The Evolution of IRS Telephone Quality Measures

Laura Rosage, Statistics of Income, Internal Revenue Service
P.O. Box 2608, Washington D.C. 20013

Key Words: Quality, Defects per opportunity, Accuracy

The Internal Revenue Service (IRS), tasked with collecting taxes from this country's citizens, deals with more Americans than any other public institution. Unfortunately, over the years, the tax law has increased in complexity and the myriad of forms has become confusing. In an effort to assist taxpayers comply with the law, the IRS established a toll-free telephone service with Customer Service Representatives (CSRs) ready to help taxpayers with their tax related questions. Since 1965, the IRS has offered this free telephone assistance to millions of taxpayers. IRS assisters handled over 55 million telephone calls in fiscal year 2003 (October 2002 through September 2003) and 15.8 million calls during filing season 2004 (January through mid-April 2004) alone. With this large volume of inquiries handled by telephone assisters, the accuracy of the information provided has a potentially large impact and is of interest to stakeholders both within and outside of IRS. The way the accuracy of telephone assistance is measured has evolved over the years from test calls, to live monitoring of telephone calls, and soon, to contact recording. This paper details the evolution of how the IRS monitors calls and the latest move from a pass/fail method of measurement to a defects-per-opportunity methodology. It discusses the strengths and limitations of each method, the overall impact on quality rates, and future plans for improvement to the measurement process.

Monitoring Taxpayer Calls for Quality Purposes

One of the IRS's major goals is to make its telephone operations a world-class customer service organization. To do this, it is necessary to track the accuracy, efficiency, and quality of the service provided by those answering the telephones. However, when first implemented, the focus of telephone service review was only to evaluate employee performance. Managers reviewed the work of CSRs and used the data gathered as feedback for employees' performance appraisals and to identify training needs. It wasn't until the late 1980s that IRS attempted to measure the overall quality of the service provided to taxpayers.

Managerial Review

IRS's first iteration of quality measurement consisted of managers reviewing their employees at each call site. The manager would sit with an individual CSR and listen in on selected telephone calls. When the call was complete, the manager was able to provide immediate feedback to the CSR on any errors made or on issues of timeliness or professionalism. While better than no review at all, there were some drawbacks to this system. First, the data gathered during telephone call monitoring were really intended for CSR performance review rather than a site or national measurement of quality. Also, because managers monitored their own employees, complete impartiality of the reviews could not be guaranteed. Additionally, since managers sat with the CSRs, the assisters were aware they were being monitored. They could alter their behavior during calls that were reviewed: Responding in a more professional manner, researching the tax issue more thoroughly, or adhering to their manual guidelines more fully. These changes could affect the quality of the call, giving IRS a skewed view of the performance of the employee, as well as the overall quality of the service provided to taxpayers. Finally, because review of the CSRs was performed at each of the call sites, there were issues with the lack of consistency of reviews from manager-to-manager and site-to-site.

Attempting to get a clearer picture of the actual service taxpayers received, IRS implemented technology that allowed managers to review CSRs remotely. Without alerting the CSR, they could listen in on and review telephone calls from the privacy of their office, rather than sitting beside the assister as he or she was on the phone. This transparency eliminated the problem of the CSRs knowing they were being monitored and modifying their behavior accordingly, but the issues of inconsistency of reviews from site-to-site still existed. Additionally, the accuracy results from this process were generally very high and at odds with the Government Accounting Office's (GAO) assessment of quality, further supporting the concern that the monitoring carried out at a local level was not impartial.

Integrated Test Call Survey System

In an effort to eliminate these concerns, the Service implemented a new program in addition to the managerial review, the Integrated Test Call Survey

System (ITCSS). This system, established in 1988, was designed to produce a national estimate of quality rather than relying on managerial review of employees to establish the measurement and to provide timely feedback to call sites. The sites could then use the feedback to target specific areas for improvement, then assess the success of their efforts. Under ITCSS, a centralized group of quality reviewers called into the toll-free IRS tax law assistance line, posed mock questions to CSRs, and rated the quality of the responses given. The creation of this centralized review process, where independent reviewers received identical training and held regular meetings on how to rate calls consistently, reduced the inconsistency and impartiality of rating the quality of service provided to taxpayers at the local level. Of course, this method of measurement also introduced other issues. Although the universe of test calls was modeled closely after the volume and topic of taxpayer inquiries, this national sample was not a sample of the universe of actual taxpayer calls, but a review of responses to fabricated questions, posed by persons other than real taxpayers. Any data gathered from this test was an artificial measurement of the accuracy of information IRS assisters provided to the public. Additionally, ITCSS measured tax law calls only. For most other types of calls coming in on the IRS toll-free lines, it is necessary to know the identity of the caller and access their tax records to completely and accurately respond to their inquiry. This would not be possible with test callers. Also, after a time, even though the test questions were changed periodically, the CSRs were often able to identify calls from quality reviewers.

