: 80% same day repairs; 95% within 24 hours. The latter is viewed by some as particularly tight. While there appeared to be no evidence that a corporate official said to falsify records, technicians may have felt some pressure to do so.

What actually happened is still under debate (and delaying a rate case). The key point is that the regulatory process can get choked through strategic maneuvering by key stakeholders. The Public Counsel's Office did not like the incentive plan that the FPSC had adopted for the telco to begin with. Their staff were unconvinced that rate-payers would benefit from the plan. The falsification report provided a wedge for attacking the plan--since quality of service was part of the incentive plan: rate-payers would be due refunds.

Thus, if the FPSC replaced the host of pass-fail standards with a single index, some of the steam would be taken out of the Public Counsel's case. The telco might argue "mitigating circumstances" since the old Rules had been jettisoned. With tens of millions of dollars at stake, there seemed no reason to replace the many standards with a single quality index. Ironically, had the weighted standard been in effect, the undue pressure might have been avoided.

In July 1992, rule-making was initiated; a meeting was scheduled for September 22 on Rule No. 25-4.080 (Weighted Measurement of Quality of Service):

The purpose of this new rule is to introduce another tool which the Commission can utilize in its effort to accurately measure the quality of service provided by local exchange telephone companies. . . . The rule authorizes the Commission to utilize a weighted index system when considering the adequacy of service . . . The system contains various quality of service measures currently contained in Commission rules and weights them according to their importance in the provision of acceptable quality local telephone service. . . . Companies shall be responsible for complying with each service standard, whether or not an overall score of seventy-five (75) is achieved when the weighted index is employed.



The proposed rule has yet to make it back to the Commissioners for a final vote. If accepted, the new rule will use the weighted index as an "additional" tool for evaluating quality. It will *not* replace the dozens of pass-fail standards.

#### 5.4 Revisions

At about the same time as rule-making was initiated, the reviews on our *Management Science* re-submission arrived. More comments required significant additional research. Again, reviewers were concerned that expert regulators rather than consumers filled out the survey instrument. Partly in response to these concerns, we obtained a sample from the largest purchaser of telephone services within the state government. Thus, these respondents were sophisticated and representative of consumers. We found close agreement between the weights of experts within the FPSC to those of these other experts.

Previously, we had been able to measure quality, but if quality improvements lead to higher costs to rate-payers, we were unable to say whether such improvements should be encouraged or discouraged. We had suggested that this would require regulators to marry our system for measuring the marginal benefits of quality with a detailed study of the marginal costs of improvement along various dimensions. Although the spirit of our weighted index was to get the regulators out of the business of micro-managing regulated companies, we understood that the cost side warranted attention.

The second addition to the initial study extended the work to address this issue. The revised survey instrument elicited tradeoffs between various dimensions of quality and the price of basic monthly service, allowing us to quantify the dollar value of a 1%-point improvement along the various dimensions of service quality. Thus, regulators could encourage a quality improvement that would be coupled with an \$X increase in price if the increase in consumer benefits associated with the improvement exceeds the disutility associated with the price increase. Interestingly enough, the FPSC experts gave a dollar value associated with quality increases that was *three times* that obtained from companies and the large demander.

We were pleased that the multi-year research project resulted in a proposed rule for adopting the weighted index system for the evaluation of service quality. However, the proposed rule utilizes the index as an added requirement rather than as a replacement indicator of pass-fail performance standards. Since the spirit of our recommendation was to move away from detailed consideration of the sixty-plus dimensions of quality, we hope that the FPSC ultimately adopts the comprehensive performance index. If the FPSC utilizes the crude, unweighted pass-fail mechanism as well as the comprehensive index during the transition period, that is their judgement call as to what is politically acceptable.

## 5.5 Customer Perceptions and Expert Trade-Offs

It is useful to underscore the role of experts in our framework. The second revision attempted to address the key conceptual issues raised by reviewers.<sup>3</sup> When expert regulators assess service quality in the interest of everyday consumers, the tacit assumption is that the experts' utility functions are highly correlated with those of everyday customers would have if they did not lack knowledge of how measurable attributes translate into realized benefits. Another justification for using experts is that research indicates that when tradeoffs among attributes do not already exist in respondents heads, the tradeoffs they construct on the spot are highly unstable. There is a substantial literature showing that experts make tradeoffs that are much less sensitive to distorting effects of measurement. The relevant issue is not whether the consumer has experience with the product. The critical concern is whether consumers have experience making tradeoffs among the particular dimensions relevant to the decision at hand. Experience means that these tradeoffs can be retrieved rather than constructed at the time of measurement (Feldman and Lynch, 1988; Fischhoff, 1992).

