United States General Accounting Office

GAO

Report to the Chairman, Subcommittee on Oversight, Committee on Ways and Means, House of Representatives

November 2002

TAX ADMINISTRATION

IRS Needs to Further Refine Its Tax Filing Season Performance Measures

G A O

Accountability * Integrity * Reliability

GAO-03-143

GAO
Accountability★Integrity★Reliability

Highlights

TAX ADMINISTRATION
IRS Needs to Further Refine Its Tax Filing Season Performance Measures

Highlights of GAO- 03-143, a report to the Subcommittee on Oversight, Committee on Ways and Means, House of Representatives

Why GAO Did This Study

The tax filing season, roughly January 1 through April 15, is when most taxpayers file their returns, receive refunds, and call or visit IRS offices or the IRS Web site with questions. To provide better information about the quality of filing season services, IRS is revamping its suite of filing season performance measures. Because the new measures are part of a strategy to improve service and because filing season service affects so many taxpayers, GAO was asked to assess whether the new measures have the four characteristics of successful performance measures graphically depicted below.

What GAO Recommends

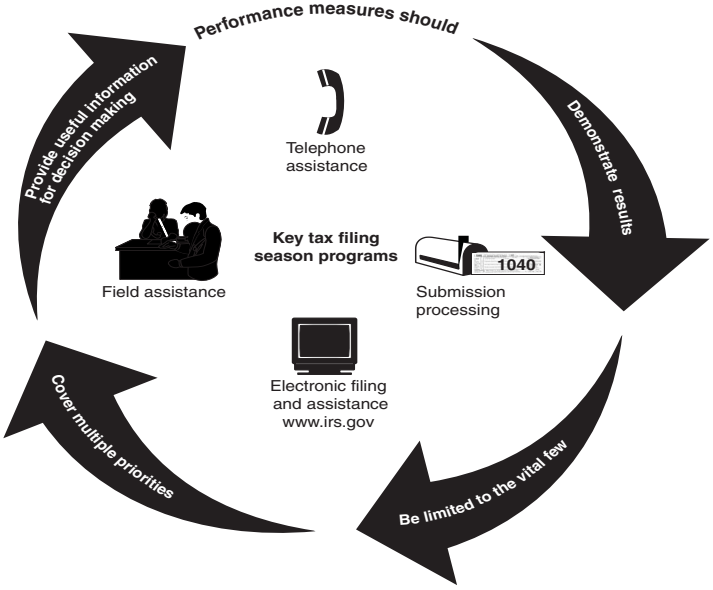
GAO is making recommendations to the Commissioner of Internal Revenue directed at taking actions to better ensure that IRS validates the accuracy of data collection methods for several measures; modifies the formulas used to compute various measures; and adds certain measures, such as cost of service, to its suite of measures.

Of GAO's 18 recommendations, IRS agreed with 12 and discussed actions that had been taken or would be taken to implement them. For 2 of those 12, the actions discussed by IRS did not fully address GAO's concerns. IRS did not agree with the other 6 recommendations.

What GAO Found

In assessing 53 performance measures across IRS's four program areas, GAO found that IRS has made significant efforts to improve its performance measurement system. Many of the measures satisfied some of the four key characteristics of successful performance measures established in earlier GAO work. Although improvements are ongoing, GAO identified instances where measures showed weaknesses including the following: (1) The objectivity and reliability of some measures could be improved so that they will be reasonably free from significant bias and produce the same result under similar circumstances. For example, survey administrators may notify Telephone Assistance's customer service representatives (CSR) too soon that their call was selected to participate in the customer satisfaction survey, which could bias CSR behavior towards taxpayers and adversely affect the measure's objectivity. In addition, the measure Electronic Filing and Assistance uses to determine the number of Web site hits was not reliable because it did not represent the actual number of times the Web site is accessed. (2) The clarity of some performance information was affected when that measure's definition and formula were not consistent. For example, the definition for "CSR response level" measure is the percentage of callers who receive service from a CSR within a specified period of time, but the measure did not include callers who received a busy signal or hung up. (3) Some suites of measures did not cover governmentwide priorities such as quality, timeliness, and cost of service. For example, Field Assistance was missing measures for timeliness and cost of service.

Performance Measures Should Have Four Characteristics



Source: GAO.

The full report, including GAO's objectives, scope, methodology, and analysis is available at www.gao.gov/cgi-bin/getrpt?GAO-03-143. For additional information about the report, contact James White, 202-512-9110 or WhiteJ@gao.gov.

Contents

Letter 1

Results in Brief 3
Background 6
Scope and Methodology 12
Filing Season Performance Measures Have Many of the Attributes of Successful Measures, but Further Enhancements Are Possible 14
Conclusions 36
Recommendations for Executive Action 37
Agency Comments and Our Evaluation 40

Appendix I **Expanded Explanation of Our Attributes and Methodology for Assessing IRS's Performance Measures** 45

Appendix II **The 53 IRS Performance Measures Reviewed** 54

Appendix III **Comments from the Internal Revenue Service** 71

GAO Comments 80

Appendix IV **GAO Contacts and Staff Acknowledgments** 81

GAO Contacts 81
Acknowledgments 81

Bibliography 82

Related Products 83

Tables

Table 1. Key Attributes of Successful Performance Measures 3
Table 2: Overview of Our Assessment of Telephone Assistance Measures 16
Table 3: Overview of Our Assessment of Electronic Filing and Assistance Measures 22

Table 4: Overview of Our Assessment of Field Assistance Measures 27
Table 5: Overview of Our Assessment of Submission Processing Measures 33
Table 6: Telephone Assistance Performance Measures 54
Table 7: Electronic Filing and Assistance Performance Measures 61
Table 8: Field Assistance Performance Measures 64
Table 9: Submission Processing Performance Measures 68

Figures

Figure 1: IRS's Mission and the Link between Its Strategic Goals and the Elements of Its Balanced Measurement System 7
Figure 2: Linkage from IRS Mission to Operating Unit Measure and Target 9
Figure 3: Performance Measures Should Have Four Characteristics 10
Figure 4: Example of Relationship among Field Assistance Goals and Measures 28

Abbreviations

CQRS	Centralized Quality Review Site
CSR	customer service representative
GPRA	Government Performance and Results Act of 1993
IRS	Internal Revenue Service
Q-Matic	Queuing Management System
TAC	Taxpayer Assistance Center
W&I	Wage and Investment

G A O
Accountability * Integrity * Reliability

United States General Accounting Office
Washington, DC 20548

November 22, 2002

The Honorable Amo Houghton
Chairman, Subcommittee on Oversight
Committee on Ways and Means
House of Representatives

Dear Mr. Chairman:

For most taxpayers, their only contacts with the Internal Revenue Service (IRS) are associated with the filing of their individual income tax returns. Most taxpayers file their returns between January 1 and April 15, which is generally referred to as the "filing season."[1] In addition to the filing itself, which can be on paper or electronic, these contacts generally involve millions of taxpayers seeking help from IRS by calling one of IRS's toll-free telephone numbers, visiting one of IRS's field assistance centers, or accessing IRS's Web site on the Internet (www.irs.gov). Between January 1 and July 13, 2002, for example, IRS received about 105 million calls for assistance over its toll-free telephone lines.[2]

As part of a much larger effort to modernize and become more responsive to taxpayers, IRS is revamping how it measures and reports its filing season performance. The new filing season performance measures are to balance customer satisfaction, employee satisfaction, and business results, such as the quality of answers to taxpayer inquiries and the timeliness of refund issuance. IRS intends to use the balanced measures to make managers and frontline staff more accountable for improving filing season performance.

Because so many taxpayers are affected by IRS's performance during the filing season and because the revamped measures are part of a strategy to improve performance, you asked us to review IRS's new set of filing

[1]Although April 15 is generally considered the end of the filing season, millions of taxpayers get extensions from IRS that allow them to delay filing until as late as October 15.

[2]IRS tracks its performance in providing filing season-related telephone service through mid-July instead of April because it receives many filing season-related calls after April 15 from taxpayers who are inquiring about the status of their refunds or responding to notices they received from IRS related to returns they filed.

season performance measures. Those measures belong to the four program areas critical to a successful filing season: telephone assistance; electronic filing and assistance; field assistance; and the processing of returns, refunds, and remittances (referred to as "submission processing"). Specifically, our objective was to assess whether the key performance measures IRS uses to hold managers accountable in the four program areas had the characteristics of a successful performance measurement system.

Previous GAO work indicated agencies successful in measuring performance had performance measures that demonstrate results, are limited to the vital few, cover multiple priorities, and provide useful information for decision making.[3] To determine whether IRS's filing season performance measures satisfy these four general characteristics, we assessed the measures using nine specific attributes.[4] Earlier GAO work cited these specific attributes as key to successful performance measures. Table 1 is a summary of the nine attributes, including the potentially adverse consequences if they are missing. All attributes are not equal and failure to have a particular attribute does not necessarily indicate that there is a weakness in that area or that the measure is not useful; rather, it may indicate an opportunity for further refinement. An expanded explanation of the nine attributes is included in appendix I.

[3]Some earlier work includes U.S. General Accounting Office, *Executive Guide: Effectively Implementing the Government Performance and Results Act*, GAO/GGD-96-118 (Washington, D.C.: June 1996) and U.S. General Accounting Office, *The Results Act: An Evaluator's Guide to Assessing Agency Annual Performance Plans*, GAO/GGD-10.1.20 (Washington, D.C.: Apr. 1998).

[4]The four characteristics are overarching, thus there is not necessarily a direct link between any one attribute and any one characteristic.

Table 1. Key Attributes of Successful Performance Measures

Attributes	Definitions	Potentially adverse consequences of not meeting attribute
Linkage	Measure is aligned with division and agencywide goals and mission and clearly communicated throughout the organization.	Behaviors and incentives created by measures do not support achieving division or agencywide goals or mission.
Clarity	Measure is clearly stated and the name and definition are consistent with the methodology used to calculate it.	Data could be confusing and misleading to users.
Measurable target	Measure has a numerical goal.	Can not tell whether performance is meeting expectations.
Objectivity	Measure is reasonably free from significant bias or manipulation.	Performance assessments may be systematically over- or understated.
Reliability	Measure produces the same result under similar conditions.	Reported performance data is inconsistent and adds uncertainty.
Core program activities	Measures cover the activities that an entity is expected to perform to support the intent of the program.	Not enough information available in core program areas to managers and stakeholders.
Limited overlap	Measure should provide new information beyond that provided by other measures.	Manager may have to sort through redundant, costly information that does not add value.
Balance	Balance exists when a suite of measures ensures that an organization's various priorities are covered.	Lack of balance could create skewed incentives when measures over-emphasize some goals.
Governmentwide priorities	Each measure should cover a priority such as quality, timeliness, and cost of service.	A program's overall success is at risk if all priorities are not addressed.

Source: Summary of information in appendix I.

We shared these attributes with various IRS officials, who generally agreed with their relevance. As discussed in greater detail in the separate scope and methodology section of this report, we took many steps to validate and ensure consistency in our application of the attributes.

We testified before the Subcommittee on Oversight on some of the interim results of our assessment in April 2002.[5]

Results in Brief

In assessing 53 performance measures across four of IRS's key filing season program areas, we found that the measures satisfied many of the nine attributes of successful performance measures previously listed in table 1. As part of its agencywide reorganization, IRS has made significant efforts to improve its performance measurement system, which is to

[5]U.S. General Accounting Office, *Internal Revenue Service: Assessment of Budget Request for Fiscal Year 2003 and Interim Results of 2002 Tax Filing Season*, GAO-02-580T (Washington, D.C.: Apr. 9, 2002).

provide useful information about how well IRS performed in achieving its goals. The improvement of this system is an ongoing process where, in some cases, IRS is only beginning to collect baseline information on which to form targets and develop other measures that would provide better information to evaluate performance results. Despite IRS's progress, we identified instances in all four program areas where the individual measures or suites of measures did not meet some of our nine attributes. Some of these instances represent opportunities for IRS to further refine its measures.

All of the 15 telephone assistance measures had some of the attributes of successful performance measures. Of the more significant problems, five measures had either clarity or reliability problems and one had an objectivity problem. For example,

- five measures did not provide managers and other stakeholders with clear information about the program's performance. For example, the definition for "customer service representative (CSR) response level" is the percentage of callers who receive service from a CSR within a specified period of time, but the formula did not include callers who received a busy signal or hung up; this limitation could lead managers and other stakeholders to conclude that IRS is providing significantly better service than it is.

All of the 13 electronic filing and assistance performance measures fulfilled some of the 9 attributes. The most significant problems involved changing targets, objectivity, and missing measures. For example,

- electronic filing and assistance changed the targets for two of its measures during fiscal year 2001, which could distort the assessment of performance because what was to be observed changed. For example, it changed the target for the "number of 1040 series returns filed electronically" from 42 million to 40 million because midyear data indicated that 42 million 1040 series returns were not going to be filed electronically. Because of the subjective considerations involved, changing the target in this situation also affected the measure's objectivity.

All of field assistance's 14 performance measures satisfied some of the attributes. Many of the more important problems involved clarity and reliability. In addition, some measures were missing, which could cause an emphasis on some program goals at the expense of a balance among all goals. For example,

- the methods used to track workload volume and staff hours expended required manual input that is subject to errors and inconsistencies, which could affect data accuracy and thus the reliability of 8 of field assistance's 14 measures.

- Field assistance did not have timeliness, efficiency, or cost of service measures.

Many of the 11 submission processing measures had the attributes of successful performance measures. Some of the more significant problems related to clarity and reliability. For example,

- one measure—"productivity"—was unclear because it is a compilation of different types of work IRS performs in processing returns, remittances, and refunds and issuing notices and letters. Managers told us that they needed specific information related to their own operations and that the measure's methodology was difficult to understand.

In all four program areas, we were unable, because of documentation limitations, to verify the linkages among IRS's goals and measures. Among other things, such linkages provide managers and staff with a road map that shows how their day-to-day activities contribute to attaining agencywide goals.

We are making recommendations to the Commissioner of Internal Revenue directed at taking actions to better ensure that IRS's filing season measures have the four characteristics of successful performance measures. For example, we are recommending that IRS modify the formulas used to compute various measures; validate the accuracy of data collection methods for several measures; and add certain measures such as cost of service, to its suite of measures.

We requested comments on a draft of this report from the Commissioner of Internal Revenue. We received written comments, which are reprinted in appendix III. In his comments, the Commissioner agreed that there were opportunities to refine some performance measures and said that our observation about the ongoing nature of the performance measurement process was on target. The Commissioner agreed with 12 of our 18 recommendations and discussed actions that had been taken or would be taken to implement them. In 2 of those cases, the actions discussed by IRS did not fully address our concerns. The Commissioner disagreed with the

other 6 recommendations. We discuss the Commissioner's comments in the "Agency Comments and Our Evaluation" section of the report.

Background

In keeping with the Government Performance and Results Act of 1993 (GPRA),[6] IRS revamped its set of filing season performance measures as part of a massive, ongoing modernization effort. Congress mandated the modernization effort in the IRS Restructuring and Reform Act of 1998[7] and intended that IRS would better balance service to taxpayers with enforcement of the tax laws. To implement the modernization mandate, the Commissioner of Internal Revenue developed a strategy composed of five interdependent components. One of those components is the development of balanced performance measures.[8]

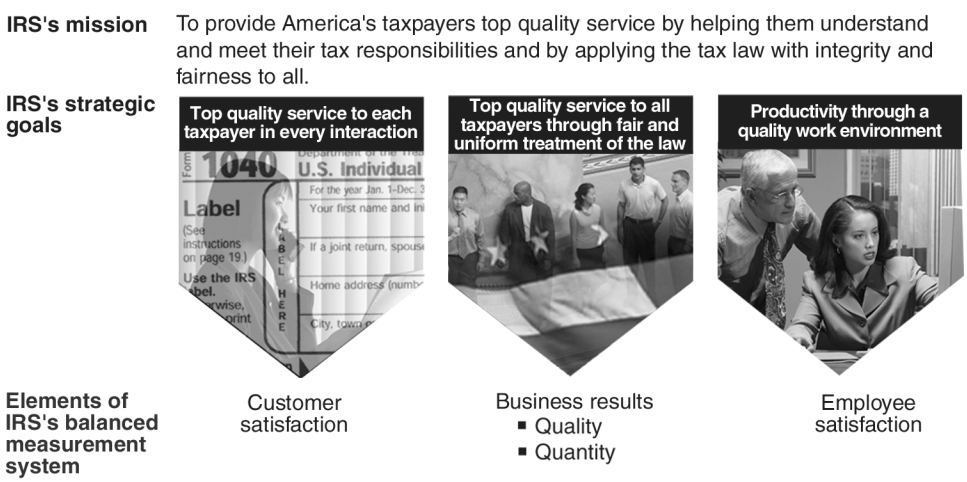
Balanced measures are to emphasize accountability for achieving specific results and to reflect IRS's priorities, which are articulated in its mission and its three strategic goals—top quality service to all taxpayers through fair and uniform application of the law, top quality service to each taxpayer in every interaction, and productivity through a quality work environment. IRS has defined three elements of balanced measures—(1) customer satisfaction, (2) employee satisfaction, and (3) business results (quality and quantity measures)—to ensure balance among its priorities. Figure 1 shows IRS's mission and the link between its strategic goals and the three elements of IRS's balanced measurement system.

[6]GPRA, P.L. 103-62, was enacted to hold federal agencies accountable for achieving program results. IRS's balanced measurement system is consistent with the intent of GPRA.

[7]IRS's Restructuring and Reform Act of 1998, P.L. 105-206, was enacted on July 22, 1998, and calls for broad reforms in areas such as the structure and management of IRS, electronic filing, and taxpayer protection and rights.

[8]The other components include revamped business practices, customer-focused operating divisions, management roles with clear responsibility, and new technology.

Figure 1: IRS's Mission and the Link between Its Strategic Goals and the Elements of Its Balanced Measurement System



Source: GAO depiction of information in IRS Publication 3561 and IRS's Progress Report (December 2001).

IRS intends to use the balanced measures to make managers and frontline staff more accountable for improving filing season performance. We reviewed the performance measures in the four programs areas that interact with taxpayers the most during the filing season—telephone assistance, electronic filing and assistance, field assistance, and submission processing. Each of these program areas is part of IRS's Wage and Investment (W&I) operating division, which generally serves taxpayers whose only income is from wages and investments.[9] Although IRS had measures of performance prior to the reorganization, IRS managers have spent much effort to revamp the filing season performance measures since that time.