Centralized Quality Review Site

Because of the limitations of ITCSS, the IRS eventually moved away from the test call system and created the Centralized Quality Review Site (CQRS) in Philadelphia. This site was established to centralize the IRS telephone review process into one location; to sample real, live calls from the universe of actual taxpayer inquiries; and to establish an estimate of the true level of service being provided to taxpayers. They were also charged with standardizing the review process of telephone calls and centralizing IRS telephone quality review data into one database.

The site, established in 1997, initially began with eight reviewers measuring the quality of tax law calls only. Over time, further types of calls were added. The CQRS now has over 50 full-time reviewers who monitor tax law calls, taxpayer account-related calls, collections calls, calls from the tax-practitioner priority line, calls from U.S. taxpayers overseas, calls from employers seeking business taxpayer identification numbers, and all Spanish-language calls, as well as requests for IRS tax forms.

The CQRS was able to establish an impartial measurement of quality for each call site and type of call by utilizing technology that enabled them to remotely monitor live taxpayer telephone calls coming into any IRS call site across the country. They reduced inconsistencies in the review process through holding regular consistency training, as well as utilizing a standard data collection instrument that gathered the same data elements for all calls and stored the information in a central database. And because they were monitoring real, live calls, the quality measurement generated from the review data produced the clearest picture of the level of service provided to taxpayers since IRS implemented quality review.

This standardization of the review process and improvement in the consistency of reviews was a major step toward accurately measuring the quality of service the IRS provided to callers. Also, with remote monitoring, neither the caller nor the CSR was aware if their particular call was selected for review. This transparency meant that the monitored response was real, typical of the type of taxpayer/CSR interaction, and not altered in any way. Despite these advances, many new issues were introduced as IRS moved to monitoring live taxpayer telephone calls. Because the telephone calls were live, reviewers were required to monitor the calls real-time. While this sounds innocent enough, real-time monitoring had a tremendous impact on reviewer resources.

Monitoring taxpayer telephone calls for quality in real-time consumes a considerable amount of resources. Initially, the largest depletion was due to dead air. A reviewer would dial into a site to monitor calls. If there weren't any available calls at that site (no taxpayers calling in or an unscheduled site closing), the reviewer wouldn't know until listening to several minutes of silence or dead air. With the acquisition of software called Custom View, which allows the reviewers to see call traffic in the sites (real-time, less a 6 second delay), this problem was virtually eliminated. However, that was not the only problem with real-time monitoring.

To select a call for review, the quality reviewer at CQRS dials into a designated telephone number for a given site and type of call and is then attached to the next incoming call. The reviewer stays with the call, as long as it is in that particular site, able to hear the

complete CSR/taxpayer interaction. Because it is a live telephone call, the reviewer experiences what the taxpayer experiences, including time on hold or waiting for the next available assister. Any hold or wait time is wasted time for a reviewer and can dramatically impact the number of telephone calls that he or she can monitor. Unfortunately, there is no way to eliminate these phenomena when monitoring live telephone calls.

Additionally, in order to sample from the entire universe of calls when monitoring in real-time, the CQRS must have reviewers scheduled during all times of day that the IRS call centers are open. Unfortunately, there are not enough review resources to cover all hours of operation, which are 7 a.m. - 2 a.m. Eastern Time, meaning some calls are not subject to quality review. The CQRS does have staff monitoring phone lines from 7 a.m.- 12 a.m., so only those calls received during the very early morning hours of 12 a.m.- 2 a.m., less than 3 percent of the total universe of taxpayer calls, are not subjected to sampling for quality review.

Another minor issue associated with real-time monitoring is that the only record of the content of the call is the reviewer's notes. If the reviewer is unable to catch something that is said during the call, it can never be re-heard or recaptured. A reviewer's determination of the accuracy of the call is dependent upon what he or she is able to hear and jot down during the call. This can become an issue if a call site objects to the reviewer's evaluation of a call. Formal rebuttals from sites, requesting a re-evaluation of monitored calls, are frequently sent to CQRS for response. Unfortunately, since the disputed calls cannot be replayed, reviewer notes are the only evidence of what occurred during the call, leaving some room for continued disagreement.

Independent of the problems associated with real-time monitoring, is the issue of call transfers. Using the current communications technology available at IRS, reviewers are unable to follow a call if it is transferred outside of the original site receiving the call. If a customer service representative receives a call that he or she is unable to answer, they must transfer that call to another assister. If that call is then routed to another call site, in the current telecommunications environment, the CQRS reviewer is unable to follow the selected call. Therefore, the reviewer cannot determine if the taxpayer received the correct answer to their inquiry. This situation is becoming increasingly more common given the current operational push for call site specialization, where assisters at a given call site are trained to answer only specific types of calls. Whether the taxpayer selected the wrong option from the automated menu or because the initial CSR who screened the call misunderstood the taxpayer's question, calls that are misrouted would have to be transferred to another site rather than to another assister within the same site. This increase in call transfers would result in an increased number of calls selected for review that the CQRS reviewers would not be able to follow to completion.

