United States General Accounting Office

GAO

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

November 1998

IRS PERSONNEL ADMINISTRATION

Use of Enforcement Statistics in Employee Evaluations

GAO/GGD-99-11

GAO

United States
General Accounting Office
Washington, D.C. 20548

General Government Division

B-279232

November 30, 1998

The Honorable Bill Archer
Chairman, Committee on Ways and Means
House of Representatives

The Honorable Nancy L. Johnson
Chairman, Subcommittee on Oversight
Committee on Ways and Means
House of Representatives

Both the law and Internal Revenue Service (IRS) policy prohibit IRS managers from using records of tax enforcement results, such as dollars collected and taxes assessed, to evaluate employee performance or establish production goals. The underlying reason for these prohibitions is that employees may try to achieve statistical benchmarks through inappropriate or abusive enforcement actions, such as unwarranted seizures or unreasonable tax assessments, instead of acting strictly on the merits of the case.

This report responds to your request for information on IRS managers' compliance with legislative and policy prohibitions and the requirement that managers certify each quarter whether violations had occurred. Accordingly, this report discusses our evaluation of (1) the extent to which IRS' certification process identified violations of law and policy, (2) IRS employees' perceptions of the use of tax enforcement results in their annual performance evaluations, (3) supervisors' use of tax enforcement results in written employee performance evaluations, and (4) IRS' efforts to revise the certification process.

Background

Since 1973, IRS Policy Statement P-1-20 has prohibited using records of tax enforcement results to evaluate enforcement officers or to impose or suggest production goals or quotas. An enforcement officer is an employee who exercises judgment with regard to determining tax liability or the ability to pay. Enforcement officers include employees directly involved in collection, examination, or criminal investigation functions as well as appeals officers and reviewers. Each function has its own set of tax enforcement results. For example, for the examination function, tax enforcement results include data such as the number of hours spent per return, amount of dollars proposed per return, and number of cases closed. For the collection function, tax enforcement results include the number of liens, levies, and seizures; amount of dollars collected; and number of cases closed.

As a general rule, the policy prohibited managers from using any statistic that measures quantity, time per case, type of disposition, or dollar value. Instead, employees are to be evaluated on the basis of the quality of the work, including the use of enforcement tools, timeliness of action, and application of tax law. However, the policy statement permitted IRS officials to use statistics on tax enforcement results when carrying out certain management activities, such as preparing long-range and financial plans, allocating resources, and evaluating program effectiveness. In addition, managers were permitted to use tax enforcement results when evaluating employee performance during a case review, for example by noting whether a revenue officer appropriately levied a bank account given the facts of the specific case. Managers were also permitted to discuss with employees the number of cases processed, the amount of time spent on cases, and the kind of results obtained, as long as the discussion was based on a review of the results that were appropriate for specific cases processed by the employees and goals and quotas were not involved.

In 1988, along with other changes, the Taxpayer Bill of Rights[1] also prohibited the use of records of tax enforcement results to evaluate employees, but the prohibition only applied to employees directly involved in collection activities and their immediate supervisors. The law specified that IRS would be in compliance with the legislative requirements as long as IRS followed its policy statement. To help ensure compliance, the law also required each district director to certify each quarter whether violations had occurred and the corrective action taken, if needed. Because IRS service centers also have collection responsibilities, IRS imposed the same certification requirement on its service center directors. Generally, the self-certification process requires managers to submit certifications to the next higher level for review. Quarterly, group managers are to submit certifications to their branch chiefs, branch chiefs are to submit certifications to their division chiefs, and division chiefs are to submit certifications to their directors. Directors, relying on lower-level managers' certifications, are to submit their certifications to the Commissioner of Internal Revenue through the Assistant Commissioner (Collection).

[1]Taxpayer Bill of Rights, Public Law 100-647, Nov. 10, 1988.

During hearings held by the Senate Committee on Finance in September 1997, witnesses alleged that IRS' focus on enforcement statistics was encouraging enforcement officers to take unnecessary and illegal enforcement actions against taxpayers. Prompted by these hearings, IRS' Internal Audit reviewed the use of enforcement statistics by collection personnel in 12 districts, as well as at national and regional offices, and found an atmosphere largely driven by statistical measures. Internal Audit also found similar conditions in its July 1998 report on the use of enforcement results by the examination functions in 12 district offices.