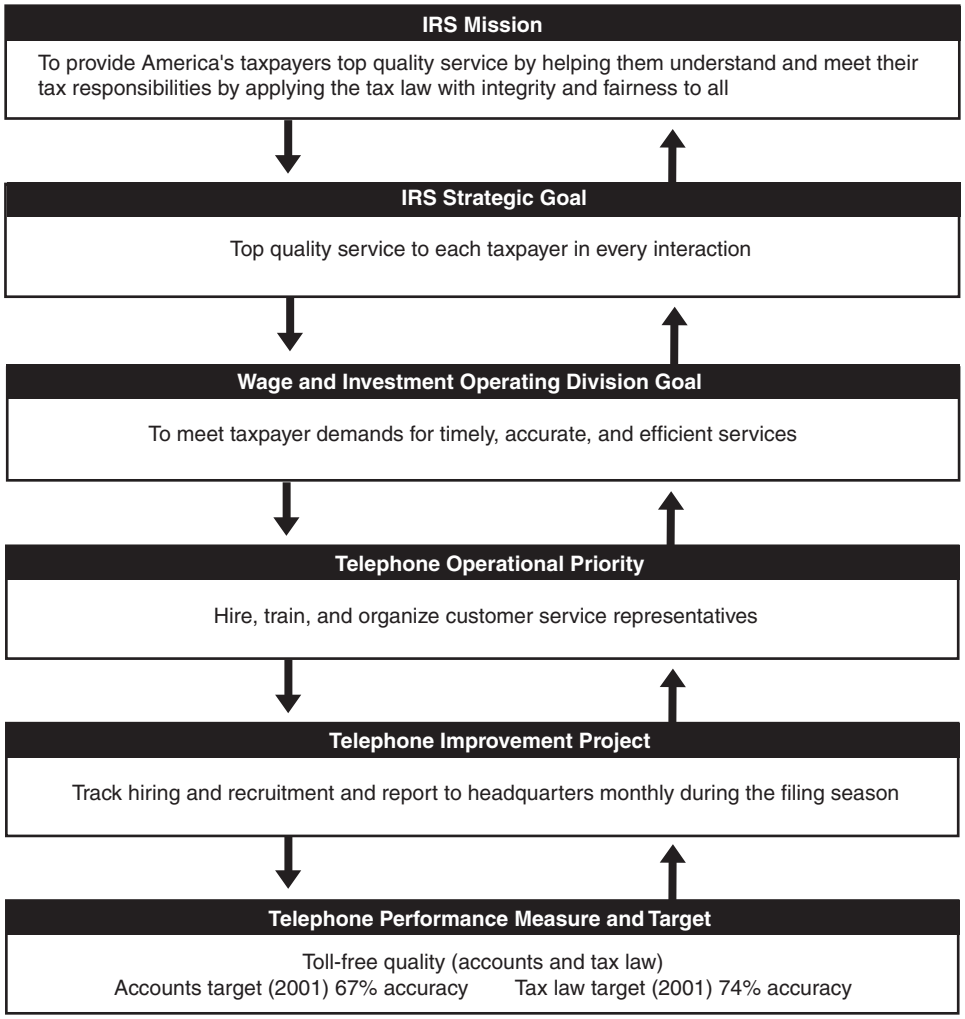
An important aspect of IRS's progress in the challenging task of improving its performance measures was the development of a new Strategic Planning, Budgeting, and Performance Management process in 2000. As part of that process, IRS prepares an annual Strategy and Program Plan

[9]As part of IRS's reorganization that took effect in October 2000, IRS established four operating divisions that serve specific groups of taxpayers. The four divisions are (1) Wage and Investment, (2) Small Business and Self-Employed, (3) Large and Mid-Size Businesses, and (4) Tax Exempt and Government Entities.

that communicates some of the various levels of IRS's goals (e.g., strategic goals, operating division goals) and many performance measures.[10] Although the Strategy and Program Plan does not document all the linkages among the various goals and performance measures, figure 2 is an example we developed to demonstrate the complete relationship from the agency level mission down to the operating unit's measures and targets.

[10]The Strategy and Program Plans we used in our analysis had actual performance information for part of the current fiscal year and planning information for the current and two subsequent fiscal years. An IRS manager said the agency plans to stop including actual information in Strategy and Program Plans prepared after fiscal year 2002.

Figure 2: Linkage from IRS Mission to Operating Unit Measure and Target

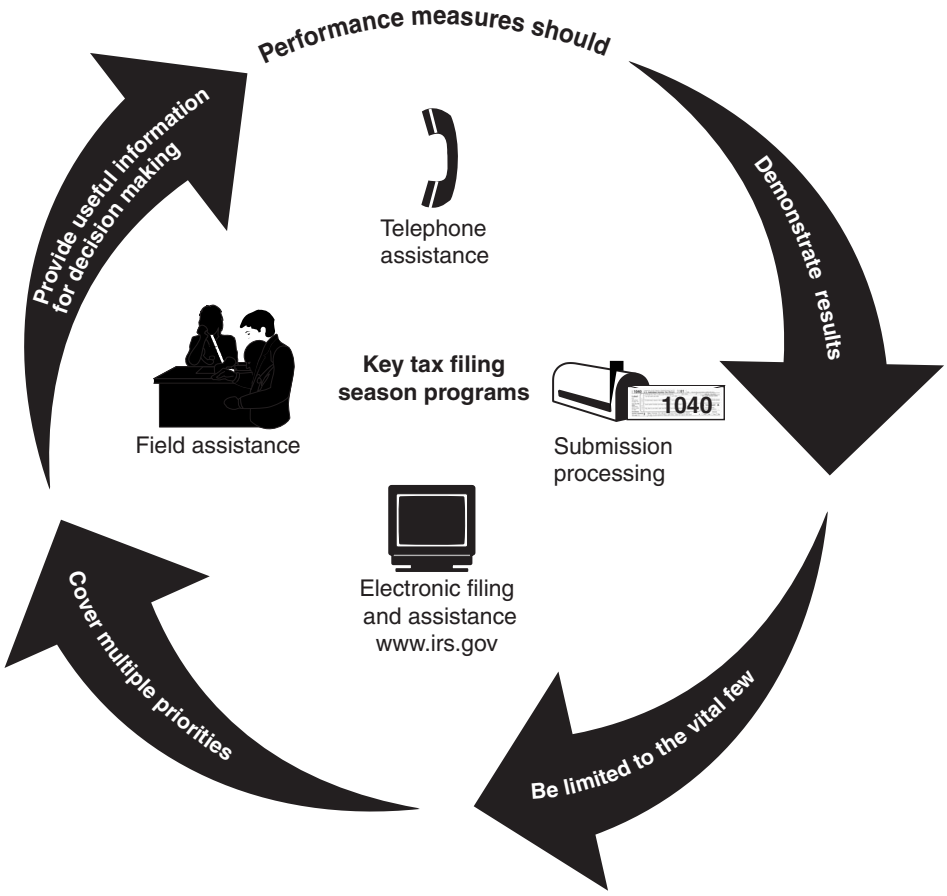


Source: GAO Analysis of IRS's Strategy and Program Plan (October 29, 2001), the W&I Business Performance Review (January 2002), IRS's Progress Report (December 2001) and IRS Publication 3561.

The Strategy and Program Plan is an important document because the Commissioner holds IRS managers accountable for the results of the performance measures contained within it. In addition, many of the measures within the document are presented to outside stakeholders, such as Congress and the public, as key indicators of IRS's performance. The Strategy and Program Plan is the source of the 53 measures we reviewed in the four programs.

As we discussed in our June 1996 guide on implementing GPRA,[11] agencies that were successful in measuring performance strived to establish performance measures that were based on four general characteristics. Those four characteristics are shown in figure 3 as applicable to the four filing season programs we reviewed and are described in more detail following the figure.

Figure 3: Performance Measures Should Have Four Characteristics



Source: GAO.

[11]GAO/GGD-96-118.

Demonstrate results. Performance measures should show an organization's progress towards achieving an intended level of performance or results. Specifically, performance goals establish intended performance, and measures can be used to assess progress towards achieving those goals.

Be limited to the vital few. Limiting measures to core program activities enables managers and other stakeholders to assess accomplishments, make decisions, realign processes, and assign accountability without having an excess of data that could obscure rather than clarify performance issues.

Cover multiple priorities. Performance measures should cover many governmentwide priorities, such as quality, timeliness, cost of service, customer satisfaction, employee satisfaction, and outcomes. Performance measurement systems need to include incentives for managers to strike the difficult balance among competing interests. One or two priorities should not be overemphasized at the expense of others. IRS's history shows why this balance is important. Because of its emphasis on achieving certain numeric targets, such as the amount of dollars collected, IRS failed to adequately consider other priorities, such as the fair treatment of taxpayers.

Provide Useful Information for Decision Making. Performance measures should provide managers and other stakeholders timely, action-oriented information in a format that helps them make decisions that improve program performance. Measures that do not provide managers with useful information will not alert managers and other stakeholders to the existence of problems nor help them respond when problems arise.

On the basis of these four characteristics of successful performance measures, we used various performance management literature to develop a set of nine specific attributes that we used as criteria for assessing IRS's filing season performance measures. The nine attributes are linkage, clarity, measurable target, objectivity, reliability, core program activities, limited overlap, balance, and governmentwide priorities. Appendix I describes these attributes in more detail.

Scope and Methodology

As previously mentioned, we focused our work on four key filing season programs—telephone assistance, electronic filing and assistance, field assistance, and submission processing—within W&I. IRS officials identified the performance measures in the Strategy and Program Plan to be the highest, most comprehensive level of measures for which they are accountable. After discussions with IRS, we decided to review all 53 measures in the Strategy and Program Plan relating to the four filing season programs. We used W&I's draft fiscal year 2001 – 2003 Strategy and Program Plan (dated July 25, 2001) to conduct our review and updated relevant information with the final plan (dated October 29, 2001). Appendix II describes each measure we reviewed in the four program areas and provides other relevant information, such as targets and potential weaknesses.

Our review focused on whether IRS's new set of filing season performance measures had the characteristics of a successful performance measurement system (i.e., demonstrated results, were limited to the vital few, covered multiple priorities, and provided useful information for decision making). For use as criteria in assessing the measures, and as detailed in appendix I, we identified nine attributes of performance measures from various sources, such as earlier GAO work, Office of Management and Budget Circular No. A-11,[12] GPRA, and IRS's handbook on Managing Statistics in a Balanced Measures System.[13] We shared our attributes with IRS officials from various organizations that have a role in developing or monitoring performance measures. Those units included IRS's Organizational Performance Division and several W&I units, such as Strategy and Finance; Planning and Analysis; Customer Account Services; and Communications, Assistance, Research, and Education. Officials in these units generally agreed with the relevance of our attributes and our assessment approach.

We applied the 9 attributes to the 53 filing season measures in a systematic manner, but some judgment was required. To ensure consistency and reliability in our application of the attributes, we had one staff person responsible for each of the four areas. That staff person prepared the

[12]Office of Management and Budget, *Preparation and Submission of Budget Estimates*, Circular No. A-11, Revised. Transmittal Memorandum No. 72 (Washington, D.C.: July 12, 1999).

[13]IRS, *Managing Statistics in a Balanced Measures System*, Handbook 105.4 (Washington, D.C.: Oct. 1, 2000).

initial analysis and at least two other staff reviewed those detailed results. Several staff reviewed the results for all four areas. We did not do a detailed assessment of IRS's methodology for calculating the measures, but looked only at methodological issues as necessary to assess whether a particular measure met the overall characteristics of a successful performance measure.

In applying the attributes, we analyzed numerous pieces of documentation, such as IRS's Congressional Budget Justification, Annual Performance Plan, and data dictionary,[14] and many other reports and documents dealing with the four IRS programs, goals, performance measures, and improvement initiatives. We interviewed IRS officials at various levels within telephone assistance, electronic filing and assistance, field assistance, and submission processing to understand the measures, their methodology, and their relationship to goals, among other things. We also interviewed officials from various IRS organizations that are involved in managing, collecting, and/or using performance data, such as the Organizational Performance Division; Strategy and Finance; Customer Account Services; Statistics of Income; and the Centralized Quality Review Site; and a representative of an IRS contractor, Pacific Consulting Group, responsible for analyzing and reporting the results of telephone assistance's customer satisfaction survey. Appendix I provides more detail on the nine attributes we used, including explanations and examples of each attribute and information on our methodology for assessing each attribute.

We conducted our review in Atlanta, Ga.; Washington, D.C.; Cincinnati, Ohio; and Memphis, Tenn. from September 2001 to September 2002 in accordance with generally accepted government auditing standards.

[14]The data dictionary is an IRS document that provides information on performance measures, such as the measure's name, description, and methodology.

Filing Season Performance Measures Have Many of the Attributes of Successful Measures, but Further Enhancements Are Possible

The 53 filing season performance measures included in our review have many of the attributes of successful performance measures, as detailed in appendix I. For example, in all four of the program areas we reviewed, most measures covered the core activities of each program and had targets in place. In addition, IRS had several on-going initiatives aimed at improving its measures, such as telephone assistance's efforts to revamp all aspects of its quality measures.

At the same time, however, the measures did not satisfy all the attributes, indicating the potential for further enhancements. The nine attributes we used to assess each measure are not equal and failure to have a particular attribute does not necessarily indicate that there is a weakness in that area. In some cases, for example, a measure may not have a particular attribute because benchmarking data are being collected or a measure is being revised. Likewise, a noted weakness, such as a measure not having clarity or being reliable, does not mean that the measure is not useful. For example, telephone assistance's "CSR level of service" measure does not meet our clarity attribute because its name and definition indicate that only calls answered by CSRs are included, but its formula includes some calls answered by automation. This defect currently does not impair the measure's usefulness because the number of automated calls is fairly insignificant. Other weaknesses, however, could lead managers or other stakeholders to draw the wrong conclusions, overlook the existence of problems, or delay resolving problems. For example, electronic filing and assistance's "number of IRS digital daily Web site hits" measure was not considered clear or reliable because it systematically overstates the number of times the Web site is accessed. In total, therefore, the weaknesses identified should be considered areas for further refinement. Such refinements are not expected to be costly or involve significant additional effort on the part of IRS because in many instances our recommendations only include modifications or increased rigor to procedures or processes already in place.

The rest of this report discusses the results of our analysis for each of the four program areas—telephone assistance, electronic filing and assistance, field assistance, and submission processing.

Telephone Assistance Measures

As shown in table 2, all 15 of IRS's telephone performance measures have some of the attributes of successful performance measures.[15] However, as summarized in this section, the measures have several shortcomings. For example, we identified opportunities to improve the clarity of five measures and the reliability of five other measures. Table 6 in appendix II has more detailed information about each telephone measure, including any weaknesses we identified and any recommendations for improvement.

[15]IRS deleted its "automated completion rate" measure in the 2002 Strategy and Program Plan and now has 14 telephone measures. However, IRS still tracks that measure.

Table 2: Overview of Our Assessment of Telephone Assistance Measures

Telephone assistance measures	Telephone assistance attributes								
	Linkage[a]	Clarity	Measurable target	Objectivity	Reliability	Core program activities[b]	Limited overlap	Government-wide priorities	Balance
Total automated calls answered		✓	✓	✓	✓	✓		✓	This attribute applies to the overall suite of measures rather than to the measures individually.[d]
Customer Service Representative (CSR)[c] calls answered		✓	✓	✓	✓	✓		✓	
CSR level of service			✓	✓	✓	✓	✓	✓	
Toll-free customer satisfaction			✓		✓	✓	✓	✓	
Toll-free tax law quality		✓	✓	✓		✓		✓	
Toll-free accounts quality		✓	✓	✓		✓		✓	
Average handle time		✓		✓	✓	✓	✓	✓	
Automated completion rate				✓	✓	✓		✓	
CSR services provided				✓	✓	✓		✓	
Toll-free tax law correct response rate		✓	✓	✓		✓		✓	
Toll-free account correct response rate		✓	✓	✓		✓		✓	
Toll-free timeliness		✓	✓	✓		✓	✓	✓	
Toll-free employee satisfaction		✓	✓	✓	✓	✓	✓	✓	
CSR response level			✓	✓	✓	✓		✓	
Average speed of answer		✓		✓	✓	✓		✓	

Note: A check mark denotes that the measure has the attribute.

[a]We were unable to verify the linkages between goals and measures because of insufficient documentation.

[b]Core program activities of telephone assistance are to provide timely and accurate assistance to taxpayers with inquiries about the tax law and their accounts.

[c]IRS also refers to CSRs as assistors.

[d]IRS considers that these measures are balanced because they address priorities, such as customer and employee satisfaction and business results. However, including measures, such as cost of service, could improve the balance of telephone assistance's program priorities.

Source: GAO analysis.

No Documentation Shows the Complete Linkage between Agencywide Goals and Telephone Measures

Although telephone assistance management stated that their goals and measures generally aligned, we were unable to verify this because no documentation shows the complete relationship. For example, some documentation may show a link from a measure to an agencywide goal, but the operating division level goals were omitted. When we attempted to create the linkage ourselves, we found it difficult to determine how some measures related to the different agencywide and operating division goals. When we asked some IRS officials to describe the complete link, they too had a difficult time and were uncertain of some connections.

Telephone assistance managers stated that staff received performance management training that should help them to understand their role in

helping the organization achieve its goals. However, having clear and complete documentation would provide evidence that linkages exist and help prevent misunderstandings. When employees do not understand the relationship between goals and measures, they may not understand how their work contributes to agencywide efforts and, thus, goals may not be achieved.

Most Telephone Measures Have Clarity

Ten of the 15 measures have clarity (e.g., "automated calls answered" clearly describes the count of all toll-free calls answered at customer service sites by automated service). However, five measures contain or omit certain data elements that can cause managers or other stakeholders to misunderstand the level of performance. For example, the "CSR response level," is defined as the percentage of callers who started receiving service from a CSR within a specified period of time. However, this may not reflect the real customer experience at IRS because the formula for computing the measure does not include callers who tried to reach a CSR but did not, such as callers who (1) hung up while waiting to speak to a CSR, (2) were provided access only to automated services and hung-up, and (3) received a busy signal.[16] (The other four measures, as noted in table 6 in appendix II, are "CSR level of service," "automated completion rate," "CSR service provided," and "toll-free customer satisfaction.")

Measures that do not provide clear information about program performance may affect the validity of managers' and stakeholders' assessments of IRS's performance, possibly leading to a misinterpretation of results or a failure to take proper action to resolve performance problems.

Most Telephone Measures Have Targets

Eleven of the 15 measures have numerical targets that facilitate the future assessment of whether overall goals and objectives were achieved. Of the four measures with no targets, three were measures for which IRS was collecting data for use in developing first-time targets and one was a measure ("automated completion rate") that IRS was no longer tracking in the Strategy and Program Plan. Although we generally disagree with the removal of the "automated completion rate" measure from the Strategy and Program Plan, as described in an upcoming section, not having targets in these instances is reasonable.

[16]There were about 30 million of these calls during in fiscal year 2001, which can have a significant impact on the "CSR response level" measure.

Data Collection Methods for Telephone Assistance's Customer Satisfaction Measure Are Not Always Objective

IRS determines customer satisfaction with its toll-free telephone assistance through a survey administered to taxpayers who speak with a CSR.[17] We observed survey collection methods in Atlanta that were not always objective; that is, the administrators did not always follow prescribed procedures for selecting calls to participate in the survey. Not following prescribed procedures produces a systematic bias that could compromise the randomness of the sample. Also, IRS procedures do not require that administrators listen to the entire call. Although administrators are instructed to notify the CSR towards the end of a call that the call was selected for the survey, this may not occur. If an administrator begins listening to a call after it has started, it can be difficult to determine the full nature of the taxpayer's question and thus whether the conversation is about to end. As a result, an administrator could prematurely notify a CSR that the call was selected for the survey, which could change the CSR's behavior towards the taxpayer and affect the results of the survey and the measure. In addition, administrators may not be able to correctly answer certain questions on the survey, which could impair any analysis of those answers. We discussed these issues with a representative of the IRS contractor (Pacific Consulting Group) responsible for analyzing and reporting the survey results who said that (1) he was aware of these problems and (2) the same problems existed at other locations.