Recording Taxpayer Calls for Quality Purposes

Until recently, the recording of taxpayer telephone calls, while legal if the act of recording is disclosed to callers, was not permitted based upon guidance from IRS Counsel. Call recording was viewed as an invasion of taxpayer privacy. However, since call recording has become standard in the customer service arena, IRS has revisited the issue and approved call recording for quality purposes only. With the aid of a vendor, IRS is now testing and piloting call recording in select call sites. Call recording is scheduled for complete installation and implementation in all IRS call sites by FY 2006. Telecommunications technology being implemented will enable IRS to record 100 percent of all incoming calls, then systematically select calls for quality review. The selected recordings would then be reviewed by CQRS reviewers and entered into the standardized database.

Because review will still occur at CQRS, all the advantages of this consistent third-party review remain. However, call recording also brings a number of additional benefits. Primarily, call recording eliminates many of the drawbacks of the real-time monitoring of telephone calls. Once reviewers are able to listen to a recording of the taxpayer/CSR interaction, there will be no more listening to dead air and no waiting on hold. They will be able to fast-forward through any wait time while the CSR is researching the taxpayer's issue. Reviewers will also be able to rewind the recording and re-listen to portions of the call, or the entire call if necessary, to more accurately assess what occurred during the call. Additionally, if the topic of the call is beyond the scope of a reviewer's training, he or she will be able to flag the call for evaluation by another reviewer with more technical expertise. Reviewers will also be able to listen to a recording at any time of day, regardless of when the call was placed, eliminating the need and added cost of an evening shift. Furthermore, the late night calls occurring

between the hours of 12 a.m. and 2 a.m., not previously subject to review, will be available for quality review sampling under call recording. All of these advances secured through the implementation of call recording allow for cost savings, resource savings, and improvements to the quality sampling and review process.

An additional and unexpected resource savings is the reduction in the number of rebuttals from sites who feel the evaluation of a call by the CQRS was incorrect. During the pilot process, those sites with call recording have been able to listen to any call received at their site, including those CQRS reviewers may have evaluated for quality purposes. Because call site managers are able to access and listen to the real CSR/taxpayer interaction rather than rely on reviewers' notes, they can immediately eliminate rebuttals for calls they believe were evaluated correctly without any CQRS involvement. Now, only those calls where there is a legitimate disagreement in the call evaluation, are forwarded to CQRS for further action, resulting in an overall decrease in the amount of resources spent on rebuttals and re-evaluations.

Another advantage of call recording is that recorded calls, once "sanitized" to remove any taxpayer-identifying information, may be used for training purposes. Recordings of real taxpayer/CSR interactions will allow IRS to train CSRs how best to respond to taxpayer issues. By listening to the same call, assisters can be taught to respond to many different situations in a consistent way. In the same manner, the recordings can also be used by CQRS managers to train reviewers to consistently evaluate the quality of the service provided to taxpayers calling the IRS.

Call recording also provides a major advancement in the way IRS is able to measure quality. Because reviewers are currently unable to follow a call if it is transferred from one site to another, it is impossible to capture everything that individual taxpayer experienced from the start of the call to the very end of the call. With the implementation of call recording at all IRS call sites, all incoming toll-free telephone calls will be recorded at each site. Therefore, if a call is transferred from one site to another, the portion of the call after being transferred will be captured and recorded at the destination site. Because all calls will be tagged with a unique identifier as they enter the IRS, multiple segments of a single call can be combined after-the-fact. Thus, for the first time since quality review began, IRS will be able to capture the complete taxpayer experience for any call in the universe.

Because call recording is still in the pilot phase and the necessary hardware and software has not yet been installed in all sites, it is difficult to identify any disadvantages or problems with using this methodology to capture and review calls for quality measures. Certainly, initial start-up costs are extremely high, but the resource savings and added benefits of call recording should eventually outweigh those one-time costs.

IRS Toll-Free Telephone Assistance Quality Measures

Once IRS monitors or records a call, how is the quality of that call measured? The methodology behind the measurement of quality has also evolved over the years.