The IRS Restructuring and Reform Act of 1998 (Public Law 105-206, July 22, 1998) repealed the Taxpayer Bill of Rights prohibitions, which only covered collection employees, and instead expanded the prohibitions to cover all employees. The act also expanded the quarterly certifications to cover each appropriate supervisor.

Results in Brief

For fiscal years 1996 and 1997, district and service center directors submitted 368 quarterly certifications that reported 11 potential violations, of which 4 were subsequently determined by IRS to be actual violations. However, we identified several systemic weaknesses that affected the reliability of the certifications. Specifically, we found (1) some confusion among IRS officials about what constituted a violation, (2) inadequate guidance about specific actions directors should take to identify violations, (3) a failure to integrate performance evaluations and the certification process, and (4) unclear guidance on sanctions that could be applied against managers for misusing tax enforcement results or submitting false certifications.

While the quarterly certifications showed few violations, our survey of a statistically representative sample of examination and collection employees showed a widespread perception that managers considered enforcement results when preparing annual performance evaluations. We estimated that 75 percent of front-line employees (revenue agents, tax auditors, and revenue officers) and 81 percent of group managers perceived that tax enforcement results affected their most recent performance evaluation. When asked about various potential bases for their perception, about 70 percent of front-line employees, who indicated that tax enforcement results affected their evaluations, said they based their perception in part on information communicated to them verbally in staff meetings or performance feedback sessions with their managers. And about 36 percent of front-line employees indicated tax enforcement results were used in their written performance evaluations.

On the basis of our review of the two most recent performance evaluations for a statistically representative sample of front-line employees, we estimated that 9 percent of employees received a written evaluation that, according to our application of IRS' revised guidelines, contained a tax enforcement result. Although this incidence rate is low relative to our employee survey results, we also found that an estimated 69 percent of the evaluations in our sample contained narrative that employees could have interpreted as inappropriate references to tax enforcement results but which did not violate IRS guidance. For example, an estimated 50 percent of the evaluations in our sample included either general or case-specific references to enforcement-related activities, such as employees' overall use of collection tools or auditing techniques, that IRS does not consider violations. In addition, we estimated that about 41 percent of the evaluations in our sample mentioned process measures dealing with the age of the cases in the employee's workload inventory and the number of cases worked within guidelines established for closing cases, which could be interpreted by employees as a surrogate for prohibited measures, such as the "number of cases closed."

As a result of IRS' Internal Audit reviews, IRS has undertaken several steps to strengthen the certification process, such as (1) expanding the number of employees covered by certifications, (2) revising its guidance on the proper use of tax enforcement results, (3) requiring an annual independent review to verify the accuracy of the certifications, and (4) requiring the preparer and reviewer to attest that the tax enforcement results were not improperly used in the evaluation. Although these actions address some of the weaknesses of the current system, IRS' revised guidance has few examples of the appropriate and inappropriate use of tax enforcement results in written evaluations. In addition, the certification form includes only a general statement of compliance rather than a statement specifically detailing what managers are certifying. Furthermore, the revised guidance does not clearly inform managers about potential sanctions for inappropriate use of tax enforcement results. In addition to these potential sources of confusion, IRS' new independent review process is geared toward identifying written violations and not violations communicated verbally, which our employee survey indicates is the most common way tax enforcement results are communicated.

Scope and Methodology

To determine the extent to which IRS' current certification process identified violations, we reviewed fiscal year 1996 and 1997 quarterly certifications filed by district and service center directors—a total of 288 for the district offices and 80 for the service centers. We also obtained summary information on the potential violations for the first quarter of fiscal year 1998. To obtain information on IRS' quarterly certification policy and procedures, we interviewed officials from IRS' National Office; the Western Regional Office; the Georgia, Kansas/Missouri, and Northern California District Offices; and the Atlanta and Fresno Service Centers. We visited these offices because of their close proximity to our offices. We also surveyed officials in all districts and service centers on the quarterly certification process they had in place as of the last quarter of calendar year 1997. Appendix I contains a summary of the violations reported in the quarterly certifications we reviewed.