IRS has taken corrective action on one of these weaknesses. Because management decided that the procedures for selecting calls to participate in the customer satisfaction survey were too difficult to follow, it revised them. Sites began using the revised sampling procedures in July 2002.

Reliability of Five Telephone Quality Measures Is Suspect

The reliability of telephone assistance's five quality measures ("toll-free tax law quality," "toll-free accounts quality," "toll-free tax law correct response rate," "toll-free account correct response rate," and "toll-free timeliness") is suspect because of potential inconsistencies in data collection that arise due to differences among individual reviewer's judgment and perceptions.[18] Although it is not certain how much variation among reviewers exists, errors could occur throughout data collection and could affect the results of the measures and conclusions about the extent to which performance goals have been achieved.

[17]CSRs answer about 24 percent of all incoming calls.

[18]As of January 2002, there were 53 quality reviewers in the Centralized Quality Review Site: 26 for tax law inquiries, 20 for account inquiries, and 7 others.

Reliability and credibility increase when performance data are checked or tested for significant errors. IRS has conducted consistency reviews in the past and found problems. It has taken steps to improve consistency, the most important of which was the establishment of the Centralized Quality Review Site (CQRS).[19] Among other controls within CQRS that are designed to enhance consistency, reviewers are to receive the same training and gather to discuss cases where the guidance is not clear. IRS has conducted one review to determine the effectiveness of CQRS and its efforts to improve consistency since IRS's October 2000 reorganization and continues to find some problems.

At the time of our review, IRS was reviewing the five quality measures as part of an ongoing improvement initiative. Since that time, it redesigned many aspects of the measures, including what is measured, how the measures are calculated, how data are collected, and how people are held accountable for quality.[20] Changes emanating from this initiative may further enhance consistency.

Telephone Measures Cover Core Program Activities

Telephone assistance's core program activities are to provide timely and accurate assistance to taxpayers with inquiries about the tax law and their accounts. IRS has at least one measure that directly addresses each of these core activities. For example, "toll-free accounts quality" is a measure that shows the percentage of accurate responses to taxpayers' account related questions.

Some Overlap Exists between Telephone Assistance Measures

The amount of overlap that exists between measures is a managerial decision. Of the 15 telephone measures we reviewed, 10 have at least partial overlap. For example, both the "CSR response level" and "average speed of answer" measures attempt to show how long a taxpayer waited before receiving service, except that the former shows the number of

[19]CQRS is responsible for monitoring the accuracy of telephone assistance. It produces various reports that show call sites what errors CSRs are making so site managers can take action to reduce those errors.

[20]IRS significantly modified its five quality measures beginning in October 2002 based on the results of its initiative, which was aimed at redesigning the way IRS measures quality to better capture the taxpayer's experience. Specifically, IRS renamed the toll-free correct response rate measures for tax law and account inquiries to "customer accuracy" for tax law or account inquiries. Plans call for the tax quality measures for tax law and account inquiries to be discontinued, but reported in fiscal year 2003 for trending and comparative purposes. IRS also eliminated the "toll-free timeliness" measure and replaced it with a new "quality timeliness" measure. Finally, IRS implemented a new measure called "professionalism."

taxpayers receiving service within 30 seconds while the latter shows the average wait time for all taxpayers. (Table 6 in appendix II has information on other overlapping measures.)

IRS officials said that overlapping measures can add value to management's decision-making process because each measure provides a nuance that can be missed if both measures were not present. For example, the "CSR calls answered" measure shows the number of taxpayer calls answered while the "CSR services provided" measure attempts to account for situations in which more than one CSR was involved in handling a single call. At the same time, however, overlapping measures (1) leave managers to sift through redundant, sometimes costly, information to determine goal achievement and (2) could confuse outside stakeholders, such as Congress.

Although we are not suggesting that IRS stop tracking or reporting any of the overlapping measures, we question whether IRS has limited the telephone measures included in the Strategy and Program Plan to the vital few. Telephone officials agreed with this assessment and stated that some of the overlapping measures will be removed from future Strategy and Program Plans.

Telephone Measures Do Not Fully Cover Governmentwide Priorities

When considering governmentwide priorities, such as quality, timeliness, cost of service, and customer and employee satisfaction, telephone assistance is missing two measures—(1) cost of service and (2) a measure of customer satisfaction for automated services, as described below.

- *Cost of Service*. According to key legislation[21] and accounting standards,[22] agencies should develop and report cost information. Besides showing financial accountability in the use of taxpayer dollars, the cost information called for can be used for various purposes, such as authorizing and modifying programs and evaluating program performance. IRS does not report the average cost to answer a taxpayer's inquiry by telephone. A cost-per-call analysis could provide a

[21]The Chief Financial Officer's Act, P.L. 101-576,underscores the importance of improving financial management in the federal government. Among other things, it calls for developing and reporting cost information.

[22]Statement of Federal Financial Accounting Standard Number 4, "Managerial Cost Accounting Concepts and Standards for the Federal Government," is aimed at providing reliable and timely information on the full cost of federal programs, their activities, and outputs.

link between program goals and costs, as required by GPRA, and help IRS management and Congress decide about future investments in telephone assistance. IRS officials said they would like to develop a cost of services measures and are trying to determine what information would be meaningful to include or exclude in the calculation.

- *Customer Satisfaction for Automated Services*. Although IRS projections show that about 70 percent of its fiscal year 2002 calls would be handled by automation, it has no survey mechanism in place to determine taxpayers' satisfaction with these automated services. IRS officials agreed this would be a meaningful measure and want to develop one for the future, but no implementation plans have been established.

Also, as previously mentioned, IRS has removed the "automated completion rate" measure from its Strategy and Program Plan. We realize, as noted in table 6 in appendix II, that this measure has limitations that need to be addressed. However, because such a large percentage of calls are handled by automation and because IRS plans to serve even more calls with automation in the future, re-inclusion of that measure in the Strategy and Program Plan may be warranted if the associated problems can be resolved.

Telephone Measures Are Balanced

Telephone assistance has measures in place for customer satisfaction, employee satisfaction, and business results and, therefore, IRS considers the measures balanced. However, including other measures, such as a cost of service measure, as previously described, could further enhance the balance of program priorities.

Electronic Filing and Assistance Measures

As shown in table 3, all 13 of electronic filing and assistance's performance measures have some of the attributes of successful performance measures. However, as summarized in this section, the measures have some shortcomings. For example, several of the measures had some overlap and two measures had shortcomings related to the changing of targets during the fiscal year. Table 7 in appendix II has more detailed information about each electronic filing and assistance measure, including any weaknesses we identified and any recommendations for improvement.

Table 3: Overview of Our Assessment of Electronic Filing and Assistance Measures

Electronic filing and assistance measures	Electronic filing and assistance attributes								
	Linkage[a]	Clarity	Measurable target	Objectivity	Reliability	Core program activities[b]	Limited overlap	Government-wide priorities	Balance
Number of 1040 series returns electronically filed		✔			✔	✔		✔	This attribute applies to the overall suite of measures rather than to the measures individually.[c]
Number of business returns electronically filed		✔	✔	✔	✔	✔	✔	✔	
Total number of electronically filed returns		✔			✔	✔	✔	✔	
Number of information returns electronically filed		✔	✔	✔	✔	✔		✔	
Percent of information returns electronically filed		✔	✔	✔	✔	✔		✔	
Percent of individual returns electronically filed		✔	✔	✔	✔	✔		✔	
Number of payments received electronically		✔	✔	✔	✔	✔		✔	
Percent of payments received		✔	✔	✔	✔	✔		✔	
Number of electronic funds withdrawls/credit card transactions		✔	✔	✔	✔	✔		✔	
Number of IRS digital daily Web site hits			✔	✔		✔	✔	✔	
Number of downloads from "IRS.GOV"		✔	✔	✔	✔	✔	✔	✔	
Customer satisfaction-individual taxpayers		✔	✔	✔	✔	✔	✔	✔	
Employee satisfaction-electronic filing and assistance		✔	✔	✔	✔	✔	✔	✔	

Note: A check mark denotes that the measure has the attribute.

[a]We were unable to verify the linkages between goals and measures because of insufficient documentation.

[b]Electronic filing and assistance's core program activities are to provide individual and business taxpayers with the capability to transact and communicate electronically with IRS.

[c]Electronic filing and assistance measures address most governmentwide priorities, such as quantity, customer satisfaction, and employee satisfaction; however, they do not cover two important priorities—quality and cost of service.

Source: GAO analysis.

Overall Alignment of Electronic Filing and Assistance's Goals and Measures Not Fully Documented

Electronic filing and assistance's 13 performance measures are aligned with IRS's overall mission and IRS's strategic goals. However, we were unable to validate whether the lower level goals, such as electronic filing and assistance's operational goals and improvement projects, are linked to the agencywide strategic level goals and operating division performance measures because there is not complete documentation available to show that linkage.

Electronic filing and assistance's managers stated that goals and measures generally align and that employee briefings were held to communicate their goals to the organization. It is essential that all staff be familiar with IRS's mission and goals, electronic filing and assistance's goals and performance measures, and how electronic filing and assistance

determines whether it is achieving its goals so that staff know how their day-to-day activities contribute to the goals and IRS's overall mission. When this is lacking, priorities may not be clear and staff efforts may not be tied to goal achievement.

Most Electronic Filing and Assistance Measures Have Clarity

All but one of electronic filing and assistance's 13 performance measures had clarity. The "number of IRS digital daily Web site hits" measure, which is defined as the number of "hits" to IRS's Web site, is not clear because its formula counts multiple hits every time a user accesses the site's home page and counts a hit every time a user moves to another page on the Web site. The formula is not consistent with the definition because it does not represent the actual number of times the Web site is accessed.

In its fiscal year 2003 Annual Performance Plan,[23] IRS acknowledged limitations with this measure as follows.

> " . . . changes in the IRS Web design may cause a decrease in the number of 'hits' recorded in both [fiscal years] 2002 and 2003. This decrease will be due to improved Web site navigation and search functions, which may reduce the amount of random exploration by users to find content. The decrease will also be due to better design of the Web pages themselves that will reduce the number of graphics and other items that are used to create the Web page, all of which are counted as 'hits' when a page is accessed."

In our report on IRS's 2001 tax filing season, we recommended that IRS either discontinue the use of "hits" as a measure of the performance of its Web site or revise the way "hits" are calculated so that the measure more accurately reflects usage.[24] IRS responded that it should continue to count "hits" as a measure of the Web site's performance because "hits" indicate site traffic and can be used to measure system performance and estimate system needs. However, officials stated that they could improve their method of counting "hits" once they had implemented a more sophisticated, comprehensive Web analytical program. According to electronic filing and assistance officials, IRS introduced its redesigned Web site in January 2001 and implemented a new analytical program, but "hits" are still being calculated the same way.

[23]The Annual Performance Plan is a key document IRS produces each year to comply with the requirements of GPRA. It highlights a limited number of IRS performance measures.

[24]U.S. General Accounting Office, *Tax Administration: Assessment of IRS's 2001 Tax Filing Season*, GAO-02-144 (Washington, D.C.: Dec. 21, 2001).

Two Electronic Filing and Assistance Measures Had Targets Changed and Lack Objectivity

Electronic filing and assistance changed the targets for two measures—"number of 1040 series returns filed electronically"[25] and "total number of returns electronically filed"—during fiscal year 2001. Changing targets could distort the assessment of performance because what was to be observed changed. No major event (such as legislation that affected the ability of many taxpayers to file electronically) happened that warranted changing the targets in the strategic plan. Instead, electronic filing and assistance changed the target for the first of those measures from 42 million returns to 40 million returns because IRS's Research Division's midyear data indicated that 42 million 1040 series returns were not going to be filed electronically. Because the number of 1040 series returns filed electronically is a subset of the total number of returns filed electronically, electronic filing and assistance also reduced the target for total electronic filings. Because of these subjective considerations, changing the targets in this situation also affected the objectivity of these measures.

Electronic Filing and Assistance Measures Are Reliable, with One Exception

Of electronic filing and assistance's 13 performance measures, we considered 12 to be reliable because the data on performance comes from sources, such as IRS's masterfile[26] and computer program runs, that are subject to validity checks. The one measure we did not consider reliable was the "number of IRS digital daily Web site hits," because it does not represent the actual number of times the Web site is accessed, as previously described.

Measures Cover Electronic Filing and Assistance's Core Program Activities

Electronic filing and assistance's core program activities are to provide individual and business taxpayers the capability to transact and communicate electronically with IRS. Electronic filing and assistance focuses on taxpayers' ability to file their returns, pay their taxes, receive assistance, and obtain information electronically. These core activities are all covered by the 13 performance measures.

Overlap Exists among Electronic Filing and Assistance Measures

Seven of the 13 electronic filing and assistance measures had partial overlap. For example, the "number of 1040 series returns electronically filed" and "percent of individual returns electronically filed" measures provide related information on a key program activity. The difference is that the former is a count of the number filed electronically while the latter is the percentage of total individual tax returns filed electronically.

[25] 1040 series returns are individual income tax returns filed on Forms 1040, 1040A, and 1040EZ.

[26] The masterfile is the system where most of IRS's taxpayer data resides.

(Table 7 in appendix II has information on other overlapping electronic filing and assistance measures.)

The amount of overlap to tolerate among measures is management's judgment. Electronic filing and assistance officials told us that each of the overlapping measures we identified provides additional information to managers. For example, the "number of 1040 series returns electronically filed" provides managers with information on the size of the electronic return workload whereas the "percent of individual returns electronically filed" tells them how they are doing in relation to IRS's long-term strategic goal of 80 percent. IRS officials also pointed out that both number and percent performance measures exist because external customers, such as the press, like to use the measures for reporting purposes.

Electronic Filing and Assistance's Measures Do Not Cover Some Governmentwide Priorities, Thus Hindering Balance

Although electronic filing and assistance's measures address several governmentwide priorities, such as quantity, customer satisfaction, and employee satisfaction, they do not cover two important priorities—quality and cost of service. As a result, its performance measurement system is not fully balanced.

Electronic filing and assistance classifies four of its performance measures as quality measures, but the measures are merely counts of certain types of electronic transactions (such as "number of payments received electronically"). On the other hand, it tracks what we consider to be quality measures (i.e., "processing accuracy"[27] and "refund timeliness, electronically filed")[28] but those measures are not in the Strategy and Program Plan. These quality measures and others, such as one that tracks the number of electronic returns rejected,[29] could be important indicators of program success or failure. For example, IRS data indicate that many electronic tax returns are rejected; a measure that captures the volume of rejects could help to focus management's attention on the cause of those rejects.

[27]"Processing accuracy" refers to the total number of returns that do not go to the error resolution system. Transactions that fail validity checks during processing are corrected through the error resolution system.

[28]"Refund timeliness, electronically filed" is the amount of time it takes for taxpayers to receive their refunds when filing electronically.

[29]Electronic returns can be rejected, for example, if taxpayers fail to include required Social Security numbers. IRS requires taxpayers to correct such errors before it will accept their electronic returns.

Also, similar to our discussion of a cost of service measure in the telephone section, a "cost-per-electronically filed return" could provide a link between program goals and costs, as required by GPRA, and help IRS management and Congress decide about future investments in electronic filing and assistance.

Field Assistance Measures

As shown in table 4, all 14 of field assistance's performance measures have some of the attributes of successful performance measures. However, as summarized in this section, the measures have several shortcomings, primarily with respect to clarity, reliability, and balance. Table 8 in appendix II has more detailed information about each field assistance measure, including any weaknesses we identified and any recommendations for improvement.

Table 4: Overview of Our Assessment of Field Assistance Measures

Field assistance measures	Field assistance attributes								
	Linkage[a]	Clarity	Measurable target	Objectivity	Reliability	Core program activities[b]	Limited overlap	Government-wide priorities	Balance
Customer satisfaction		✔	✔	✔	✔	✔	✔	✔	This attribute applies to the overall suite of measures rather than to the measures individually.[c]
Return preparation contacts			✔	✔		✔		✔	
Geographic coverage			✔	✔	✔	✔	✔	✔	
Return preparation units				✔		✔		✔	
Tax Assistance Centers (TAC) total contacts (includes return preparation contacts)			✔	✔		✔	✔	✔	
Forms contacts			✔	✔		✔	✔	✔	
Tax law contacts				✔		✔	✔	✔	
Account contacts				✔		✔	✔	✔	
Other contacts			✔	✔		✔	✔	✔	
Tax law accuracy				✔	✔	✔	✔	✔	
Accounts/notices accuracy				✔	✔	✔	✔	✔	
Return preparation accuracy				✔	✔	✔	✔	✔	
Employee satisfaction		✔	✔	✔	✔	✔	✔	✔	
Alternate contacts		✔		✔		✔	✔	✔	

Note: A check mark denotes that the measure has the attribute.

[a]We were unable to verify the linkages between goals and measures because of insufficient documentation.

[b]Core program activities of field assistance are to provide face-to-face assistance, education, and compliance services.

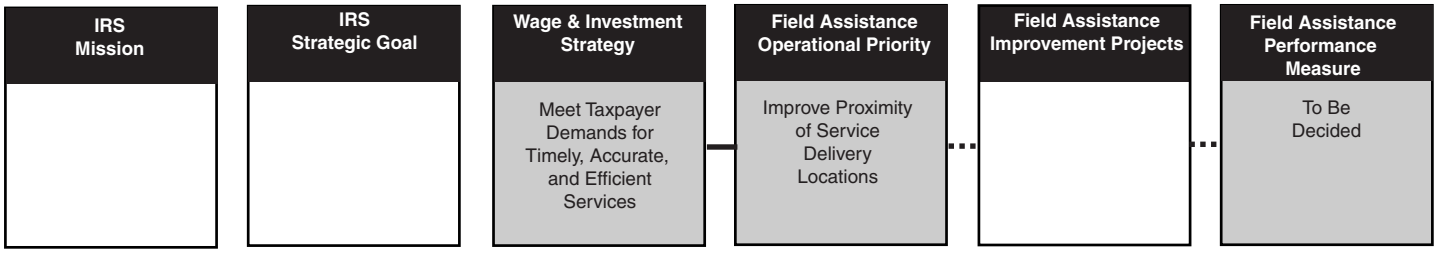
[c]Although field assistance continues to develop its suite of performance measures, important measures of timeliness, efficiency or productivity, and cost of service are missing and impair balance.

Source: GAO analysis.

Relationship between Goals and Field Assistance Measures Not Complete

Field assistance recognizes the importance of creating a clear relationship between goals and measures and has developed a template that shows some of that relationship. Figure 4 is an excerpt of the template, with the completed portions, as of October 2002, shown in gray.

Figure 4: Example of Relationship among Field Assistance Goals and Measures



Source: GAO's analysis of field assistance's business plan template.

Although the template demonstrates a noteworthy effort to show a clear link between goals and measures, it omits the link to IRS's mission, IRS's strategic goals, and field assistance's improvement projects. These links are important because they serve as the bridge between long-term strategic goals and short-term daily operational goals, which can, among other things, be used for holding IRS and the field assistance program accountable for achieving those goals. Also, officials told us that the completed template would only cite the type of performance measure—employee satisfaction, customer satisfaction, or business results—not the specific measure and target. The link to the specific measure provides additional information needed to clearly communicate the alignment of goals and measures throughout the agency, and the target communicates the level of performance the operating division hopes to achieve.

