HOW THE QUALITY OF RESPONSES THE IRS PROVIDES TO TAXPAYER INQUIRIES IS MEASURED

Kevin Cecco and Rachael Hoopengardner, Internal Revenue Service
Presented at the 2002 American Statistical Association

The Internal Revenue Service (IRS) has developed a process by which quality among services the IRS provides to taxpayer is measured. With the assistance of the Statistics of Income (SOI) Division of the IRS, this process has been centralized, implemented, and improved on to consistently/adequately measure the quality of service provided to taxpayers. The results of this quality review process are captured on a data base that can be accessed by all persons associated with measuring quality. This paper will describe the historical development and standardization of the quality measuring process, as well as the future of quality measurement within the IRS.

IRS currently measures quality for at least fifteen different types of telephone customer service product lines. IRS has been measuring the quality of service it provides to taxpayer inquiries since the late 1980's. However, within the past several years, it has made major advances in the quality measurement process.

The initial quality review system relied on "mock" questions to determine the quality of service taxpayers received. This was done through a group of reviewers making telephone calls to the IRS with made-up questions. Reviewers would then rate the quality or accuracy of the information provided to them by Customer Service Representatives (CSR's) during the call. IRS found that the initial review system had flaws. One problem was that quality reviewers were reviewing responses to fictitious questions. Another was that these test calls were planted, meaning that review was not taken as a sample of the universe. Therefore, measuring the quality rate and associated variance with this estimate became difficult.

Another method IRS used for determining the quality of service provided to taxpayers was to have a quality reviewer sit next to a CSR and listen into both the taxpayer calling for help and the CSR and rate the accuracy of the service the taxpayer received. This method posed problems as well. The CSR always knew which call was monitored and could therefore modify behavior/responses based on this knowledge. It was impossible to determine the actual behavior/responses of a CSR on a taxpayer call that was not monitored.

In the early 1990's, IRS monitored nine different programs for quality. There were two main types of programs: telephone calls and closed paper cases. Each of the nine programs had different reviewing processes and procedures; and many had a data base for compiling data used solely for that program. Each of the nine programs had a set of attributes that the organization used to define quality. Some attributes were common to all programs; however, no program had a set of attributes identical to another program. Quality reviews were performed within each site.

Having nine quality programs, each with a reviewing procedure and a data base to collect reviewing information also posed problems for IRS in measuring quality. Since the review was done within each site, there was a perceived bias in the resulting quality review from a site. This made it difficult to make site-to-site comparisons for a particular product. Also, since review was done locally in each site, procedures and processes could be interpreted differently. It was difficult to determine whether differences existed due to the "bias" from an in-site review or if the quality was actually different between sites.

Another problem was that, since each product had a different set of attributes on which a case was reviewed for quality, it was difficult to make comparisons of quality between products. These problems made it difficult to attain a national level accuracy/quality rate when combining sites or products. These difficulties brought about the notion of standardizing the quality review process. The two main goals of standardization were to have both closed paper cases and telephone cases reviewed in a single site, and to change the telephone review to a remote monitoring system.

How IRS Measures Quality

In order to achieve the standardization of the quality review process, IRS wanted to create a Centralized Quality Review System (CQRS). The CQRS would standardize the quality review process by centralizing the review process to a single location. In addition, IRS hoped to streamline the process by making a single Data Collection Instrument (DCI) by which all cases were measured for accuracy. IRS wanted to develop one universal data base that would contain data for the nine different quality systems. A single data base would facilitate feedback to the sites, allowing the sites to improve on deficiencies in providing quality service. This single data base, along with centralization, would enable IRS to generate consistent, unbiased accuracy rates at the national, regional, or site level.

The centralization and standardization effort was divided into two phases: phase one, the centralizing of the telephone review process; phase two, the centralization of the closed paper cases.

In 1996, IRS formed a team tasked to review each of the existing nine quality systems and the quality measurement system associated with each program. The team reviewed the attributes within each program and eliminated portions of the review process deemed unnecessary. The objective was to create a single DCI containing attributes applicable to all products.

In order for an attribute to exist in the new DCI, the attribute needed to meet two qualifications. First, an attribute needed to measure results and not the procedural portion of a telephone call unless the procedural attribute pertained directly to the resolution of the taxpayer inquiry. The second qualification was that the attribute needed to have potential for providing meaningful feedback in improving the accuracy rate at a site. Attributes that did not meet these two qualifications were eliminated from the review.

Attributes that remained in the review process were then separated into two categories--header and quality; and, within quality there are three subsections--disclosure, customer relations, and inquiry resolution. Header section categories include such items as the "case type" and "length of call." Items within the quality section consist of such attributes as "actively listening" and "covered disclosure." All attributes within the quality section are considered "fatal flaws." A fatal flaw means that, if one attribute within this quality section is marked as "inaccurate," the entire case is marked as "inaccurate."

Centralized Quality Review System

The Centralized Quality Review System (CQRS) is located in Philadelphia, PA. This is referred to as the CQRS site. Within the CQRS site, there are several managers whose responsibility is to oversee the telephone review process and ensure that proper training and implementation take place. There are also many quality reviewers at the CQRS site, each of whom is uniformly trained to review a single product (some are trained on multiple products based on knowledge and experience). Some reviewers receive specialized training as "experts" on specific product topics (e.g., complex tax law questions).

This telephone review is done through remote monitoring. IRS wanted to replace test calls and side-by-side monitoring with live, remote monitoring from a centralized site where deficiencies would be recorded on a single, universal data base. Remote monitoring is where a quality reviewer can tap into a telephone line and listen to a live telephone call from a taxpayer and review the call for accuracy. Neither the taxpayer nor the CSR is specifically notified of the review. However, both the taxpayer and CSR know that any call is subject to quality review. The following products are reviewed at the CQRS site: Tax Law, Accounts, Practitioner Priority Service, Automated Collection System, Spanish, Rmail, and Email.