To determine IRS employees' perceptions of the use of tax enforcement results, we surveyed a statistically representative sample of 1,104 of 20,974 IRS front-line employees and group managers in the collection and examination functions. The survey included questions on whether employees perceived that tax enforcement results influenced their evaluations, the basis for those perceptions, and the use of tax enforcement results to establish group goals. We also analyzed the comments provided voluntarily by IRS employees on the survey. Appendixes II through VI contain our sampling methodology and survey results. All results are reported at the 95-percent confidence level. Sampling errors for all of our estimates are less than plus or minus 10 percentage points unless otherwise reported.

To determine how tax enforcement results were used in written employee evaluations, we reviewed the 2 most recent evaluations for a statistically representative sample of 300 of 18,292 examination and collection front-line employees. We reviewed the evaluations to determine whether tax enforcement results were used to rate the employee in violation of IRS guidance. We discussed potential violations with IRS officials and considered their opinions on whether they were violations. We also reviewed the evaluations to determine whether the narratives used tax enforcement results to generally characterize employee performance or to discuss specific case examples. Appendix II contains our sampling methodology and appendix VII contains the results of our review of employee evaluations. All results are reported at the 95-percent confidence level and, unless noted otherwise, the data have sampling errors of no more than 8 percent.

To evaluate IRS' efforts to revise the certification process, we discussed IRS proposals with senior-level officials who were involved in developing the new process. We reviewed the existing and proposed guidance provided to managers and identified the differences between the existing and proposed certification processes. We also met with National Treasury Employees Union (NTEU) officials to obtain their perspective on the existing and proposed processes. We compared the proposed changes with the results of our review of the current process to determine if the limitations we noted were being addressed.

There were some limitations to our analysis. We did not verify employees' responses to our survey. However, we discussed these responses with senior IRS officials. We also did not verify IRS' personnel database from which we drew our sample.

We did our work between December 1997 and August 1998 in accordance with generally accepted government auditing standards.

We requested comments on a draft of this report from the Commissioner of Internal Revenue. The comments we received are in appendix VIII and are evaluated at the end of this letter.

IRS' Certification Process Identified Few Violations but Had Systemic Weaknesses

Our review of quarterly certifications for fiscal years 1996 and 1997 found that directors reported 11 potential violations in 368 quarterly certifications (see app. I). IRS determined that four of the incidents were actual violations, two were unfounded allegations, and five were reported because the directors believed the actions could be misconstrued as violations. However, there are some indications that the certification process may not have identified all violations. For example, for the first quarter of fiscal year 1998, directors reported 28 potential enforcement results violations, more than double the total reported for the prior 8 quarters. According to IRS officials, the increase in reported potential violations occurred as a result of the September 1997 Senate hearings on IRS abuses, which made managers more aware of their responsibilities for reporting potential violations. IRS had not completed its investigation of these potential violations at the time we completed our audit work.

Also, we identified several systemic weaknesses that affected the reliability of the certification process. First, our analysis of the potential violations that were reported indicated that there was confusion among IRS officials about what constituted a violation. For example, one certification noted that a district director had asked the National Office official, who was responsible for providing guidance on the use of tax enforcement results, whether a group could establish a team goal to increase dollars collected by 10 percent as part of an organizational pilot project. Although the National Office official agreed that this was acceptable because neither awards nor individual evaluations would be based on achieving the goal, IRS' Chief Counsel subsequently determined that the group goal was a production quota or goal and, as such, a violation.

Second, IRS guidance did not specify how the certifications were to be supported. For example, IRS policy did not mandate what action directors should take to support their certifications, such as reviewing a sample of evaluations or indicating on the certification the number and type of documents reviewed. As a result, National Office officials may not have been in a position to determine whether directors performed enough oversight to validate their certifications. We found that officials at all levels with certification responsibility said that they reviewed some documents to determine whether managers complied with the tax enforcement results policy. However, the nature and scope of their review varied widely. For example, some officials reviewed all performance evaluations and supporting documents for violations while other officials reviewed a random sample or reviewed a small, judgmental sample.

Third, IRS did not integrate the employee performance evaluation and certification processes. For example, although IRS procedures require managers to submit an evaluation to a higher level of management for review and approval before the evaluation is given to the employee, they do not require the reviewer to assess the evaluation for potential tax enforcement results violations. Although the preparers' and reviewers' signatures on the evaluation form signified that the evaluation complied with procedures for conducting evaluations, the procedures did not specifically refer to the tax enforcement results policy or the certification process. Such a requirement could heighten managerial awareness of policy and legal prohibitions at the time the evaluation is prepared.