Many Field Assistance Measures Lack Clarity

Many of field assistance's measures lack clarity. For example, the "geographic coverage" measure is unclear, even to IRS officials, because it is not evident by its name or definition what is or is not included in the measure's formula. Specifically, officials debated whether or not the measure included alternate sites[30] and kiosks.[31] Similarly, the formula only considers the location of Taxpayer Assistance Centers (TAC), not their hours of operation or services provided. Although we saw no evidence that this lack of clarity led to adverse consequences, it could. For example,

[30]Alternate sites are staffed with field assistance employees and offer limited face-to-face services, such as preparing returns and distributing forms. Field assistance has about 50 alternate sites, such as temporary sites in shopping malls and libraries. Alternate sites are currently not included in the "geographic coverage" measure.

[31]Kiosks are automated machines that taxpayers can use to obtain certain forms, answers to frequently asked questions, and general IRS information in English and Spanish. Kiosks are currently not included in the "geographic coverage" measure.

management or other stakeholders may determine that TACs are needed in certain areas of the country to improve geographic coverage when, in fact, alternate sites and/or kiosks are already serving those areas. IRS officials said that they have plans to revise the formula to include alternate sites and kiosks. (The other measures that lack clarity, as described in table 8 of appendix II, are "return preparation contacts," "return preparation units," "TACs total contacts," "forms contact," "tax law contacts," "account contacts," "other contacts," "tax law accuracy," "accounts/notices accuracy," and "return preparation accuracy.")

All Field Assistance Measures Are Objective and Have Targets That Are Either in Place or Being Established

We determined that all of field assistance's 14 performance measures are objective because, to the greatest extent possible, they are free of significant bias or manipulation and indicate specifically what is to be observed, in which population or conditions, and in what timeframes. Of the 14 measures, 7 have targets in place to help determine whether overall goals and objectives were achieved. Of the seven measures without targets, three were being baselined (i.e., IRS was collecting data for use in setting first-time targets). The remaining four measures were being designed at the time of our review. Targets will be set for these measures upon completion of data collection.

Data Collection Process Affects Reliability of Several Field Assistance Measures

Eight of field assistance's 14 performance measures are based on a data collection process that is subject to inconsistencies and human error, meaning that the same results may not be produced in similar circumstances. All TAC employees are to use Form 5311 (Field Assistance Activity Report) to manually report their daily hours and type of assistance provided. Supervisors are to review the forms for accuracy and forward them for manual input into the Resources Management Information System.[32] These layers of manual input are subject to error and can hinder data reliability that could (1) lead managers or other stakeholders to draw inappropriate conclusions about program performance, (2) not alert them to the existence of problems, or (3) not help them respond when problems arise. For example, as we noted in our report on IRS's 2001 tax filing season, our calculations showed that the data reported by TACs did not account for the wait times of about 661,000 taxpayers, or about 13 percent of taxpayers served.[33] IRS expects to minimize this human error by equipping all of its TACs with an on-line automated tracking and reporting

[32]The Resources Management Information System is the primary management information system that field assistance uses to track workload volume and staff hour expenditures.

[33]GAO-02-144.

system known as the Queuing Management System (Q-Matic). This system is expected, among other things, to more efficiently monitor customer traffic flow and wait times and eliminate staff time spent completing Form 5311.[34]

IRS has taken steps to solve data reliability problems with field assistance's customer satisfaction measure. In a May 2000 report, the Treasury Inspector General for Tax Administration concluded that IRS had not established an adequate management process to ensure that the survey yielded accurate, reliable, and statistically valid results.[35] To field assistance's credit and with the help of a vendor, it (1) completed major revisions to the customer satisfaction survey, such as using a different index scale; (2) included space for written comments, which were to be provided to managers on a routine basis; and (3) improved controls to ensure the survey is available to all taxpayers. However, problems arose regarding the manner in which the vendor was providing site managers with data containing cumulative responses and, as of June 2002, the vendor had temporarily stopped providing feedback to site managers and was in the process of determining a more usable format to relay information to managers. The improved data collection method is being implemented and IRS anticipates an increase in the precision with which it measures field assistance customer satisfaction.

Field Assistance Measures Cover Core Program Activities with Limited Overlap

Field assistance's measures cover its core program activities with limited overlap. Field assistance identifies its core program activities as face-to-face assistance, education, and compliance services, which include such activities as preparing returns, answering tax law questions, resolving account and notice inquiries, and supplying forms and publications. For example, field assistance has an "accounts contact" measure (counts the number of contacts made) and an "accounts accuracy" measure (measures the accuracy of the responses) to reflect both the quantity and quality of its accounts-related assistance.

[34]Of about 420 TACs, 123 had Q-Matic as of June 2002. IRS officials stated that installation and networking of Q-Matic in all offices is scheduled to be complete by September 30, 2005. In the meantime, IRS plans to pilot an installed and networked Q-Matic system in all the TACs that are located in one of IRS's seven management areas during the first quarter of 2003.

[35]Treasury Inspector General for Tax Administration, *Walk-in Customer Satisfaction Survey Results Should Be Qualified If Used for the GPRA*, 2000-10-079 (Washington, D.C.: May 17, 2000).

Field assistance identified some overlap between two measures, "return preparation contacts" and "return preparation units." It has decided, for Strategy and Program Plan purposes, to discontinue the "contacts" measure (which counts the number of customers assisted) and keep the "units" measure (which counts the number of returns prepared) because the "units" measure better reflects the amount of return preparation work done.[36] Field assistance will continue tracking the "contacts" measure outside of the Strategy and Program Plan in order to determine customer demand for service at particular sites. We concur with IRS's plans to track the "contacts" measure outside of the Strategy and Program Plan because it is a diagnostic tool that can be used for analysis purposes.

Field Assistance Is Missing Some Measures Needed to Balance Governmentwide Priorities

Field assistance continues to develop its suite of performance measures. As part of that effort, it is beginning to deploy important quality measures, such as "tax law accuracy." However, other important measures of timeliness, efficiency, and cost of service are missing, which impairs balance.

- *Timeliness*. Before fiscal year 2001, field assistance had a performance measure that officially tracked how long customers waited to receive service from an employee. According to managers, it was discontinued because employees were serving taxpayers as quickly as possible in order to meet timeliness goals, which negatively affected service quality.[37] In March 2002, management went further and (1) eliminated its requirement for TACs not equipped with Q-Matic to submit biweekly wait-time reports and (2) doubled, from 15 to 30 minutes, the wait-time interval to be used by TACs with Q-Matic in computing the percentage of customers served on time. Officials said that they took these steps because employees continued to feel pressured to hurry assistance despite the discontinuance of the official timeliness measure. However, one purpose of balanced measures is to avoid an inappropriate emphasis on just one aspect of performance. The presence of a quality measure should provide a disincentive for employees to ignore quality in favor of timeliness. Similarly, in the absence of a timeliness performance measure, (1) field assistance may not be balancing its

[36]The number of units would generally be larger than the number of contacts. For example, if a taxpayer received help in preparing his or her return and his or her child's return, field assistance would count that service as one return preparation contact and two return preparation units.

[37]TACs monitor timeliness, but IRS does not report the measure in the Strategy and Program Plan.

customers' needs for timely service with their needs for accurate information and (2) IRS is not held accountable for timeliness to stakeholders, such as the Congress.

- *Efficiency*. Efficiency, or productivity as it is often referred to, shows how efficiently IRS's resources are transformed into the production of field assistance services. Field assistance officials said they would like to develop an efficiency measure, but no plans are in place. Among other things, having an efficiency measure would help managers identify performance strengths and weaknesses.

- *Cost of Service*. As required by GPRA, agencies should have performance measures that correlate the level of program activity and program cost. Without such a measure in field assistance, officials do not know how much it costs to provide face-to-face service. Field assistance officials said that they would like to develop a cost of service measure, but they are not certain how to calculate it.

Submission Processing Measures

As shown in table 5, all 11 of submission processing's performance measures have many of the attributes of successful performance measures. However, as summarized in this section, we identified several opportunities for improvement, especially in the area of reliability. Table 9 in appendix II has more detailed information about each submission processing measure, including any weaknesses we identified and any recommendations for improvement.

Table 5: Overview of Our Assessment of Submission Processing Measures

Submission processing measures	Submission processing attributes								
	Linkage[a]	Clarity	Measurable target	Objectivity	Reliability	Core program activities[b]	Limited overlap	Government-wide priorities	Balance
Individual 1040 series returns filed (paper)		✔	✔	✔	✔	✔	✔	✔	This attribute applies to the overall suite of measures rather than to the measures individually.[c]
Number of individual refunds issued (paper)		✔	✔	✔	✔	✔	✔	✔	
Employee satisfaction		✔	✔	✔	✔	✔	✔	✔	
Refund timeliness–individual (paper)		✔	✔	✔		✔	✔	✔	
Notice error rate		✔	✔			✔	✔	✔	
Refund error rate-individual (paper)		✔	✔	✔		✔	✔	✔	
Letter error rate		✔	✔			✔	✔	✔	
Deposit timeliness (paper)		✔	✔	✔	✔	✔	✔	✔	
Deposit error rate		✔	✔			✔	✔	✔	
Refund interest paid (per $1 million of refunds)		✔	✔	✔	✔	✔	✔	✔	
Submission processing productivity			✔	✔	✔	✔	✔	✔	

Note: A check mark denotes that the measure has the attribute.

[a]We were unable to verify the linkages between goals and measures because of insufficient documentation.

[b]Core program activities of submission processing are to efficiently and accurately process returns, remittances, and refunds and issue notices and letters.

[c]Submission processing measures cover various governmentwide priorities, such as efficiency, timeliness, and accuracy; however, submission processing's measures did not include a measure for customer satisfaction or for showing how much it costs to process the average return.

Source: GAO analysis.

Alignment between IRS's Goals and Submission Processing Measures Is Uncertain

No formal documentation exists to show how submission processing's 11 measures are aligned with IRS's mission, its agencywide goals, and its operating division goals. Despite this lack of formal documentation, submission processing officials said, and we generally concur, that some linkage does exist. Without complete documentation, however, we could not verify all the linkages. Submission processing officials stated that staff and managers are aware of the link between measures and goals because the submission processing organization has taken action to help ensure that staff understand the measures and their role in supporting IRS's overall mission and strategic and operating goals. For example, according to submission processing officials, they visited all eight W&I processing centers in 2001 to talk directly with staff and managers about the importance of balanced performance measures in ensuring that IRS meets its goals. Complete documentation of the linkages between goals and measures could further enhance understanding of those goals and measures with managers and staff.

Submission Processing Measures Have Clarity, with One Exception

All but one of the submission processing measures have clarity and provide information to enable executives, other managers, and outside stakeholders to properly assess performance against goals. The one exception is the productivity measure.

Managers in different processing centers told us that they did not use the productivity measure to provide them with performance information or to help them assess performance because, among other things, the measure does not provide specific information about their unit's or center's performance or their contribution to overall productivity. This is because the measure, as designed, is a compilation of different types of work IRS performs in processing returns, remittances, and refunds and issuing notices and letters. As a result, unit managers used different productivity measures specific to their own processes to help them identify how to increase their area's productivity. However, according to IRS officials, the productivity measure is useful and provides adequate information to some IRS executives.

From our perspective, although the productivity measure may be meaningful to executives, the fact that field managers use other measures and profess not to understand the current productivity measure indicates that the current measure does not provide those managers with useful information that would alert them to problems and help them respond when problems arise. In addition, because the measure is calculated by compiling and weighting different types of processing work per staff year expended, it may be too confusing to be useful to outside stakeholders, such as Congress.

All Submission Processing Measures Have Targets and Most Are Objective

All 11 of submission processing's measures have measurable targets and most are objective (i.e., reasonably free of significant bias or manipulation). For example, the "notice error rate" had a target of 8.1 percent for fiscal year 2001. The "deposit timeliness" measure appears to be objective, for example, because the Integrated Submission and Remittance Processing System[38] automatically calculates data on which the measure is based. However, the "notice error rate" and "letter error rate" measures are not objective because the coding required as part of data collection by individual reviewers is subject to much interpretation that could systematically bias the results of the measures. In October

[38]The Integrated Submission and Remittance Processing System is the system IRS uses to process tax returns and remittances.

2002, the Treasury Inspector General for Tax Administration reported, based on a review at two processing centers, that the "deposit error rate" measure was not objective, because the associated sampling plan was not consistently implemented.[39] The Treasury Inspector General for Tax Administration recommended that IRS take steps to ensure consistent implementation, and IRS reported that steps have been taken.

Five Submission Processing Measures Lack Reliability

Five measures are subject to consistency problems that affect the reliability of the measures. Those measures are "refund timeliness—individual (paper)," "notice error rate," "refund error rate," "letter error rate," and "deposit error rate." Specifically, the five measures are based on a data collection process, which according to the Director of Submission Processing, involves about 80 staff who identify, interpret, and analyze errors at the eight W&I processing centers. The "notice error rate" and "letter error rate" measures also involve coding that is subject to further interpretation.

Submission processing managers recognized that staff inconsistently coded notice and letter errors during the 2001 filing season. Neither IRS nor we know the extent to which such inconsistencies exist because no routine studies are done to validate the accuracy of data collection. Reliability and credibility increase when such studies are done. Submission processing initiated studies beginning in June 2001 to improve reliability, but has not established any improvement goals.

Submission Processing Measures Cover Core Program Activities without Overlap

Each of submission processing's measures directly pertains to one of the core program activities of submission processing's business operations—timely, efficiently, and accurately processing returns, remittances, and refunds and issuing notices and letters—without redundancy or overlap. For example, the "refund error rate—individual (paper)" measure directly pertains to one of submission processing's core program activities, processing refunds, and does not overlap with any of the other 11 measures.

Unlike the other three program areas we reviewed, submission processing has two customers—taxpayers, to whom IRS issues refunds and sends notices, and the Department of the Treasury, for which IRS deposits

[39]Treasury Inspector General for Tax Administration, *The Internal Revenue Service Needs to Improve Oversight of Remittance Processing Operations*, 2003-40-002 (Washington, D.C.: Oct. 7, 2002).

remittances. Therefore, for some measures, such as "refund timeliness," IRS views taxpayers as the customer, while for other measures, such as "deposit timeliness," IRS views Treasury as the customer. Submission processing officials believe that this dual-customer perspective provides a complete view of their operations and the measures cover all aspects of their operations while still being limited to a manageable number.

Submission Processing Measures Cover Various Governmentwide Priorities, but Are Not Fully Balanced

Submission processing's measures cover various governmentwide priorities, such as efficiency, timeliness, and accuracy. However, at the time of our review, submission processing measures lacked balance because they did not include a measure for customer satisfaction or a measure showing how much it costs to process a return.

Although submission processing officials believe that some existing measures, such as "notice error rate" and "refund timeliness," provide information related to the customer's experience, they recognize that directly obtaining customers' perspectives would be more accurate than assuming their experience based on such measures. Thus, submission processing is obtaining customer satisfaction information as part of IRS's corporate customer satisfaction survey, which IRS expects will be available by the 2003 filing season.

Similar to the other three program areas, submission processing does not have a cost of service measure.[40] Among other things, not having a cost of service measure affects IRS's ability to adequately compare different types of processing, such as paper versus electronic. In our view, because IRS does not take into account the cost to process a particular type of return, managers cannot fully understand the effectiveness of their unit.

Conclusions

Because the filing season affects so many taxpayers, IRS's performance is important. Having successful performance measures that demonstrate results, are limited to the vital few, cover multiple program priorities, and provide useful information to decision makers will help IRS management and stakeholders, such as Congress, make decisions about how to fund and improve return processing and assistance to taxpayers.

[40]Submission processing did have some data related to the average direct labor cost to process some paper returns in 1999.

Despite the challenge of developing a set of 53 measures that satisfy our criteria, IRS has made significant progress. As developed to date, the measures satisfy many of our nine attributes for successful performance measures. For example, in all four of the program areas we reviewed, most measures covered the core activities of each program and had targets in place. IRS also has several on-going improvement initiatives, such as the effort to redesign all aspects of its telephone assistance quality measures.

Although the measures satisfied many of the nine attributes, our evaluation also showed that they do not have all the characteristics of successful performance measures. The most significant weaknesses include (1) the inability of some measures to provide clear information to decision makers about program performance, (2) data collection methods that hamper objectivity and reliability, and (3) measures to cover governmentwide priorities that are missing from the Strategy and Program Plan. Although such weaknesses do not mean that the measures are not useful, IRS risks basing program and resource allocation decisions on inadequate or incomplete information and is less accountable until the weaknesses are addressed.

Correcting these weaknesses is important in order to (1) create a results-oriented environment that demonstrates and tracks how IRS's programs and activities contribute to achieving its mission and strategic goals, (2) avoid creating an excess of data that could obscure key information needed to identify problem areas and assess goal achievement, (3) form a balanced environment that takes the core program activities of the program into account, and (4) provide managers and other stakeholders with critical information on which to base their decisions.

Recommendations for Executive Action

We recommend that the Commissioner of Internal Revenue direct the appropriate officials to do the following:

Take steps to ensure that agencywide goals clearly align with operating division goals and performance measures for each of the four areas reviewed. Specifically, (1) clearly document the relationship among agencywide goals, operating division goals, and performance measures (the other three program areas may want to consider developing a template similar to the one field assistance developed, shown in figure 4) and (2) ensure that the relationship among goals and measures is communicated to staff at all levels of the organization.

Make the name and definition of several field assistance measures (i.e., "geographic coverage," "return preparation contacts," " return preparation

units," "TACs total contacts," "forms contacts," "tax law contacts," "account contacts," "other contacts," "tax law accuracy," "accounts/notices accuracy," and "return preparation accuracy") more clear to indicate what is and is not included in the formula.

As discussed in the body of this report and in appendix II, modify the formulas used to compute various measures to improve clarity. If formulas cannot be implemented in time for the next issuance of the Strategy and Program Plan, then modify the name and definition of the following measures so it is clearer what is or is not included in the measure.

- Remove automated calls from the formula for the "CSR level of service" measure.
- Revise the "CSR response level" measure to include calls from taxpayers who tried to reach a CSR but did not, such as those who (1) hung-up while waiting to speak to a CSR, (2) were provided access only to automated services and hung up, and (3) received a busy signal.
- Analyze and use new or existing data to determine why calls are transferred and use the data to revise the "CSR services provided" measure so that it only reflects transferred calls in which the caller received help from more than one CSR (i.e., exclude calls in which a CSR simply transferred the call and did not provide service.)
- Either discontinue use of the "number of IRS digital daily Web site hits" measure or revise the way "hits" are calculated so that the measure more accurately reflects usage.
- Revise field assistance's "geographic coverage" measure by ensuring that the formula better reflects (1) the various types of field assistance facilities, including alternate sites and kiosks; (2) the types of services provided by each facility; and (3) the facility's operating hours.
- Revise submission processing's "productivity" measure so it provides more meaningful information to users.

Refrain from making changes to official targets, such as electronic filing and assistance did in fiscal year 2001, unless extenuating circumstances arise. Disclose any extenuating circumstances in the Strategy and Program Plan and other key documents.