Past Methodology of Telephone Quality Measurement

Before FY 2004, there was a single measure for the quality of telephone calls coming into the IRS toll-free telephone assistance service. This measure was reported internally to IRS executives, and externally to Congress, GAO, and the Office of Management and Budget (OMB). Quality for a call was measured as pass or fail, where if one element, or "attribute," of a call was incorrect, the entire call was counted as incorrect. An attribute is any individual element of the call that is rated for accuracy by IRS reviewers. Depending upon the taxpayer's issue, some attributes of a call may be: Did the assister greet the taxpayer courteously and professionally? Did the assister verify the taxpayer's social security number? Did the assister give the taxpayer the correct answer to their question? Did the assister provide their identification number? Certain attributes, while required elements during a call, may not affect the correctness of the answer provided to the taxpayer. Also, not all attributes are applicable to every type of call, so one particular call may have very few attributes, while another may have several. Regardless of the number of attributes for any given call, a single call could only have the possibility of scoring 0 percent or 100 percent. Under the pass/fail methodology previously, if an assister answered the taxpayer's question correctly, but forgot to provide his or her identification number at the start of the call, as required under IRS procedures, the call was scored as 0 percent for quality measurement purposes. While this practice encouraged attention to details on the

part of the telephone assisters, it presented an unclear measure of the quality of service provided to taxpayers, especially to the external users of the data.

Current Methodology of Telephone Quality Measurement

In an attempt to construct a more accurate picture of the quality of the service provided to taxpayers, the pass/fail methodology was retired and a new measurement system was implemented for FY 2004. This new method of measurement, defects-per-opportunity, was designed to distinguish between wrong answers and procedural errors that do not affect the accuracy of the answer provided to the taxpayer. IRS's single measurement for quality was separated into five individual measures:

- Customer Accuracy — Did the assister give the taxpayer the right answer?
- Regulator Accuracy — Did the assister follow all IRS regulations according to the tax code?
- Procedural Accuracy — Did the assister follow all internal IRS procedures for this type of call?
- Timeliness — Did the assister respond to the taxpayer in a timely manner?
- Professionalism — Did the assister respond to the taxpayer in a courteous and professional manner?

Given the nature of the measures, Customer Accuracy, Timeliness, and Professionalism are reported externally; whereas, Regulatory and Procedural Accuracy are measures intended for IRS use only. The five measures are each calculated as a percentage: the number of correct attributes divided by the total number of applicable attributes. Because Customer Accuracy has only one applicable attribute for any type of call — Did the taxpayer receive the correct answer? — a single call still only has the possibility of scoring 0 percent or 100 percent. However, with the elimination of all other non-applicable attributes, this measure of accuracy is now a very clear representation of the quality of the service provided to taxpayers. Each of the other four measures generally has multiple applicable attributes for each call, thus a single call can now score 0 percent, 100 percent, or anywhere in between.

Using FY 2003 data, Customer Accuracy was calculated using both methods:

FY 2003 data	Pass/Fail	Defects-per-Opportunity
Tax Law	80.10% (+/- 0.66%)	81.97% (+/- 0.63%)
Accounts	68.43% (+/- 0.44%)	88.15% (+/- 0.30%)

There was little effect on Tax Law calls, but the difference in the accuracy of Account calls is significant. This is due to the fact that for Account calls, telephone assisters are generally required to perform many internal procedures where an error may occur that does not affect the accuracy of the answer provided to the taxpayer. Under the old pass/fail methodology, this would have caused the entire call to be counted as incorrect. Now, attributes relating to internal procedures are included in the measures of Procedural Accuracy and Regulatory Accuracy and are no longer included in Customer Accuracy, providing IRS executives, as well as Congress, GAO, and OMB, a clearer picture of the quality of service provided through the toll-free telephone assistance service — that assisters actually gave callers a correct answer approximately 88 percent of the time rather than the 68 percent previously reported.

Future of IRS Telephone Quality Measures

Over the years, the way IRS monitors telephone calls and measures quality has undergone continuous improvement. With the implementation of call recording, the Service will have taken the next step in the process. Beyond call recording, there are plans in motion to combine national quality review performed by CQRS and local managerial review into one standard database. This will be the last step in the standardization of the review process. Once completed, managers and quality reviewers will be reviewing calls using the same attributes and standards and all quality data will be stored in a single place. This will provide individual call sites with additional data for error and trend analysis, allowing them to identify specific areas where additional training might improve quality.

In another move to improve the quality process, individual attributes from reviewed calls will soon be directly linked to telephone assisters' critical job elements (CJEs). CJEs are the specific items managers use to rate the performance of their employees. For example, call attributes regarding courtesy and professionalism will be linked with the professionalism CJEs for telephone assisters. The

attribute for whether or not the taxpayer received the correct answer will be linked with the technical knowledge CJEs for assisters. Through this linkage, managers will be able to use their reviews to quantify the performance of their employees rather than relying solely on qualitative data and subjective judgment.

Conclusion

A significant goal of the IRS is to make its telephone operations a world-class customer service organization. By improving the way the level of service provided to taxpayers is measured, IRS can not only better determine how closely they have come to achieving that goal, but can also identify areas for further improvement. With this continuous cycle measurement and improvement, we hope this goal of providing world class customer service to taxpayers will soon become reality.