Our approach is supported by Fischhoff, who surveyed a large body of applied policy research which uses "contingent valuation" methodology. That research attempts to elicit citizens values in the hope that spending on public policy programs can reflect the priorities of consumers. The repeated finding is that when consumers' tradeoffs are elicited for goods that are not customarily traded in any marketplace, consumers do not have articulated values relevant to those decisions. The result is that measured tradeoffs have indefensible properties.

One could still argue that consumers should identify the criteria to be measured and that the role of regulators might be circumscribed to judging the quality of credence attributes--those characteristics which are rarely learned, even after consumption. In fact, this process is very close to what actually happens. Consumers call in their complaints to companies and regulators. Such data are reported. The complaints are invariably phrased in terms of "benefits." Legal constraints force regulators to reverse engineer those complaints and to trace them back to problems on attributes that underlie those benefits. The problem is that when laws and rules have been written pertaining to these "objective" attributes (eg. percent of inter-office calls completed), they become "credence" attributes for consumers. Typical consumers do not understand the links between technical attributes and benefits.

When comparing expert regulators and novice consumers, we expect the former to generate weights that would be strongly related to those elicited from a "trained" representative sample of consumers. Consistency across companies, the FPSC, and a large state agency which buys telecommunications services support our view that the weights provide a good first cut at prioritizing the dimensions of quality.

---

<sup>3</sup>I am indebted to John Lynch for formulating the points related to customer vs. expert perceptions.

## 6. Concluding Observations

What have we learned? The regulatory process is run by lawyers, with the aid of accountants and engineers. Economists have input into the process, but the deference is underwhelming. Perhaps, this is appropriate. After all, as a profession, we have developed precious little that helps decision-makers identify and reward quality.

Are there other lessons from this case study? Regulatory review and academic review have similar properties. Each has well established criteria, although the weights given each will differ dramatically. Academics put a premium on elegance (although simplicity sometimes wins the day). Certainly, regulators will emphasize simplicity over complexity. Both seek robustness of results. The conclusions need to stand up to possible changes in initial conditions. Academics place a premium on the new and innovative, while regulators emphasize continuity. After all, the assets are so long-lived that switching policies can wreck havoc with decision-making. But there is also an element of continuity that academics respect. The accepted paradigm will not be displaced easily, so the policy conclusions had better square with prior views with regards to the setting. It is "okay" for the results to be counterintuitive, so long as they are based on maximizing behavior by individual (generally, well-informed) agents!

Both processes are designed to kill bad or useless ideas. Thus, review lags are not only reasonable, but necessary if the contribution is to be evaluated carefully and thoughtfully. Rejections by an editor involve some randomness: the particular reviewer does not really understand the paper (alternatively, the points are not expressed in a logical and careful manner), the reviewer has an irrational grudge against a line of research (my favorite excuse), or the reviewer is on sabbatical and the disorganized editor lets you languish in purgatory for an unseemly amount of time. Whatever the reasons for the lag and rejection, the process is honored and is widely believed to improve the contributions to the scientific literature.

Good reviewers provide detailed feedback when the paper shows promise. Original versions of papers are seldom ready for prime time. A parallel process occurs in the adversarial regulatory setting. Various stakeholders will identify limitations to the proposed policy. Alternatively, one group might stonewall an idea if implementation would be injurious to their position.

So our tale still has no ending. The most comprehensive (and rigorous) expression of our ideas is still under review at a highly ranked journal. The feedback has been thorough and we have tried to be responsive to reviewer suggestions. Similarly, we still wait for Commission passage of a rule that just adds the proposed index to the regulatory tool kit for evaluating performance. In both cases, the lags seem long--but they are also understandable, given the stakes in both instances.