Modify procedures for the toll-free customer satisfaction survey, possibly by requiring that administrators listen to the entire call, to better ensure that administrators (1) notify CSRs that their call was selected for the survey as close to the end of a call as possible and (2) can accurately answer the questions they are responsible for on the survey.

Implement annual effectiveness studies to validate the accuracy of the data collection methods used for the five telephone measures ("toll-free tax law quality," "toll-free accounts quality," "toll-free tax law correct response rate," "toll-free account correct response rate," and "toll-free timeliness") subject to potential consistency problems. The studies could determine the extent to which variation exists in collecting data and recognize the associated impact on the affected measures. For those measures, and for the five submission processing measures that already have effectiveness studies in place ("refund timeliness-individual (paper)," "notice error rate," "refund error rate—individual (paper)," "letter error rate," and "deposit error rate"), IRS should establish goals for improving consistency, as needed.

Ensure that plans to remove overlapping measures in telephone and field assistance are implemented.

As discussed in the body of this report, include the following missing measures in the Strategy and Program Plan in order to better cover governmentwide priorities and achieve balance.

- In the spirit of provisions in the Chief Financial Officers Act of 1990 and Financial Accounting Standards Number 4, develop a cost of services measure using the best information currently available for each of the four areas discussed in this report, recognizing data limitations as prescribed by GPRA. In doing so, adhere to guidance, such as Office of Management and Budget Circular A-76, and consider seeking outside counsel to determine best or industry practices.
- Given the importance of automated telephone assistance, develop a customer satisfaction survey and measure for automated assistance.
- Put the "automated completion rate" measure back in the Strategy and Program Plan after revising the formula so that calls for recorded tax law information are not counted as completed when taxpayers hang up before receiving service.
- Add one or more quality measures to electronic filing and assistance's suite of measures in the Strategy and Program Plan. Possible measures include "processing accuracy," "refund timeliness, electronically filed," and "number of electronic returns rejected."
- Re-implement field assistance's timeliness measure.
- Develop a measure that provides information about field assistance's efficiency.

Agency Comments and Our Evaluation

The Commissioner of Internal Revenue provided written comments on a draft of this report in a letter dated November 1, 2002, which is reprinted in appendix III. The Commissioner was pleased to see that many of the measures had the attributes for successful performance and agreed that others presented opportunities for further refinement. He stated that the report was objective and balanced and that our observation of the on-going nature of the performance measurement process was on point. Furthermore, he noted that the attributes we developed can be used as a checklist when performance measures are developed in the future.

Of our 18 recommendations, IRS

- agreed with 10 and cited planned corrective actions that were responsive to those recommendations;
- cited actions taken or planned in response to 2 that did not fully address our concerns; and
- disagreed with 6.

The following discussion focuses on the recommendations with which IRS disagreed or for which we believe additional action is necessary to address our concerns.

In response to our recommendation about clarifying the name and definition of several field assistance measures, IRS said that the recently updated data dictionary addressed our concerns. We reviewed the updated data dictionary. The modifications are substantial and provide significant additional information about the measures. However, the definitions remain unclear. Specifically, the definitions should either define a taxpayer assistance center or state whether or not alternate sites, such as kiosks and mobile sites, are included.

IRS did not agree that automated calls should be removed from the formula for the "CSR level of service" measure. IRS said that including the count of callers who choose an automated service while waiting for CSR service is appropriate. IRS's response does not accurately characterize all the calls answered by automation that are included in the "CSR level of service" measure. Rather than choosing an automated service while waiting for a CSR, some callers complete an automated service after hearing an announcement that, due to high call volume, only automated services are available—a choice is not involved. We believe that the "CSR level of service" measure, because of its name and the way it is calculated, could be misleading and might misrepresent taxpayers' access to CSR's. For example, increasing the percentage of calls served through

automation because a CSR was not available—meaning that CSR's were actually more difficult to reach—would improve the "CSR level of service" measure, thus giving the impression that access to CSR's had improved when it had actually gotten worse. Calls answered through automation, regardless of the type of assistance (CSR or automation) the caller was originally seeking, should be reflected in an automated-level-of-service measure, such as "automated service completion rate."

IRS did not agree that it should modify the "CSR response level" measure to include calls in which the caller hung up before receiving service or got a busy signal. IRS said that altering the measure would deviate from industry standards and hinder IRS's ability to gauge success in meeting this "world class service" goal. We support IRS's efforts to gauge its progress toward providing world class customer service by telephone. However, IRS's use of the same telephone wait-time measure used by others may actually hinder a meaningful comparison of IRS with industry leaders. The "CSR response level" measure shows, for the callers who reached a CSR, the percentage that waited 30 seconds or less. According to IRS officials, when taxpayers call IRS attempting to reach a CSR, they are much less likely to reach one than when they call a recognized telephone service leader (i.e., callers to IRS are more likely to hang up while waiting to speak to a CSR, hang up after being given access to only automated service because a CSR is not available, or receive a busy signal). Therefore, when the "CSR response level" measure (which excludes these hang-ups and busy signals) is used by IRS, the measure may represent the experience of a significantly smaller percentage of the total callers that attempted to reach an a CSR than when the same measure is used by industry leaders, thus potentially overstating the ease with which callers reached IRS CSR's. Data we obtained from IRS suggest that there were about an equal number of hang-ups and busy signals as calls answered in this measure in 2001.

In response to our recommendation about implementing annual studies to validate the accuracy of various data collection methods and establishing goals for improving consistency, IRS said that it (1) has an ongoing process to ensure proper administration of the collection methods for the telephone measures cited in our recommendation, (2) does not agree that an annual independent review by non-CQRS analysts is merited, and (3) does not agree that it should incorporate consistency improvement goals in the Strategy and Program Plan process. As we noted in our report, telephone assistance's CQRS has some controls in place to monitor consistency. However, we believe that reliability and credibility increase when performance data are checked or tested for significant errors, which

IRS currently does not do. We did not recommend that non-CQRS analysts do these reviews; who does the reviews is for IRS to decide. Also, we recognized in our report that submission processing has an on-going process to verify consistency and that it has found problems. Because that review process has found some problems, we believe that establishing goals for improving consistency in submission processing is warranted. Because telephone assistance does not have a review process in place, we do not know whether improvement goals are needed, but noted that they could be. We did not recommend that these goals become a part of the Strategy and Program Plan process. Instead, these goals should become part of the review process and be made known to staff who are performing the work.

IRS disagreed with our recommendation that it put the "automated completion rate" measure back in the Strategy and Program Plan. Instead, IRS said it would continue to track and monitor that rate as a diagnostic measure. IRS told us that its decision is based on the fact that data on automated calls are not good enough to merit the attention the measure would have at the Strategy and Program Plan level. We recognize that there are data weaknesses with this measure. That is why our recommendation calls for IRS to revise the formula before returning the measure to the Strategy and Program Plan. Because serving more callers through automation is important to IRS's strategy for improving taxpayer service, we believe that IRS needs a measure of the level of service provided by automation in its Strategy and Program Plan to balance its measure of the level of service provided by CSRs. Other than counts of the number of calls served, IRS has no measure of its effectiveness in serving taxpayers through automation. Without such a measure, IRS risks poorly serving the increasing number of taxpayers being served through automation while possibly improving access for a declining number of callers who need to speak with a CSR.

IRS does not believe that adding one or more quality measures to electronic filing and assistance's suite of measures in the Strategy and Program Plan would enhance the electronic filing program. It noted that it tracks the quality of electronic filing outside the Strategy and Program Plan and that quality has been consistently high. We recognize that electronic filing and assistance tracks quality outside the Strategy and Program Plan. However, we disagree with IRS's position that adding quality measures to that plan would not enhance the program. According to IRS officials, measures in the Strategy and Program Plan are the highest, most comprehensive level of measures for which they are accountable. In addition, many of those measures are made available to outside

stakeholders. By not elevating these measures of quality to the Strategy and Program Plan, electronic filing and assistance risks not being held to any quality standards. Furthermore, not having quality measures hampers balance among electronic filing and assistance's suite of measures and is not consistent with IRS's balanced measurement program or the intent of IRS's Restructuring and Reform Act of 1998.

IRS disagreed with our recommendation that it re-implement field assistance's timeliness measure. IRS said that although timeliness goals are important in providing service to taxpayers, they are detrimental to quality service because field assistance employees tend to rush customers when traffic is high. This position is inconsistent with IRS's balanced measurement program and the intent of IRS's Restructuring and Reform Act of 1998. Although the accuracy of assistance is an important measure of quality, the timeliness of that assistance is also an important and balancing aspect of quality. Without this balancing emphasis, staff could theoretically take excessive time providing quality tax law assistance to a few taxpayers regardless of the impact on the wait-time for other taxpayers. We agree that Q-Matic is the best source of this information and support IRS's plans to implement it nationwide. IRS also stated that it could use feedback from its customer satisfaction surveys to obtain information about the "promptness of service." As we noted in our report, problems arose in the manner with which the feedback was provided from the vendor and the vendor had stopped providing feedback to site managers until the problems could be resolved. Even when those problems are resolved, a timeliness measure based on actual IRS data versus taxpayers' perceptions would be meaningful.

Regarding our recommendation about implementing an efficiency measure in field assistance, IRS said that it will be testing a system for use as a "diagnostic tool" to monitor and evaluate the strengths and weaknesses of various productivity measures. However, IRS's response was silent as to whether or when it would establish a field assistance productivity measure. Maintaining and enhancing organizational productivity is a fundamental agency management responsibility. The extent to which IRS's field assistance organization is meeting this basic responsibility needs to be visible to IRS, Treasury, and congressional stakeholders in the form of an organizational performance measure, rather than a "diagnostic tool," which is generally visible only to IRS managers.

We are sending copies of this report to the Chairmen and Ranking Minority Members of the Senate Committee on Finance and the House

Committee on Ways and Means and the Ranking Minority Member of this Subcommittee. We are also sending copies to the Secretary of the Treasury; the Commissioner of Internal Revenue; the Director, Office of Management and Budget; and other interested parties. We will make copies available to others on request. In addition, the report will be available at no charge on the GAO Web site at http://www.gao.gov.

This report was prepared under the direction of David J. Attianese, Assistant Director. Other major contributors are acknowledged in appendix IV. If you have any questions about this report, contact Mr. Attianese or me on (202) 512-9110.

Sincerely yours,

James R. White
Director, Tax Issues

Appendix I: Expanded Explanation of Our Attributes and Methodology for Assessing IRS's Performance Measures

Performance goals and measures that successfully address important and varied aspects of program performance are key to a results-oriented, balanced work environment. Measuring performance allows organizations to track the progress they are making toward their goals and gives managers critical information on which to base decisions for improving their programs. Organizations need to have performance measures that (1) demonstrate results, (2) are limited to the vital few, (3) cover multiple program priorities, and (4) provide useful information for decision making in order to track how their programs and activities can contribute to attaining the organization's goals and mission. These four characteristics are important to accurately reveal the strengths and weaknesses of a program since measures are often the key motivators of performance and goal achievement.

For use as criteria to determine whether the Internal Revenue Service's (IRS) performance measures in four key program areas—telephone assistance, electronic filing and assistance, field assistance, and submission processing—demonstrate results, are limited to the vital few, cover multiple program priorities, and are useful in decision making, we developed nine attributes of performance goals and measures based on previously established GAO criteria. In addition, we considered key legislation, such as the Government Performance and Results Act of 1993 (GPRA) and the IRS Restructuring and Reform Act of 1998, and performance management literature cited in the bibliography and related products sections at the end of this report. Our nine attributes may not cover all the attributes of successful performance measures; however, we believe these are some of the most important. We shared these attributes with IRS officials responsible for performance measurement issues, such as the Acting Director of the Organizational Performance Division; and several officials in the Wage and Investment (W&I) operating division, such as the Director of Strategy and Finance; the Chief of Planning and Analysis; the Director of Customer Account Services; and the Director of Field Assistance. These officials generally agreed with the relevance of the attributes and our review approach.

We applied these attributes to the 53 filing season measures in W&I's fiscal year 2001-2003 Strategy and Program Plan in a systematic manner, but some judgment was required. To ensure consistency and reliability in our application of the attributes, we had one staff person responsible for each of the four areas. That staff person prepared the initial analysis and at least two other staff reviewed those detailed results. Several staff reviewed the results for all four areas. Inherently, the attributes described below are not weighted equally. Weaknesses identified in a particular attribute do not,

in and of themselves, mean that a measure is ineffective or meaningless. Instead, weaknesses identified should be considered areas for further refinement.

Detailed information on each attribute, including an explanation, examples, and the methodology we used to assess that attribute with respect to the measures covered by our review, follows.

Attributes of Successful Performance Measures

1. Is there a relationship between the performance goals and measures and an agency's goals and mission? *(Referred to as "linkage")*

Explanation: Performance goals and measures should align with an agency's goals and mission. A cascading or hierarchal linkage moving from top management down to the operational level is important in setting goals agencywide, and the linkage from the operational level to the agency level provides managers and staff throughout an agency with a road map that (1) shows how their day-to-day activities contribute to attaining agencywide goals and mission and (2) helps define strategies for achieving strategic and annual performance goals. As agencies develop annual performance goals as envisioned by GPRA, they can serve as a bridge that links long-term goals to agencies' daily operations. For example, an annual goal that is linked to a program and also to a long-term goal can be used both to (1) hold agencies and program offices accountable for achieving those goals and (2) assess the reasonableness and appropriateness of those goals for the agency as a whole. In addition, annual performance planning can be used to better define strategies for achieving strategic and annual performance goals.

Linkages between goals and measures are most effective when they are clearly communicated to all staff within an agency so that everyone understands what the organization is trying to achieve and the goals it seeks to reach. Communicating goals and their associated measures is a continuous process and supports the basis for everything the agency does each day. Communication creates a "line of sight" throughout an agency so that everyone understands what the organization is trying to achieve and the goals it seeks to reach.

Example: Submission processing's "notice error rate" measure determines the percentage of incorrect notices issued to taxpayers by submission processing employees. The target set for this measure in 2001 was 8.1 percent. This measure could be used to support the "notice redesign" improvement project as well as the operational priority to "prioritize

notices and monitor and control notice issuance." It also is used to support one of W&I's goals— "to meet taxpayer demands for timely, accurate, and efficient services." This W&I strategy aligns with IRS's strategic goal, "top quality service to all taxpayers through fair and uniform application of the law," which in turn, supports IRS's mission to "provide America's taxpayers top quality service by helping them understand and meet their tax responsibilities and by applying the tax law with integrity and fairness to all."

Methodology: We compared IRS's measures with its targets, improvement projects, operational priorities, operating division goals, and agencywide goals and mission as documented in the Strategy and Program Plan. We also interviewed operational/unit managers and managers responsible for the Strategy and Program Plan about linkages and reviewed training materials.

2. Are the performance measures clearly stated? *(Referred to as "clarity")*

Explanation: A measure has clarity when it is clearly stated and the name and definition are consistent with the methodology used for calculating the measure. A measure that is not clearly stated (i.e., contains extraneous or omits key data elements) or that has a name or definition that is inconsistent with how it is calculated can confuse users and could cause managers or other stakeholders to think that performance was better or worse than it actually was.

Example: Telephone assistance's "average handle time" measure shows the average number of seconds Customer Service Representatives (CSRs) spent assisting callers. Its definition and formula are consistent with the name of the measure and clearly note that the measure includes talk and hold times and the time a CSR spends on work related to a call after the call is terminated.

Methodology: We compared the name of the measure, the written definition of the measure, and the formula or methodology for computing the measure. In several instances, we discussed certain components of the definition and formula with IRS officials to better understand its meaning and purpose. For example, we discussed components of telephone assistance's quality measures with staff in Customer Account Services, and

staff in the Centralized Quality Review Site. We also reviewed on-line information available to field assistance managers from the Queuing Management System (Q-Matic).[1] We spoke to managers at different levels within each of the four areas we reviewed and asked them about the information they received and how they used it. In addition, we used some of the results of a random telephone survey of managers we conducted in 2001 at 84 of IRS's 413 Taxpayer Assistance Centers (TAC) to solicit their views on the services provided at those offices.

3. Do the performance measures have targets, thus allowing for easier comparison with actual performance? *(Referred to as "measurable target")*

Explanation: Where appropriate, performance goals and measures should have quantifiable, numerical targets or other measurable values. Numerical targets or other measurable values facilitate future assessments of whether overall goals and objectives were achieved because comparisons can be easily made between projected performance and actual results. Some goals are self-measuring (i.e., they are expressed objectively and are quantifiable) and therefore do not require additional measures to assess progress. When goals are not self-measuring, performance measures should translate those goals into observable conditions that determine what data to collect to learn whether progress was made toward achieving goals. The measures should have a clearly apparent or commonly accepted relation to the intended performance or have been shown to be reasonable predictors of desired behaviors or events. If a goal cannot be expressed in an objective, specific, and measurable form, GPRA allows the Office of Management and Budget to authorize agencies to develop alternative forms of measurement.[2]

[1]Q-Matic is an automated tracking and reporting system that is expected to more efficiently monitor customer traffic flow and wait times and eliminate staff time completing Form 5311. Of about 420 TACs, 123 had Q-Matic as of June 2002.

[2]An alternative form of measurement may be either (1) separate, descriptive statements of a minimally effective program or (2) a successful program, expressed with sufficient precision and in such terms that would allow for an accurate, independent determination to be made of how actual performance compares with the goals stated. An example would be the polio vaccine and how its value to society is judged by experts through a peer review.

Example: Electronic filing and assistance's "percent of individual returns electronically filed" had a numerical target of 31 percent in fiscal year 2001.

Methodology: We determined that a goal or measure had a measurable target when expected performance could be compared with actual results, and in general, was not changed during the measurement period. Each of the measures we reviewed was listed in the Strategy and Program Plan, which provides projections or targets for the current and two subsequent fiscal years. We verified that the target was measurable. When the Strategy and Program Plan did not show a target, we contacted appropriate IRS officials to determine why.

4. Are the performance goals and measures objective? *(Referred to as "objectivity")*

Explanation: To the greatest extent possible, goals and measures should be reasonably free of significant bias or manipulation that would distort the accurate assessment of performance. They should not allow subjective considerations or judgments to dominate the outcome of the measurement. To be objective, performance goals and measures should indicate specifically what is to be observed, in which population or conditions, and in what timeframe and be free of opinion and judgment. Objectivity is important because it adds credibility to the performance goals and measures by ensuring that significant bias or manipulation will not distort the measure.

Example: The "customer satisfaction" measure for telephone assistance has the potential for bias and therefore may not be objective. Survey administrators are instructed to notify the CSR towards the end of the call that his or her call was selected to participate in the survey. A potential problem arises because administrators are not required to listen to the entire call, and it can be difficult to determine when the call is about to end. Therefore, if a CSR is notified prior to the end of the call that the call was selected for survey, the CSR could change behavior towards the taxpayer, thus affecting the results of the survey and the measure.

Methodology: We reviewed information in IRS guidance or procedures, data collection instruments, reports, and other documents. We held discussions about objectivity with various staff and officials, such as data owners and analysts, within each of the four areas we reviewed. Because our interviews raised questions about the objectivity of some measures for telephone assistance, we monitored some taxpayer calls and interviewed

an official from IRS's customer satisfaction survey contractor, Pacific Consulting Group.