The Statistics of Income (SOI) Division of IRS develops the sampling plans that the CQRS site uses to sample incoming telephone calls for each of the products. SOI provides the CQRS site a sampling plan biannually, filing season and non-filing season, that is designed to allow each product to achieve the desired level of confidence and precision of the sample estimates. The plan is developed biannually due to the large discrepancy in volume and types of inquiries received from taxpayers on these product lines during the two times of the year. The sampling plan details which product line, which site, and which application each reviewer is to monitor, covering all hours of operation of these product lines. This ensures that samples will be both valid and random.

Since centralizing the quality review process, IRS has encountered problems with the sampling of taxpayer telephone calls. One problem was "dead air." When a reviewer, scheduled to monitor a specific site on a specific application at a certain time of day, tapped into that site to monitor the next incoming call, there was, at times, no traffic or volume coming through that site. The reviewer had to wait on that application for a call to review. This was not an effective use of our resources. Nevertheless, the quality reviewer could not abandon the site without knowing the reason for no traffic. One reason for "dead air" was site closure. A site would be closed due to "x" or "y," and a reviewer was scheduled to monitor that site and the reviewer had no way of knowing that the site was closed.

IRS overcame this problem through the installation of Custom View, one of its newest technological advances in quality review. Custom View allows reviewers to pull up site information and see in "real time" the traffic coming into that site. Within a site, the traffic/volume is broken down by application. Now, when a reviewer is scheduled to monitor a site and is only finding "dead air," that reviewer can determine whether or not it is due to site closure. It also enables a reviewer to move to applications within the same site with heavier volume. This decreases the amount of "dead air" time and increases the productivity of the reviewers.

Another problem IRS has with quality review monitoring is the month of December. The volume of incoming taxpayer inquiries drastically decreases during December due to holidays, making it difficult, if not impossible, to achieve the desired level of confidence and precision for that month. Further intensifying the problem is the large amount of requested leave from the quality reviewers during the month, also due to holidays. IRS is still working to determine how best to correct this problem.

Quality Review Data Base

The CQRS site needed a data base on which to store all of the information collected from the quality review. With the assistance of SOI, the Quality Review Data Base (QRDb) was developed as the universal data base for all CQRS review. The QRDb is a user-friendly database that resides on the Internet and allows IRS users to access established reports or to generate user-defined reports. This system produces both weighted and unweighted reports of accuracy, as well as the precision margin of any weighted accuracy rate. The data base generates site-level, regional-level, and national-level reports. It also generates these reports by various time periods, such as month, planning period (which is a combination of several months similar to a quarter), and fiscal year. Each user is given a specified level of accessibility to these reports. Site-level managers are only allowed to access reports for the sites they manage. Regional-level managers are allowed to access reports of the entire region, as well as site-level reports of any site within their regions. And national-level managers are allowed access to all reports on the QRDb. The QRDb allows the CQRS to produce consistent, unbiased national accuracy rates for any and all products within the CQRS.

Future of Quality at IRS

IRS has several goals for the future of quality measurement. The CQRS would like to implement phase two of the centralization, which would entail bringing all closed paper review to the CQRS. Samples would be pulled at each individual site, but the CQRS site would receive these sampled cases to review. This would ensure that closed paper cases had a consistent, unbiased review.

Another goal of IRS is to institute call recording based on the SOI sampling plan in which calls would be selected and recorded at the CQRS site and reviewed at a later time. This would help IRS ensure that every incoming call has equal opportunity for being selected into the sample. Currently, quality reviewers have to work schedules that mimic the hours of operation at call sites. The difficulty with that is that the CQRS site is located in Philadelphia, and, therefore, quality reviewers are also located in Philadelphia. Covering late hours of site operation for sites on the West Coast is difficult. Also, Saturday hours are difficult to adequately sample, given the work schedules of quality reviewers. With call recording, quality reviewers could work normal hours, and the call recording system would record live calls across the operating hours. Call recording would also correct the problem of undersampling in December, mentioned previously.

Another improvement with the quality review process is that, in the future, there will be a new method of measuring quality. As stated previously, all of the quality attributes in the DCI are fatal flaws. So, if a CSR spends 45 minutes on a telephone call assisting a taxpayer on an inquiry and at the end of the call makes a minor mistake, the entire call is marked “inaccurate.” Currently, a call can only be marked as an “accurate” or “inaccurate” telephone call. It is believed that some of these "fatal flaws" are not critical to the taxpayer receiving the correct answer to the inquiry. Therefore, IRS is modifying the way in which the final calculation of the quality rate of a telephone call is made. In the future, only the attributes critical to the taxpayer receiving accurate service will be considered “fatal flaws.” A case will be divided into five subcategories on which a case is reviewed. Those subcategories are procedural, regulatory, timeliness, professionalism, and customer impact. Only attributes within the customer impact category will be considered fatal flaws. This will change the way in which a case is rated for accuracy. A case could have a quality rate ranging from 0 to 100 percent and not just a 0 or a 100 percent. Executives within the IRS believe this new method will more accurately reflect the quality of service taxpayers are actually receiving.

Finally, another change is the continuing growth of the CQRS site. As new products are added to the centralized review process, more reviewers will be needed to cover all of the products monitored. And the CQRS will be continually modified and improved on as the IRS determines new and better ways to serve taxpayers and also better methods of measuring that service.