### Bibliography

- Berg, Sanford V. and John G. Lynch, Jr. (1992), "The Measurement and Encouragement of Telephone Service Quality," *Telecommunications Policy*, (April), 210-224.
- Buzas, Thomas E. and John G. Lynch, Jr. (1988), "A Formula for the Comprehensive Evaluation of Local Telephone Companies: Report to the Florida Public Service Commission: Preliminary Report," (January).
- Buzas, Thomas E., John G. Lynch, Jr., and Sanford V. Berg (1991), "Issues in the Measurement of Telephone Service Quality," in *After the Breakup: Assessing the New Post AT&T Divestiture Era*, ed. by Barry Cole, Columbia University Press, 268-276.
- Buzas, Thomas E., John G. Lynch, Jr., and Sanford V. Berg (1992), "Regulatory Measurement and Evaluation of Telephone Service Quality," revised and resubmitted to *Management Science*, 1-52.
- Feldman, J. M. and J. G. Lynch, Jr. (1988), "Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior," *Journal of Applied Psychology*, 73, 421-435.
- Fischhoff, B. (1992), "Value Elicitation: Is There Anything in There?" in M. Hechter, L. Cooper, and L. Nadel (eds.), *Values*. Stanford, CA: Stanford University Press.
- Louviere, Jordan (1984), "Hierarchical Information Integration: A New Method for the Design and the Analysis of Complex Multiattribute Judgement Problems," in *Advances in Consumer Research*, Vol. 11, ed. Thomas Kinnear, Provo, UT.: Association for Consumer Research, 148-155.
- Noam, Eli M. (1991), "The Quality of Regulation in Regulating Quality: A Proposal for an Integrated Incentive Approach to Telephone Service Performance," *Price Caps and Incentive Regulation in Telecommunications*, M. A. Einhorn, ed. Boston: Kluwer Academic Publishers, 167-189.

# A New Index of Telephone Service Quality: Academic and Regulatory Review

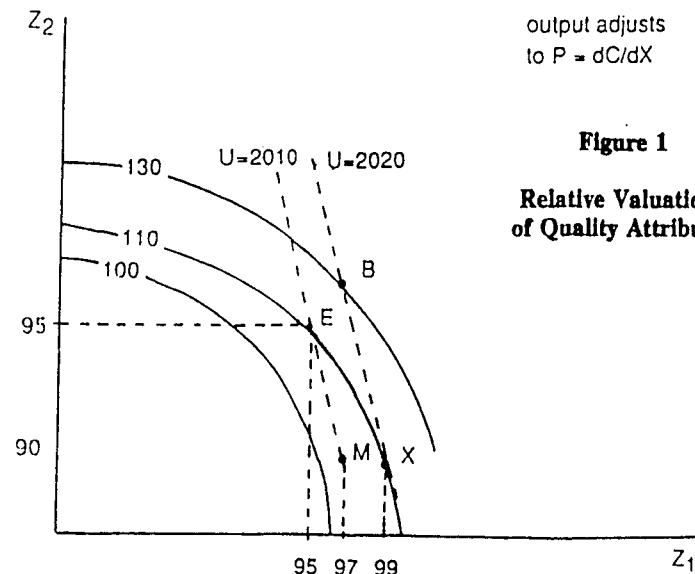
by Sanford V. Berg

## Abstract

Historically, state regulatory commissions have used large batteries of pass-fail standards to evaluate the quality of local telephone service. This paper describes the development of an alternative approach to measuring, evaluating, and rewarding service quality. The ideas presented here have evolved through two different processes--academic research and regulatory implementation. These processes have some similarities and differences which complicated our objective of developing an index which was both analytically sound and policy relevant.

To be presented at the  
Columbia Institute for Tele-Information  
Conference on  
Quality Choices in Network Industries  
April 23, 1993

1. Introduction
2. Overview of Current Regulatory Schemes and the Proposed Index
3. Scientific Review and the Creation of Knowledge
4. Regulatory Review and Procedural Fairness
5. Case Study of a New Quality Index
6. Concluding Observations



## 2.2 Relative Valuations of Service Characteristics

Substandard vs. Superstandard

(I)  $Z_1 > Z_1^*$ ,  $Z_2 < Z_2^*$  eg. point X

(II)  $Z_1 = Z_1^*$ ,  $Z_2 = Z_2^*$  eg. point E

$$Q_* = W_0 + W_1(Z_1 - Z_1^*) + W_2(Z_2 - Z_2^*)$$

$$Q_* - (W_0 - W_1Z_1^* - W_2Z_2^*) = W_1Z_1 + W_2Z_2$$

$$Q = W_1Z_1 + W_2Z_2$$

Acceptable Quality

$$Q = Z_2 + (5/2)Z_1$$

$$Q = 332.5$$

Same Quality from (97, 90) (point M)

Higher Quality if  $Q = 337.5$  (point X)

Issue of Incremental Costs

**Figure 2  
Regulatory and Academic Reviews**

1986

FPSC/PURC Discussions

1987

Methodology Developed

FPSC Data Collected

1988

Draft Report

Telco Data Analysis

Draft Article

1989

PURC Seminar  
"Final" Report

FPSC/PURC Meeting

Revised

Submitted/Conferences

1990

FPSC Hearings on Service  
Interruptions

Data Collection

Rule Making

Revised Report

Reviews Received

1991

Public Counsel Files

PURC Testimony

Recommendation Deferred

Revised

Resubmitted

1992

FPSC Staff Accepts Index

Reviews Received

Revised

1993

Rule Making

Resubmitted

1994