5. To what extent do the performance goals and measures provide a reliable way to assess progress? *(Referred to as "reliability")*

Explanation: Reliability refers to whether measures are amenable to applying standard procedures for collecting data or calculating results so that they would be likely to produce the same results if applied repeatedly to the same situation. Errors can occur at various points in the collection, maintenance, processing, and reporting of data. Significant errors would affect conclusions about the extent to which performance goals have been achieved. Likewise, errors could cause the measure to report performance at either a higher or lower level than is actually being attained. Reliability is increased when verification and validation procedures, such as checking performance data for significant errors by formal evaluation or audit, exist.

Example: Field assistance's "return preparation contacts" measure tracks the total number of customers assisted with return preparation by IRS. This measure may not be reliable because it involves a significant amount of manual entry on Form 5311 (Field Assistance Activity Report) even at sites with the Q-Matic system. In addition to the potential for error associated with manual entry, the instructions for filing Form 5311 require that service time be recorded in whole hours, which can misconstrue actual service times and is less exact than the data in Q-Matic, which records service times in minutes.

Methodology: We looked for weaknesses in IRS's guidance or procedures, data collection instruments, reports, and other documents that might cause errors. We discussed potential weaknesses with various officials, such as account data analysts, within each of the four areas we reviewed. Because these efforts revealed the potential for errors in measuring telephone performance, we monitored employees preparing data collection instruments for assessing telephone quality and customer satisfaction in Atlanta. Likewise, we monitored field assistance staff helping taxpayers and reporting their time using both the automated Q-Matic system and Form 5311.

6. Do the performance measures sufficiently cover a program's core activities? *(Referred to as "core program activities")*

Explanation: Core program activities are the activities that an entity is expected to perform to support the intent of the program. Performance measures should be scoped to evaluate the core program activities. Limiting the number of performance measures to the core program activities will help identify performance that contributes to goal achievement. At the same time, however, there should be enough performance measures to ensure that managers have the information they need about performance in all the core program activities. Without such information, the possibility of achieving program goals is less likely.

Example: The core program activities for submission processing include (1) processing returns, (2) depositing remittances, (3) issuing refunds, and (4) sending out notices and letters. Each of submission processing's 11 measures correspond to one of those core activities. For example, the "number of individual 1040 series returns filed (paper)" measure corresponds to processing returns and the "letter error rate" measure corresponds with sending out notices and letters.

Methodology: We determined the core program activities of each of the four areas we reviewed based on IRS documentation and discussions with IRS officials. We reviewed the suite of performance measures for each of the four areas to determine whether measures existed that covered each core program activity. We determined whether any measures were missing or other pieces of information were needed to better manage programs by using judgment and questioning IRS officials. In addition, we reviewed the results of a questionnaire that we had used during a review of IRS's 2001 filing season to ask TAC managers about information needed to manage their program.

7. Does there appear to be limited overlap among the performance measures? *(Referred to as "limited overlap")*

Explanation: Measures overlap when the results of measures provide basically the same information. A measure that overlaps with another is unnecessary and does not benefit program management. Unnecessary or overlapping measures not only can cost money but also can cloud the bottom line in a results-oriented environment by making managers or other stakeholders sift through unnecessary or redundant information. Some measures, however, may overlap partially and provide stakeholders some new information. In those cases, management must make a judgment as to whether having the additional information is worth the cost and possible confusion it may cause.

Example: Telephone assistance's "toll-free average speed of answer" and "toll-free CSR response level" measures attempt to show how long a taxpayer waited before receiving assistance. The difference between the two measures is that the latter shows the percentage of taxpayers receiving assistance within 30 seconds while the former shows the average time taxpayers waited for service. These two measures are likely to be correlated and thus partially overlap. However, the amount of overlap between measures is management's discretion.

Methodology: Within each of the four areas we reviewed, we looked at the suite of measures and compared the measures' names and definitions. We also looked at the correlations between measures' results. When two measures seemed similar, we discussed the potential for overlap with IRS officials.

8. Does there appear to be a balance among the performance goals and measures, or is there an emphasis on one or two priorities at the expense of others? *(Referred to as "balance")*

Explanation: Balance exists when a suite of measures ensures that an organization's various priorities are covered. IRS considers its measures to be balanced when they address customer satisfaction, employee satisfaction, and business results (quality and quantity). Performance measurement efforts that overemphasize one or two priorities at the expense of others may skew the agency's performance and keep its managers from understanding the effectiveness of their programs in supporting IRS's overall mission and goals.

Example: Submission processing has an employee satisfaction measure and several business results measures, such as "deposit timeliness." As of October 2002, however, it had not fully implemented a customer satisfaction measure, which resulted in an unbalanced process that can overlook something as important as the customer's perspective.

Methodology: For each of the four areas, we ensured that a measure existed for each component. If measures did not exist for certain components, we contacted IRS officials to find out why and to see what plans IRS has to ensure balance in the future.

9. Does the program or activity have performance goals and measures that cover governmentwide priorities? *(Referred to as "governmentwide priorities")*

Explanation: Agencies should develop a range of related performance measures to address governmentwide priorities, such as quality, timeliness, efficiency, cost of service, and outcome. A range is important because most program activities require managers to balance these priorities among other demands. When complex program goals are broken down into a set of component quantifiable measures, it is important to ensure that the overall measurement of performance does not become biased because measures that assess some priorities but neglect others could place the program's success at risk.

Example: Electronic filing and assistance provides the capability for taxpayers to transact and communicate electronically with IRS. The 13 measures we reviewed included, for example, the number or percent of returns filed, the number of hits to or downloads from IRS's Web site, and employee and customer satisfaction. The Strategy and Program Plan did not have any measures on the program's quality or timeliness. Not having these measures means that management may not be sufficiently balancing competing demands.

Methodology: We analyzed the suite of measures in the Strategy and Program Plan for each of the four areas we reviewed. Based on discussions with IRS officials and our own judgment, we identified measures that appeared to be missing. We discussed those identified with IRS officials.

Appendix II: The 53 IRS Performance Measures Reviewed

The following four tables provide information on the 53 performance measures we reviewed in the four program areas within the Internal Revenue Service's (IRS) Wage and Investment (W&I) operating division that are critical to a successful filing season. Among other things, the tables show how each of the 53 measures matched up against the attributes in appendix I. The attributes not addressed in the tables are (1) "linkage," because sufficient documentation did not exist to validate linkages with any of the measures and (2) "balance," because that attribute does not apply to specific measures but, rather, to a program's entire suite of measures. When reviewing the suite of measures, we found some instances where additional measures are warranted; the additional measures are generally not cited in these tables.

Telephone Assistance Performance Measures

Of the 53 performance measures in our review, 15 are for telephone assistance.[1] Table 6 has information about each of the 15 telephone measures.

Table 6: Telephone Assistance Performance Measures

Measure name and definition[a]	FY 2001 target and actual	Weaknesses of measure and consequences	Recommendations
Total automated calls answered A count of all toll-free calls answered at telephone assistance centers by an automated system (e.g., Telephone Routing Interactive System) and Tele-Tax.[b]	Target: 85,000,000 calls answered Actual: 104,228,052 calls answered	Some overlap with automated completion rate measure. Both attempt to show how many automated calls were answered, but the automated completion rate tries to show the percentage that completed automated service successfully. Overlap could cloud the bottom line and obscure performance results.	See note 1 to the table.
Customer Service Representative (CSR) calls answered The count of all toll-free calls answered at telephone assistance centers.	Target: 31,500,000 calls answered Actual: 32,532,503 calls answered	Some overlap with CSR services provided measure. Both attempt to show how many calls CSRs answered, but CSR services provided tries to count calls requiring the help of more than one CSR as more than one call. Overlap could cloud the bottom line and obscure performance results.	See note 1 to the table.

[1]IRS deleted its "automated completion rate" measure in the 2002 Strategy and Program Plan and now has only 14 telephone measures. However, IRS still tracks this measure.

Measure name and definition[a]	FY 2001 target and actual	Weaknesses of measure and consequences	Recommendations
CSR level of service The relative success rate of taxpayers who call for toll-free services reaching a CSR.	Target: 55% Actual: 53.7%	Formula lacks clarity because it includes some automated calls, which overstates the number of calls answered by CSRs and thus the level of service being provided by CSRs.[c] Definition lacks clarity because it does not disclose inclusion of some automated calls, which could lead to misinterpreted results or a failure to take proper action to resolve performance problems.	Remove automated calls from the formula.
Toll-free customer satisfaction Customer's perception of service received, with a rating of "4" being the best.	Target: 3.45 average score Actual: 3.45 average score	Not clear because survey only applies to calls handled by CSRs. Satisfaction is not measured for calls handled by automation, which accounted for 76 percent of all calls in fiscal year 2001. Potential bias exists (not objective) because administrators are not required to listen to the entire call, (1) CSRs could be prematurely notified that their call was selected for the survey, thus changing their behavior towards the caller and affecting the results of the survey and (2) administrators may not be able to correctly answer certain questions on the survey, which could impair the accuracy of the data.	Develop a customer satisfaction survey for automated assistance. Modify procedures for the toll-free customer satisfaction survey, possibly by requiring that administrators listen to the entire call, to better ensure that administrators (1) notify CSRs that their call was selected for the survey as close to the end of a call as possible and (2) can accurately answer the questions they are responsible for on the survey.

Measure name and definition[a]	FY 2001 target and actual	Weaknesses of measure and consequences	Recommendations
Toll-free tax law quality[d] Evaluates the correctness of answers given by CSRs to callers with tax law inquiries as well as CSRs' conformance with IRS administrative procedures, such as whether the CSR gave his or her identification number to the taxpayer.	Target: 74% Actual: 75.21%	A reliability weakness exists because evaluations are based on judgments that are potentially inconsistent. No routine studies to determine effectiveness of procedures to ensure consistency of data collection. Possible inconsistencies affect the accuracy of the measure and conclusions about the extent to which performance goals have been achieved. Some overlap with toll-free tax law correct response rate. Both attempt to show the percentage of callers receiving accurate responses to tax law questions, but toll-free tax law quality includes CSR conformance with administrative procedures in computing that percentage. Overlap could cloud the bottom line and obscure performance results.	Implement annual effectiveness studies to validate the accuracy of data collection methods and establish goals for improving consistency, as needed. See note 1 to the table.
Toll-free accounts quality[e] Evaluates the correctness of answers given by CSRs to callers with account-related inquiries as well as CSRs' conformance with IRS administrative procedures, such as whether a CSR gave his or her identification number to the taxpayer.	Target: 67% Actual: 69.17%	A reliability weakness exists because evaluations are based on judgments that are potentially inconsistent. No routine studies to determine effectiveness of procedures to ensure consistency of data collection. Possible inconsistencies affect the accuracy of the measure and conclusions about the extent to which performance goals have been achieved. Some overlap with toll-free account correct response rate. Both attempt to show the percentage of callers receiving accurate responses to account questions, but toll-free accounts quality includes CSR conformance with administrative procedures in computing that percentage. Overlap could cloud the bottom line and obscure performance results.	Implement annual effectiveness studies to validate the accuracy of data collection methods and establish goals for improving consistency, as needed. See note 1 to the table.
Average handle time The average number of seconds CSRs spent assisting callers. It includes talk and hold times and the time a CSR spends on work related to a call after the call is terminated.	Target: not available Actual: 609 seconds	Target to be set upon completion of baseline data collection.[f]	None.

Measure name and definition[a]	FY 2001 target and actual	Weaknesses of measure and consequences	Recommendations
Automated completion rate The percentage of total callers who completed a selected automated service.	Target: not available Actual: not available	Formula lacks clarity because it assumes that all callers seeking recorded tax law information, including those who hang up before receiving service, received the information they needed, which could produce inaccurate or misleading results. Not clear because definition does not disclose the previously mentioned assumption, which could lead to misinterpreted results or a failure to take proper action to resolve performance problems. Measure removed from the Strategy and Program Plan; target not available. Some overlap with total automated calls answered. Both attempt to show how many automated calls were answered, but automated completion rate tries to show the percentage that completed an automated service successfully. Overlap could cloud the bottom line and obscure performance results.	Revise the measure so that calls for recorded tax law information are not counted as completed when callers hang up before receiving service. Put this measure back in the Strategy and Program Plan after revising the formula so that calls for recorded tax law information are not counted as completed when taxpayers hang up before receiving service. See note 1 to the table.
CSR services provided The count of all calls handled by CSRs.	Target: not available Actual: 35,799,122 calls answered	Not clear because definition does not disclose that IRS counts all calls transferred from one CSR to another as receiving an additional service, which could lead to misinterpreted results or a failure to take proper action to resolve performance problems. IRS does not have complete information on why calls were transferred. Thus, IRS cannot identify appropriate steps to reduce any inefficiency associated with transferred calls. Target to be set upon completion of baseline data collection[f] Some overlap with CSR calls answered. Both attempt to show how many calls CSRs answered, but CSR services provided tries to count calls requiring the help of more than one CSR as more than one call. Overlap could cloud the bottom line and obscure performance results.	Analyze and use new or existing data to determine why calls are transferred and use the data to revise the measure so that it only reflects transferred calls in which the caller received help from more than one CSR (i.e., exclude calls in which a CSR simply transferred the call and did not provide service). See note 1 to the table.

Measure name and definition[a]	FY 2001 target and actual	Weaknesses of measure and consequences	Recommendations
Toll-free tax law correct response rate[g] Evaluates the correctness of answers given by CSRs to callers with tax law inquiries.	Target: 81.6% Actual: 79.53%	A reliability weakness exists because evaluations are based on judgments that are potentially inconsistent. No routine studies to determine effectiveness of procedures to ensure consistency of data collection. Possible inconsistencies affect the accuracy of the measure and conclusions about the extent to which performance goals have been achieved. Some overlap with toll-free tax law quality. Both attempt to show the percentage of callers receiving accurate responses to tax law questions, but toll-free tax law quality includes CSR conformance to administrative procedures in computing that percentage. Overlap could cloud the bottom line and obscure performance results.	Implement annual effectiveness studies to validate the accuracy of data collection methods and establish goals for improving consistency, as needed. See note 1 to the table.
Toll-free account correct response rate[h] Evaluates the correctness of answers given by CSRs to callers with account-related inquiries.	Target: 90.8% Actual: 88.72%	A reliability weakness exists because evaluations are based on judgments that are potentially inconsistent. No routine studies to determine effectiveness of procedures to ensure consistency of data collection. Possible inconsistencies affect the accuracy of the measure and conclusions about the extent to which performance goals have been achieved. Some overlap with toll-free accounts quality. Both attempt to show the percentage of callers receiving accurate responses to account questions, but toll-free accounts quality includes CSR conformance with administrative procedures in computing that percentage. Overlap could cloud the bottom line and obscure performance results.	Implement annual effectiveness studies to validate the accuracy of the data collection methods and establish goals for improving consistency, as needed. See note 1 to the table.
Toll-free timeliness[i] The successful resolution of all issues resulting from the caller's first inquiry (telephone only).	Target: 82% Actual: 82.8%	A reliability weakness exists because evaluations are based on judgments that are potentially inconsistent. No routine studies to determine effectiveness of procedures to ensure consistency of data collection. Possible inconsistencies affect the accuracy of the measure and conclusions about the extent to which performance goals have been achieved.	Implement annual effectiveness studies to validate the accuracy of data collection methods and establish goals for improving consistency, as needed.

Measure name and definition[a]	FY 2001 target and actual	Weaknesses of measure and consequences	Recommendations
Toll-free employee satisfaction The percentage of survey participants that answered with a 4 or 5 (two highest scores possible) to the question "considering everything, how satisfied are you with your job?"	Target: 55% Actual: 46%	None observed.	None.
CSR response level The percentage of callers who started receiving service from a CSR within a specified period of time.	Target: 49% Actual: 40.8%	Not clear because formula does not include calls that received a busy signal or resulted in a hang-up before a CSR came on the line, and the definition does not disclose that exclusion. Performance may be overstated and the real customer experience not reflected. Some overlap with average speed of answer. Both attempt to show how long callers waited before receiving service, except that CSR response level shows the number of callers receiving service within 30 seconds. Overlap could cloud the bottom line and obscure performance results.	Revise measure to include calls from taxpayers who tried to reach a CSR but did not, such as those who (1) hung-up while waiting to speak to a CSR, (2) were provided access only to automated services and hung up, and (3) received a busy signal. See note 1 to the table.
Average speed of answer The average number of seconds callers waited in queue before receiving service from a CSR.	Target: not available Actual: 295 seconds	Target to be set upon completion of baseline data collection.[f] Some overlap with toll-free CSR response level. Both attempt to show how long callers waited before receiving service, except that CSR response level shows the number of callers receiving service within 30 seconds. Overlap could cloud the bottom line and obscure performance results.	See note 1 to the table.

Note 1: We identified this measure as having partial overlap with another measure. Telephone assistance officials generally agreed with our assessment and stated that some of these overlapping measures will be removed from future Strategy and Program Plans. The following recommendation applies to several measures as noted in the table: "ensure that plans to remove overlapping measures are implemented."

[a]The names of some measures have been modified slightly from the official names used by IRS for ease of reading and consistency purposes. For example, we replaced the word "assistor" with CSR. Also, the definitions of the measures listed in the table come from various IRS sources, including interviews.

[b]The Telephone Routing Interactive System is an interactive routes callers to CSRS or automated services and provides interactive services. Tele-Tax is a telephone system that provides automated services only.

[c]About 780,000 automated calls were included in the formula during the 2001 filing season. If they had not been included, the CSR level of service would have decreased by about 1 percentage point. The effect could be more significant in the future because IRS plans to increase the number of calls handled through automation.

[d]IRS plans to discontinue the "toll-free tax law quality" measure in fiscal year 2004.

[e]IRS plans to discontinue the "toll-free accounts quality" measure in fiscal year 2004.

[f]Although these measures did not have a measurable target in place, IRS is taking reasonable steps to develop a target.

[g]IRS changed the name of the "toll-free tax law correct response rate" measure to "customer accuracy for tax law inquiries" beginning in October 2002.

[h]IRS changed the name of the "toll-free account correct response rate" measure to "customer accuracy for account inquiries" beginning in October 2002.

[i]IRS discontinued the "toll-free timeliness" measure beginning in October 2002, and replaced it with a new "quality timeliness" measure.

Source: GAO comparison of IRS's December 13, 2000, July 25, 2001, and October 29, 2001, Strategy and Program Plans with the attributes in appendix I and an Embedded Quality Discussion Document (7/23/02), which discusses the changes IRS plans for its telephone assistance quality measures.

Electronic Filing and Assistance Performance Measures

Of the 53 performance measures in our review, 13 are for electronic filing and assistance.[2] Table 7 has information about each of the 13 measures.

Table 7: Electronic Filing and Assistance Performance Measures

Measure name and definition[a]	FY 2001 target and actual	Weaknesses of measure and consequences	Recommendations
Number of 1040 series returns electronically filed (millions) The number of Forms 1040, 1040A, and 1040EZ filed electronically.	Target: 40.0 Actual: 40.0	Target changed during filing season from 42.0 to 40.0. Changing the target in this instance was subjective in nature and resulted in an objectivity weakness as well. Some overlap with percent of individual returns electronically filed. Both measures show the extent of electronic filing by individuals—one in absolute numbers, the other as a percent of total filings. Overlap could cloud the bottom line and obscure performance results.	Refrain from making changes to official targets unless extenuating circumstances arise. Disclose any extenuating circumstances in the Strategy and Program Plan and other key documents. See note 1 to the table.
Number of business returns electronically filed (millions) The number of Forms 941, 1041, and 1065 filed electronically.	Target: 3.7 Actual: 1.66	None observed.	None.
Total number of electronically filed returns (millions). The number of Forms 1040, 1040A, 1040EZ, 941, 1041 and 1065 filed electronically.	Target: 43.7 Actual: 41.7	Target changed during filing season from 45.7 to 43.7. Changing the target in this instance was subjective in nature and resulted in an objectivity weakness as well.	Refrain from making changes to official targets unless extenuating circumstances arise. Disclose any extenuating circumstances in the Strategy and Program Plan and other key documents.

[2]IRS has since added three measures ("number of information returns filed by magnetic tape," "percent of information returns filed by magnetic tape," and "customer satisfaction-business") that were not part of our review. In addition, electronic filing and assistance is developing new performance measures and goals because it is in the midst of a major reorganization. When the reorganization is completed, electronic filing and assistance will no longer be responsible for all the operational programs for which it was responsible in 2001 and 2002. Electronic filing and assistance will remain responsible for strategic services, Internet development services, and development services. The IRS organizations assuming responsibility for electronic filing and assistance's operational programs will be responsible for the related performance measures and goals.

Measure name and definition[a]	FY 2001 target and actual	Weaknesses of measure and consequences	Recommendations
Number of information returns electronically filed (millions) The total number of information returns filed electronically. Includes Forms 1098, 1099, 5498, and W-2G and Schedules K-1. Excludes Forms W-2 and 1099-SSA/RRB received from the Social Security Administration.	Target: 334.0 Actual: 322.8	Some overlap with percent of information returns electronically filed. Both measures show the extent of electronic filing —one in absolute numbers, the other as a percent of total filings. Overlap could cloud the bottom line and obscure performance results.	See note 1 to table.
Percent of information returns electronically filed The percentage of total information returns filed electronically.	Target: 24.4% Actual: not available[b]	Some overlap with number of information returns electronically filed. Both measures show the extent of electronic filing —one in absolute numbers, the other as a percent of total filings. Overlap could cloud the bottom line and obscure performance results.	See note 1 to table.
Percent of individual returns electronically filed The percentage of total 1040 series tax returns (Forms 1040, 1040A, and 1040EZ) filed electronically.	Target: 31% Actual: 32%	Some overlap with number of 1040 series returns electronically filed. Both measures show the extent of electronic filing by individuals—one in absolute numbers, the other as a percent of total filings. Overlap could cloud the bottom line and obscure performance results.	See note 1 to table.
Number of payments received electronically (millions) All individual and all business tax payments made through the electronic federal tax payment system (EFTPS).	Target: 64.4 Actual: 53.8	Some overlap with percent of payments received electronically. Both measures show the extent to which payments are received electronically—one in absolute numbers, the other as a percent of total receipts. Overlap could cloud the bottom line and obscure performance results.	See note 1 to table.
Percent of payments received electronically The percentage of all individual and business tax payments made through EFTPS.	Target: 30% Actual: not available[b]	Some overlap with number of payments received electronically. Both measures show the extent to which payments are received electronically—one in absolute numbers, the other as a percent of total receipts. Overlap could cloud the bottom line and obscure performance results.	See note 1 to table.
Number of electronic funds withdrawals/credit card transactions (millions) The total number of credit card and direct debit payments processed through EFTPS.	Target: 1.0 Actual: 0.63	Some overlap with number and percent of payments received electronically. The payments covered by this measure are included in the universe of payments covered by the other two measures. Overlap could cloud the bottom line and obscure performance results.	See note 1 to table.
Number of IRS digital daily Web site hits (billions) The number of hits to IRS's Web site.	Target: 2.0 Actual: 2.3	Measure is not clear and lacks reliability because, for example, initial access counts as multiple hits and movement throughout the Web site will count as additional hits.	Either discontinue use of this measure or revise the way "hits" are calculated so that the measure more accurately reflects usage.

Measure name and definition[a]	FY 2001 target and actual	Weaknesses of measure and consequences	Recommendations
Number of downloads from "IRS .GOV" (millions) The total number of tax forms downloaded from IRS's Web site.	Target: 311 Actual: 309	None observed.	None.
Customer satisfaction – individual taxpayers The percentage of taxpayers who respond "very satisfied" with individual E-file products.	Target: 76% Actual: 83%	None observed.	None.
Employee satisfaction – Electronic filing and assistance The percentage of survey participants that answered with a 4 or 5 (two highest scores possible) to the question "considering everything, how satisfied are you with your job?"	Target: 66% Actual: 38%	None observed.	None.

Note: We identified this measure as having partial overlap with another measure. Electronic filing and assistance officials told us that each of the overlapping measures we identified provides additional information to managers. Determining whether or not to remove overlapping measures is management's discretion.

[a]The names of some measures have been modified slightly from the official names used by IRS for ease of reading and consistency purposes. The definitions of the measures listed in the table come from various IRS sources, including interviews.

[b]Despite setting a target, actual data were not available because electronic filing and assistance did not begin tracking the measure until 2002.

Source: GAO comparison of IRS's December 13, 2000, July 25, 2001, and October 29, 2001, Strategy and Program Plans with the attributes in appendix I.

Field Assistance Performance Measures

Of the 53 performance measures in our review, 14 are for field assistance. Table 8 has information about each of the 14 field assistance measures.

Table 8: Field Assistance Performance Measures

Measure name and definition[a]	FY 2001 target and actual	Weaknesses of measure and consequences	Recommendations
Customer satisfaction From surveys established in 1998, an index was created to represent overall customer satisfaction with field assistance services, with a "7" being the best.[b]	Target: 6.5 average score Actual: 6.4 average score	None identified.	None.
Return preparation contacts Total number of customers assisted with tax return preparation, including electronic and non-electronic tax return preparation at taxpayer assistance centers (TAC).	Target: 979,206 Actual: 1,009,387	Name, definition, and formula of measure are not clear. Significant manual data collection process impedes reliability because of the potential for errors and inconsistencies that could affect the accuracy of the measure and conclusions about the extent to which performance goals have been achieved. Some overlap with return preparation units measure. Both measures attempt to show number of services provided, but the contact measure takes the number of taxpayers served into account and the units measure counts the number of returns prepared for those taxpayers served. Overlap could cloud the bottom line and obscure performance results.	Make the name and/or definition of the measure more clear to indicate what is and is not included in the formula. See note 1 to the table. See note 2 to the table.

Measure name and definition[a]	FY 2001 target and actual	Weaknesses of measure and consequences	Recommendations
Geographic coverage Percentage of W&I taxpayer population with distinct characteristics, behaviors, and needs for face-to-face assistance within a 45-minute commuting distance from a TAC.	Target: 70% Actual: 74%	Name, definition, and formula of measure are not clear; uncertainties exist among IRS officials about what is and is not included in the measure. The formula does not include all facilities, which could lead to misinterpreted results or a failure to properly identify alternative facility types to resolve access problems. Because the formula does not include all facilities, it is difficult for decision makers to determine if, when, and where additional TACs are needed.	Make the name and/or definition of the measure more clear to indicate what is and is not included in the formula. Revise the formula to better reflect (1) the various types of field assistance facilities, including alternate sites and kiosks; (2) the types of services provided by each facility; and (3) the facility's operating hours.
Return preparation units Actual number of tax returns prepared, in whole or in part, in a TAC or alternative site. (Multiple returns may be prepared for a single customer.)	Target: not available Actual: not available	Name, definition, and formula of measure are not clear. Target to be set upon completion of data collection.[c] Significant manual data collection process impedes reliability because of the potential for errors and inconsistencies that could affect the accuracy of the measure and conclusions about the extent to which performance goals have been achieved. Some overlap with return preparation contacts. Both measures attempt to show number of services provided, but the contact measure takes the number of taxpayers served into account and the units measure counts the number of returns prepared for those taxpayer's served. Overlap could cloud the bottom line and obscure performance results.	Make the name and/or definition of the measure more clear to indicate what is and is not included in the formula. See note 1 to the table. See note 2 to the table.
TACs total contacts Total number of customers assisted, including number of customers assisted with tax return preparation, at TACs and alternate sites and via mobile services. All face-to-face, telephone, and correspondence contacts are included.	Target: 9,116,099 Actual: 9,681,330	Name, definition, and formula of measure are not clear. Significant manual data collection process impedes reliability because of the potential for errors and inconsistencies that could affect the accuracy of the measure and conclusions about the extent to which performance goals have been achieved.	Make the name and/or definition of the measure more clear to indicate what is and is not included in the formula. See note 1 to the table.

Measure name and definition[a]	FY 2001 target and actual	Weaknesses of measure and consequences	Recommendations
Forms contacts Total number of customers actually assisted by employees at TACs, alternate sites, and via mobile services by (1) providing forms from stock or (2) using a CD-ROM.	Target: 2,331,000 Actual: 2,388,039	Name, definition, and formula of measure are not clear. Significant manual data collection process impedes reliability because of the potential for errors and inconsistencies that could affect the accuracy of the measure and conclusions about the extent to which performance goals have been achieved.	Make the name and/or definition of the measure more clear to indicate what is and is not included in the formula. See note 1 to the table.
Tax law contacts Total number of customers assisted in TACs, alternate sites, and via mobile services with inquiries involving general tax law questions, non-account related IRS procedures, preparation or review of Forms W-7, Individual Taxpayer Identification Number documentation verification or rejection, a form request where probing requiring technical tax law training takes place, and assisting customers with audit reconsideration.	Target: not available Actual: 1,787,338	Name, definition, and formula of measure are not clear. Target to be set upon completion of data collection.[c] Significant manual data collection process impedes reliability because of the potential for errors and inconsistencies that could affect the accuracy of the measure and conclusions about the extent to which performance goals have been achieved.	Make the name and/or definition of the measure more clear to indicate what is and is not included in the formula. See note 1 to the table.
Account contacts Total number of customers assisted in TACs, alternate sites, and via mobile services with inquiries involving account related inquiries including math error notices, Integrated Data Retrieval System work, payments not attached to a tax return, CP2000 inquiries, Individual Taxpayer Identification Number issues requiring account research, the issuance of Form 809 receipts, and account related procedures.	Target: not available Actual: not available	Name, definition, and formula of measure are not clear. Target to be set upon completion of data collection.[c] Significant manual data collection process impedes reliability because of the potential for errors and inconsistencies that could affect the accuracy of the measure and conclusions about the extent to which performance goals have been achieved.	Make the name and/or definition of the measure more clear to indicate what is and is not included in the formula. See note 1 to the table.
Other contacts Total number of customers assisted in TACs, alternate sites, and via mobile services with Form 2063, U.S. Departing Alien Income Tax statement, date stamping tax returns when the customer is present, non-receipt or incorrect W-2 inquiries, general information such as Service Center address and directions to other agencies.	Target: 3,869,000 Actual: 4,496,566	Name, definition, and formula of measure are not clear. Significant manual data collection process impedes reliability because of the potential for errors and inconsistencies that could affect the accuracy of the measure and conclusions about the extent to which performance goals have been achieved.	Make the name and/or definition of the measure more clear to indicate what is and is not included in the formula. See note 1 to the table.

Measure name and definition[a]	FY 2001 target and actual	Weaknesses of measure and consequences	Recommendations
Tax law accuracy The quality of service provided to TAC customers. Specifically, the accuracy of responses concerning issues involving tax law.	Target: not available Actual: not available	Name, definition, and formula of measure are not clear. Target to be set upon completion of data collection.[c]	Make the name and/or definition of the measure more clear to indicate what is and is not included in the formula.
Accounts/notices accuracy The quality of service provided to TAC customers. Specifically, the accuracy of responses and/or IDRS transactions concerning issues involving account work and notices.	Target: not available Actual: not available	Name, definition, and formula of measure are not clear. Target to be set upon completion of data collection.[c]	Make the name and/or definition of the measure more clear to indicate what is and is not included in the formula.
Return preparation accuracy The quality of service provided to TAC customers. Specifically, the accuracy of tax returns prepared in a TAC.	Target: not available Actual: not available	Name, definition, and formula of measure are not clear. Target to be set upon completion of data collection.[c]	Make the name and/or definition of the measure more clear to indicate what is and is not included in the formula.
Employee satisfaction The percentage of survey participants that answered with a 4 or 5 (two highest scores possible) to the question "considering everything, how satisfied are you with your job."	Target: 62% Actual: 51%	None observed.	None.
Alternate contacts Total number of customers assisted at kiosks, mobile units, and alternate sites. It includes all face-to-face (including return preparation), telephone, and correspondence contacts.	Target: not available Actual: not available	Target to be set upon completion of data collection.[c] Significant manual data collection process impedes reliability because of the potential for errors and inconsistencies that could affect the accuracy of the measure and conclusions about the extent to which performance goals have been achieved.	See note 1 to the table.

Note 1: IRS expects to minimize this potential for errors and inconsistency by equipping all of its TACS with an on-line automated tracking and reporting system known as the Queuing Management System (Q-Matic). This system is expected, among other things, to more efficiently monitor customer traffic flow and eliminate staff time spent completing Form 5311. Because IRS is in the process of implementing Q-Matic, we are not making any recommendation.

Note 2: We identified this measure as having partial overlap with another measure. Field assistance officials agreed with our assessment and stated that they plan to remove the "return preparation contacts" measure from the Strategy and Program Plan. The following recommendation applies to two measures, as noted in the table: "ensure that plans to remove overlapping measures are implemented."

[a]The names of some measures have been modified slightly from the official names used by IRS for ease of reading and consistency purposes. The definitions of the measures listed in the table come from various IRS sources, including interviews.

[b]Field assistance implemented a new customer satisfaction survey in fiscal year 2002. The index was changed, and a rating of "5" is now best.

[c]Although these measures did not have a measurable target in place, IRS is taking reasonable steps to develop a target.

Source: GAO comparison of IRS's December 13, 2000, July 25, 2001, and October 29, 2001, Strategy and Program Plans with the attributes in appendix I.

Submission Processing Performance Measures

Of the 53 performance measures in our review, 11 are for submission processing.[3] Table 9 has information about each of the 11 submission processing performance measures.

Table 9: Submission Processing Performance Measures

Measure name and definition[a]	FY 2001 target and actual	Weaknesses of measure and consequences	Recommendations
Individual 1040 series returns filed (paper)[b] The number of Forms 1040, 1040A, and 1040EZ filed at the eight W&I submission processing centers.	Target: 87,869,000 Actual: 74,972,667	None observed.	None.
Number of individual refunds issued (paper)[b] The number of individual refunds issued by the eight W&I submission processing centers after the initial filing of a return.	Target: 48,000,000 Actual: 45,456,534	None observed.	None.
Employee satisfaction The percentage of survey participants that answered with a 4 or 5 (two highest scores possible) to the question "considering everything, how satisfied are you with your job."	Target: 60% Actual: 54%	None observed.	None.

[3] IRS is developing a measure of customer satisfaction for submission processing.

Measure name and definition[a]	FY 2001 target and actual	Weaknesses of measure and consequences	Recommendations
Refund timeliness – individual (paper)[b] The percentage of refunds issued to taxpayers within 40 days of the date IRS received the individual income tax return.	Target: 96.1% Actual: 96.75%	Potential reliability weakness because data collected manually and evaluations of data based on judgment. Possible inconsistencies affect the accuracy of the measure and conclusions about the extent to which performance goals have been achieved.	Based on the results of effectiveness studies, establish goals to improve consistency, as needed.
Notice error rate The percentage of incorrect submission processing master file notices issued to taxpayers (includes systemic errors).[c]	Target: 8.1% Actual: 14.84%	Potential reliability weakness because data collected manually and evaluations of data based on judgment. Possible inconsistencies affect the objectivity of the measure and conclusions about the extent to which performance goals have been achieved.	Based on the results of effectiveness studies, establish goals to improve consistency, as needed.
Refund error rate – individual (paper)[b] The percentage of refunds that have errors caused by IRS involving, for example, a person's name or refund amount (includes systemic errors).[c]	Target: 13.6% Actual: 9.75%	Potential reliability weakness because data collected manually and evaluations of data based on judgment. Possible inconsistencies affect the accuracy of the measure and conclusions about the extent to which performance goals have been achieved.	Based on the results of effectiveness studies, establish goals to improve consistency, as needed.
Letter error rate The percentage of letters with errors issued to taxpayers by submission processing employees (includes systemic errors).[c]	Target: 11.9% Actual: 13.10%	Potential reliability weakness because data collected manually and evaluations of data based on judgment. Possible inconsistencies affect the objectivity of the measure and conclusions about the extent to which performance goals have been achieved.	Based on the results of effectiveness studies, establish goals to improve consistency, as needed.
Deposit timeliness (paper)[b] Lost opportunity cost of money received by IRS but not deposited in the bank by the next day, per $1 billion of deposits, using a constant 8% annual interest rate.	Target: $746,712 Actual: $878,867	None observed.	None.

Measure name and definition[a]	FY 2001 target and actual	Weaknesses of measure and consequences	Recommendations
Deposit error rate The percentage of payments misapplied based on the taxpayer's intent.	Target: 4.9% Actual: not available[d]	Objectivity weakness because sampling plan not consistently implemented. Potential reliability weakness because data collected manually and evaluations of data based on judgment. Possible inconsistencies affect the accuracy of the measure and conclusions about the extent to which performance goals have been achieved.	See note 1 to the table. Based on the results of effectiveness studies, establish goals to improve consistency, as needed.
Refund interest paid (per $1 million of refunds) The amount of refund interest paid per $1 million of refunds issued.	Target: $112 Actual: $128.63	None observed.	None.
Submission processing productivity The weighted workload or work units processed per staff year expended.	Target: 28,787 Actual: 28,537	Not clear because (1) definition is not clearly stated, (2) managers do not understand their unit's contribution to the formula and (3) unit managers do not use the measure to assess performance.	Revise the measure so it provides more meaningful information to users.

Note 1: We are not making a recommendation regarding the objectivity weakness for the "deposit error rate" measure because the Treasury Inspector General for Tax Administration recommended that IRS take steps to ensure that the sampling plan is being implemented consistently, and IRS reported that steps have been taken.

[a]The names of some measures have been modified slightly from the official names used by IRS for ease of reading and consistency purposes. The definitions of the measures listed in the table come from various IRS sources, including interviews.

[b]"Paper" means that returns filed electronically (or their resulting refunds) are not included in the measure.

[c]A systemic error is an error caused by a computer programming error as opposed to an IRS employee.

[d]IRS could not provide actual data on this measure due to discrepancies in its data.

Source: GAO comparison of IRS's December 13, 2000, July 25, 2001, and October 29, 2001, Strategy and Program Plans with the attributes in appendix I.

Appendix III: Comments from the Internal Revenue Service

Note: GAO comments supplementing those in the report text appear at the end of this appendix.

DEPARTMENT OF THE TREASURY
INTERNAL REVENUE SERVICE
WASHINGTON, D.C. 20224

COMMISSIONER

November 1, 2002

Mr. James R. White
Director, Tax Issues
U.S. General Accounting Office
441 G Street, N.W.
Washington, D.C. 20548

Dear Mr. White:

I appreciate your recognition of the substantial progress we have made in implementing our balanced measures and our strategic planning process. We issued our first Strategy and Program Plan (SPP) and performance measures in Fiscal Year (FY) 2000, which I believe was great progress in a short period of time. We continue to gain experience and focus on the key attributes of performance measures as we use them in our day-to-day operations. Your observation that this is an ongoing process is exactly on point. The observations of your staff will benefit us as we continue to improve our performance measures. I believe your report is an insightful review of the measures we developed for use in FY 2001. We recently completed our SPP and the related performance measures for FYs 2003-2004. We will consider your suggestions as we review our current plan and develop plans for our next SPP cycle.

I am particularly impressed with the detailed definitions, explanations, and examples your staff developed for the nine attributes of successful performance measures. I believe the Wage and Investment (W&I) Division can use these standards as a helpful checklist when they develop future performance measures.

I also was pleased to note your observation that our measures had many of the attributes for successful performance. This indicates that we appropriately developed and properly targeted key performance measures. I also agree the measures that did not satisfy all of the attributes will give us opportunities for further refinement rather than invalidate their overall value. As you noted, we have several initiatives underway to continue improving these measures. Overall, your report is objective and balanced.

I want to share some additional points for your consideration:

- Although the filing season is our busiest and most visible period, our performance measures are for the entire fiscal year.
- The report is focused on the performance measures and their relationship to the SPP without any mention of the importance of the Operating Units' Business

See comment 1.

2

Plans. The Business Plan is a derivative of the SPP that is linked to tactical actions, resource allocations, and performance milestones that drive the day-to-day activities and goals of the Operating Units. The Business Plan is the primary vehicle for accountability through the Business Performance Review Process and individual performance appraisals.

See comment 2.

- Few of our performance measures are isolated measures. No individual measures can adequately reflect the broad range of our responsibilities and our mission. We manage our programs by reviewing performance measures, diagnostic measures, performance indicators, and numerous other data sources to ensure a broad perspective of our service to our customers.

I have addressed the recommendations in more detail below:

Recommendations for Executive Action

We recommend that the Commissioner of Internal Revenue Service direct the appropriate officials to do the following:

Recommendation 1
Take steps to ensure that the agencywide goals clearly align with the operating division goals and performance measures for each of the four areas reviewed. Specifically, (1) clearly document the relationship among agencywide goals, operating division goals, and performance measures (the other three program areas may want to consider developing a template similar to the one Field Assistance developed, shown in figure 4) and (2) ensure that the relationship among goals and measures is communicated to staff at all levels of the organization.

Response
We agree with this recommendation and in the next SPP, we will review the performance measures for the four W&I areas to ensure that we align and document their relationship to operating division goals and agencywide goals. The Operating Units' Business Plans communicate the relationship of SPP goals and measures throughout the organization. Staff at all levels should recognize their role in delivering the Business Plan. The four program areas reviewed in the W&I Division distributed information on the SPP and Business Plan through annual leadership conferences at each site for FYs 2002 and 2003.

See comment 3.

Recommendation 2
Make the name and definition of several field assistance measures (i.e., "geographic coverage," "return preparation contacts," "return preparation units," "TACs total

3

contacts," "forms contacts," "tax law contacts," "account contacts," "other contacts," "tax law accuracy," "account/notice accuracy," and "return preparation accuracy") more clear to indicate what is and is not included in the formula.

Response
Field Assistance recently updated the data dictionary for FY 2003. The updated dictionary addresses your recommendation on clarity and specifically identifies what is or is not included in the formulas. We gave a copy of the updated document to your staff. We have updated the data dictionary to include the purpose of the performance measurement, the data limitations associated with data gathering of the measure, and calculation changes from the prior year. It also provides a complete description of the methodology used in capturing the data, the critical path of how the measure originates and moves through the process, and the level of reviews to ensure quality. Field Assistance uses the current data dictionary in reporting measures to all levels of the organization.

Recommendation 3
As discussed in the body of this report and in appendix II, modify the formulas used to compute various measures to improve clarity. If formulas cannot be implemented in time for the next issuance of the SPP, then modify the name and definition of the following measures so it is clearer what is or is not included in the measure.

Recommendation 3(a)
Remove automated calls from the formula for the "CSR level of service" measure.

Response
We published the definition of this measure in the SPP, Data Dictionary, Measures Matrix, and numerous other sources. We believe that including the count of callers who choose an automated service while waiting for CSR service is appropriate. The formula accurately reflects the percentage of customers that wanted to speak to a CSR and subsequently received service. While we are promoting the use of automated services as an alternative to CSR service, we expect increases to occur before a customer enters the CSR queue. The growth in automation service while in queue for CSR service should remain small or decrease. We do not believe that this measure merits further change.

Recommendation 3(b)
Revise the "CSR response level" measure to include calls from taxpayers who tried to reach a CSR but did not, such as those who (1) hung up while waiting to speak to a CSR, (2) were provided access only to automated services and hung up, and (3) received a busy signal.

4

Response
We do not agree that we should modify this measure. The methodology and 30 second threshold for this measure is in accordance with the industry standard. This measure only applies to services answered and should not include abandon calls, automated service disconnects, or busy signals. Altering this measure would deviate from the industry standards and hinder our ability to gauge success in meeting this "world class service" goal.

Recommendation 3(c)
Analyze and use new or existing data to determine why calls are transferred and use the data to revise the "CSR services provided" measure so that it only reflects transferred calls in which the caller received help from more than one CSR (i.e., exclude calls in which a CSR simply transferred the call and did not provide service.)

Response
We agree in concept with your recommendation. We are continuing to examine previously collected data on transferred calls from FY 2002. We are also studying the anticipated impact that our new Toll Free Operating Strategy will have on this measure. We specifically designed this strategy to simplify the scripts and telephone menus to make the customer's self-selection process easier and more efficient. After assessing the impact of the Toll Free Operating Strategy, we will then review the recommendation for possible change in FY 2004.

Recommendation 3(d)
Either discontinue use of the "number of IRS digital daily Web site hits" measure or revise the way "hits" are calculated so that the measure more accurately reflects usage.

Response
Due to privacy restrictions associated with the use of "cookies," we cannot track the actual web site use. Instead, for FY 2003, we will implement three new diagnostic indicators related to the web site. These indicators (page view, unique visitors, and visits) will give us additional information to track the system performance and gauge the traffic on the web site. We will monitor these indicators for a year and decide whether to include them as performance measures in the 2004-2005 SPP.

We will also continue to measure the number of hits and downloads to the web site. However, we will clarify the definition of "hits" to reflect that each file requested by a visitor registers as a hit and several hits can occur on each page.

5

Recommendation 3(e)
Revise Field Assistance's "geographic coverage" measure by ensuring that the formula better reflects (1) the various types of field assistance facilities, including alternate sites and kiosks; (2) the types of services provided by each facility; and (3) the facility's operating hours.

Response
We agree that we should have revised the geographic coverage description to include more than just Taxpayer Assistance Centers (TAC). We are working with representatives from the Office of Program Evaluation and Risk Analysis to modify the formula to ensure that the formula reflects the appropriate elements by June 30, 2003. In addition, we will use the model to assist in determining the locations for different delivery options.

Recommendation 3(f)
Revise Submission Processing's "productivity" measure so it provides more meaningful information to users.

Response
We recognize that this measure needs improvement. The broad range of returns and documents processed and numerous other variables that can impact efficiency drives the complexity of the measure. The current measure seeks to account for those differences to ensure equity and fairness in the measurement process. We have looked at alternative ways to measure productivity but have not found a suitable replacement for this measure. We will continue our efforts to develop a more meaningful productivity measurement.

Recommendation 4
Refrain from making changes to official targets, such as Electronic Filing and Assistance did in FY 2001, unless extenuating circumstances arise. Disclose any extenuating circumstances in the SPP and other key documents.

Response
We agree that we should only make changes to official targets under the circumstances you describe and that disclosing these changes is appropriate. This approach is consistent with our overall practice.

Recommendation 5
Modify procedures for the toll-free customer satisfaction survey, possibly by requiring that the administrators listen to the entire call, to better ensure that the administrators (1) notify CSRs that their call was selected for the survey as close to the end of the call

6

as possible and (2) can accurately answer the questions they are responsible for on the survey.

Response
We agree we can improve this process. We will instruct the administrators to listen to each call from its beginning to as close to the conclusion as practical. Formalizing this practice will also enable the administrators to accurately answer the questions on the survey.

Recommendation 6
Implement annual effectiveness studies to validate the accuracy of the data collection methods used for the five telephone measures ("toll-free tax law quality," "toll-free accounts quality," "toll-free tax law correct response rate," "toll-free account correct response rate," and "toll-free timeliness") subject to potential consistency problems. The studies could determine the extent to which variation exists in collecting data and recognize the associated impact on the affected measures. For those measures, and for the five Submission Processing measures that already have effectiveness studies in place ("refund timeliness-individual (paper)," "notice error rate," "refund error rate-individual (paper)," "letter error rate," and "deposit error rate"), IRS should establish goals for improving consistency, as needed.

Response
We have ongoing processes to ensure that we properly administer the collection methods for the five telephone measures to minimize potential consistency problems. We do not agree that an annual independent review by a non-CQRS analyst is merited. Members of the Treasury Inspector General for Tax Administration (TIGTA) perform in-depth oversight activities annually covering these collection methods. While we will work to improve consistency, we do not agree that we should incorporate a consistency improvement goal in the SPP process.

Recommendation 7
Ensure that plans to remove overlapping measures in Telephone and Field Assistance are implemented.

Response
We will continue our process of reviewing measures identified as overlapping and deleted those that truly were redundant.

7

Recommendation 8
As discussed in the body of this report, include the following missing measures in the SPP in order to better cover governmentwide priorities and achieve balance.

Recommendation 8(a)
In the spirit of provisions in the Chief Financial Officer's Act of 1990 and Financial Accounting Standards Number 4, develop a cost of services measure using the best information currently available for each of the four areas discussed in this report, recognizing data limitations as prescribed by GPRA. In doing so, adhere to guidance, such as Office of Management and Budget Circular A-76, and consider seeking outside counsel to determine best or industry practices.

Response
Development of cost of services measures for Telephone Assistance, Electronic Filing and Assistance, Field Assistance, and Submission Processing is dependent on Servicewide deployment of the Integrated Financial System (IFS). The first release of IFS, scheduled for October 2003, will facilitate financial reporting and financial audits. The second release of IFS, planned for March 2005, will include Property and Performance Management. At this time, the development of cost of services measures is directly linked to having a mechanism that provides cost information for performance activities. The Service is moving towards this goal with successful implementation of the IFS system.

Recommendation 8(b)
Given the importance of automated telephone assistance, develop a customer satisfaction survey and measure for automated assistance.

Response
We agree that measuring customer satisfaction with automated services is important. Our newer interactive Internet services have satisfaction surveys incorporated in the program. We are continuing to upgrade our automated services and will be implementing telephone system architectural changes as part of the Customer Communications Engineering Study. We will review your recommendation to evaluate the benefit of programming and implementing a customer satisfaction survey system based on outdated delivery systems.

Recommendation 8(c)
Put the "automated completion rate" measure back in the SPP after revising the formula so that calls for recorded tax information are not counted as completed when taxpayers hang up before receiving service.

8

Response
We continue to track and monitor the "automated completion rate" as a diagnostic measure. We do not plan to modify the formula nor do we intend to reinstate it as a measure in the SPP.

Recommendation 8(d)
Add one or more quality measures to Electronic Filing and Assistance's suite of measures in the SPP. Possible measures include "processing accuracy," "refund timeliness, electronically filed," and "number of electronic returns rejected."

Response
The quality of electronic filing has consistently been high due to the pre-submission checks integrated into the system. We do track and monitor numerous diagnostic indicators that reflect the quality of electronic filing. We use this data to determine if there are error trends that need to be addressed. We do not believe incorporating these indicators as a performance measure in the SPP would enhance the electronic filing program.

Recommendation 8(e)
Re-implement Field Assistance's timeliness measure.

Response
Field Assistance agrees that timeliness goals are important in providing service to taxpayers; however, we found that this is detrimental to quality service in TACs because the employees tend to rush the customers when traffic is high. Realistic expectations provide a framework for our workers to provide appropriate service to the taxpayer with the goal of taking the requisite time to provide complete and accurate assistance. We will continue to use positive and negative feedback from customers responding to the "promptness of service" section of the satisfaction survey as a gauge of service. In addition, we are still tracking wait-times in locations equipped with the Queuing Management System (Q-Matic) System. The Q-Matic System is an on-line automated tracking and reporting system. We agree errors occur when manual methods of tracking workload volume and staff hours are used. In order to minimize reporting errors and better track wait-time, we plan to equip all of our TACs with this system. We can have Q-Matic installed and networked at all TACs nationwide by the end of FY 2004 with the planned funding.

9

Recommendation 8(f)
Develop a measure that provides information about Field Assistance's efficiency.

Response
Field Assistance is implementing a performance monitoring system to monitor productivity measures. We will use this system as a diagnostic tool to identify organizational performance measures strengths and weaknesses and not as an evaluative tool as we are a Section 1204 organization. We will test the system during FY 2003 to determine the validity and usefulness of the data captured. At the end of the fiscal year, we will decide whether to continue with the current system, or modify it.

Again, I appreciate your observations and recommendations. If you have questions or comments, please call Floyd Williams, Director, Legislative Affairs, at (202) 622-3720.

Sincerely,

Charles O. Rossotti

Charles O. Rossotti

GAO Comments

1. We recognize that IRS's performance measures cover entire fiscal years. We reviewed 53 of the measures for all of fiscal year 2001, and we reported the full year's results in appendix II.

2. We reviewed the business plans for all four program areas we reviewed. Although we did not comment specifically about the business performance review process in the report, we noted in the background and field assistance sections that the business plans communicate part of the relationship among the various goals and measures.

3. Figure 4 shows an excerpt of field assistance's business unit plan. As noted in the figure, the template used to communicate the relationship between goals and measure is missing some key components. Figure 2 is our attempt to show the complete relationship among IRS's various goals and measures—it is based on multiple documents.

Appendix IV: GAO Contacts and Staff Acknowledgments

GAO Contacts

James White (202) 512-9110
Dave Attianese (202) 512-9110

Acknowledgments

In addition to those named above, Bob Arcenia, Healther Bothwell, Rudy Chatlos, Grace Coleman, Evan Gilman, Ron Heisterkamp, Ronald Jones, John Lesser, Allen Lomax, Theresa Mechem, Libby Mixon, Susan Ragland, Meg Skiba, Joanna Stamatiades, and Caroline Villanueva made key contributions to this report.

Bibliography

To determine whether the Internal Revenue Service's (IRS) performance goals and measures in four key program areas demonstrate results, are limited to the vital few, cover multiple program priorities, and provide useful information in decision making, we developed attributes of performance goals and measures. These attributes were largely based on previously established criteria found in prior GAO reports; our review of key legislation, such as the Government Performance and Results Act of 1993 (GPRA) and the IRS Restructuring and Reform Act of 1998; and other performance management literature. Sources we referred to for this report follow.

101st Congress. *Chief Financial Officer's Act of 1990*. P.L. 101-576. Washington, D.C.: January 23, 1990.

103rd Congress. *Government Performance and Result Act of 1993*. P.L. 103-62. Washington, D.C.: January 5, 1993.

103rd U.S. Senate. *The Senate Committee on Government Affairs GPRA Report.* Report 103-58. Washington, D.C.: June 16, 1993.

105th Congress. *IRS Restructuring and Reform Act. P.L. 105-206*. Washington, D.C.: July 22, 1998.

Internal Revenue Service. *Managing Statistics in a Balanced Measures System*. Handbook 105.4. Washington, D.C.: October 1, 2000.

The National Partnership for Reinventing Government. *Balancing Measures: Best Practices in Performance Management*. Washington, D.C.: August 1, 1999.

Office of Management and Budget, *Preparation and Submission of Budget Estimates*. Circular No. A-11, Revised. Transmittal Memorandum No. 72. Washington, D.C.: July 12, 1999.

Office of Management and Budget. *Circular A-76*, Revised. Supplemental Handbook, Performance of Commercial Activities. Washington, D.C.: March 1996 (Revised 1999).

Office of Management and Budget. *Managerial Cost Accounting Concepts and Standards for the Federal Government*. Statement of Federal Financial Accounting Standards, Number 4. Washington, D.C.: July 31, 1995

Related Products

U.S. General Accounting Office. *Internal Revenue Service: Assessment of Budget Request for Fiscal Year 2003 and Interim Results of 2002 Tax Filing Season*. (GAO-02-580T). Washington, D.C.: April 9, 2002.

U.S. General Accounting Office. *Tax Administration: Assessment of IRS's 2001 Tax Filing Season*. (GAO-02-144). Washington, D.C.: December 21, 2001.

U.S. General Accounting Office. Human Capital*: Practices That Empowered and Involved Employees* (GAO-01-1070). Washington, D.C.: September 14, 2001.

U.S. General Accounting Office. *Managing For Results: Emerging Benefits From Selected Agencies' Use of Performance Agreements* (GAO-01-115). Washington, D.C.: October 30, 2000.

U.S. General Accounting Office. *Agency Performance Plans: Examples of Practices That Can Improve Usefulness to Decisionmakers* (GAO/GGD/AIMD-99-69). Washington, D.C.: February 26,1999.

U.S. General Accounting Office. *The Results Act: An Evaluator's Guide to Assessing Agency Annual Performance Plans* (GAO/GGD-10.1.20). Washington, D.C.: April 1,1998.

U.S. General Accounting Office. Executive Guide: Effectively Implementing the Government Performance and Results Act (GAO/GGD-96-118). Washington, D.C.: June 1996.

U.S. General Accounting Office. *Executive Guide: Improving Mission Performance Through Strategic Information Management and Technology* (GAO/AIMD-94-115). Washington, D.C.: May 1, 1994.

GAO's Mission

The General Accounting Office, the investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through the Internet. GAO's Web site (www.gao.gov) contains abstracts and full-text files of current reports and testimony and an expanding archive of older products. The Web site features a search engine to help you locate documents using key words and phrases. You can print these documents in their entirety, including charts and other graphics.

Each day, GAO issues a list of newly released reports, testimony, and correspondence. GAO posts this list, known as "Today's Reports," on its Web site daily. The list contains links to the full-text document files. To have GAO e-mail this list to you every afternoon, go to www.gao.gov and select "Subscribe to daily E-mail alert for newly released products" under the GAO Reports heading.

Order by Mail or Phone

The first copy of each printed report is free. Additional copies are $2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:

U.S. General Accounting Office
441 G Street NW, Room LM
Washington, D.C. 20548

To order by Phone:	Voice:	(202) 512-6000
	TDD:	(202) 512-2537
	Fax:	(202) 512-6061

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm
E-mail: fraudnet@gao.gov
Automated answering system: (800) 424-5454 or (202) 512-7470

Public Affairs

Jeff Nelligan, managing director, NelliganJ@gao.gov (202) 512-4800
U.S. General Accounting Office, 441 G Street NW, Room 7149
Washington, D.C. 20548

PRINTED ON RECYCLED PAPER