Fourth, IRS guidance did not clearly inform managers that they could be subject to sanctions or the kinds of sanctions that could be imposed if they either misused tax enforcement results or signed a false or misleading certification. Neither the policy guidance nor the certification included an acknowledgement that managers could be subject to disciplinary action for the improper use of tax enforcement records, so managers might not have known the importance or potential consequences of their actions.

Furthermore, the lack of clear guidance could lead to inconsistent disciplinary actions. The Guide for Penalty Determinations listed in IRS' Interim Handbook of Employee Conduct and Ethical Behavior did not specify the types of disciplinary actions that could be imposed on managers who violate tax enforcement results policy. According to an IRS official, the guide is not intended to be an exhaustive listing of all offenses, and one general category can be used for penalty determinations in matters not otherwise covered. The penalties for this general category ranged from a reprimand, to a 14-day suspension, to a removal. Disciplinary action for the four violations IRS identified in the fiscal year 1996 and 1997 certifications ranged from counseling to removing the manager from his/her position.

Employees Perceive That Tax Enforcement Results Affect Evaluations

Our employee survey showed a widespread perception on the part of collection and examination front-line employees that tax enforcement results are considered by managers when preparing performance evaluations. We estimated that, overall, about 75 percent of front-line employees (i.e., revenue agents and tax auditors in the examination function and revenue officers in the collection function) and 81 percent of examination and collection group managers believe that one or more enforcement measures were considered to some extent by their managers when preparing their most recent performance evaluation.

For those revenue officers who indicated that tax enforcement results affected their evaluation, an estimated 58 percent of them cited "number of cases closed"—a measure of productivity—as a tax enforcement result influencing their evaluations. Revenue agents and tax auditors who indicated that tax enforcement results affected their evaluations cited measures of revenue production more frequently as factors that influenced their evaluations than did revenue officers. At least an estimated 61 percent of revenue agents and tax auditors said that "additional dollars proposed per return" influenced their evaluations, while an estimated 29 percent of revenue officers said that "average dollars collected per return" influenced their evaluations. Appendix III shows the point-estimates intervals for each type of enforcement result by type of employee and manager.

An estimated 77 percent of collection group managers who indicated that tax enforcement results affected their evaluations believed that the number of cases closed affected their evaluations, compared with an estimated 39 percent who believed that the average dollars collected per return affected their evaluations. Examination group managers who indicated that tax enforcement results affected their evaluations cited workload measures about as frequently as revenue production measures. About 74 percent of examination group managers identified tax enforcement results dealing with the average hours per return and about 76 percent identified additional dollars proposed per return as affecting their evaluations.

IRS' quarterly certification process focused on the use of enforcement results in formal written performance evaluations. However, our survey results indicated that although many employees had multiple bases for their perceptions on the use of tax enforcement statistics, most based their perception on verbal information received from their managers. We estimated that about 70 percent of front-line employees, who believed that tax enforcement results affected their evaluations, based their perceptions on verbal communication with their managers during group meetings and performance feedback sessions. An estimated 36 percent based their perception on written information contained in their evaluations. Likewise, managers who believed that tax enforcement results affected their evaluations had multiple bases for their perceptions. An estimated 83 percent based their perception on verbal communication with their branch and division chiefs, while an estimated 51 percent based their perception on written communications. Appendix IV contains detailed information on the basis for front-line employee and group manager perceptions.

IRS' managers are prohibited from using tax enforcement results to establish group production quotas or goals because of the possibility that group goals will be interpreted as individual goals or expectations. About one-half of front-line employees and group managers indicated that performance goals were established for their groups. Where group goals were established, at least two-thirds of the front-line employees and managers indicated that the goals included one or more enforcement results. Appendix V provides more detailed information on which enforcement results were included in group goals.

In voluntary written comments to our questionnaires, front-line employees and managers included additional reasons why they believe enforcement results influenced their performance evaluations. The comments generally centered on the following scenarios: (1) supervisors or managers implied that an employee's performance evaluation was influenced by tax enforcement results; (2) the general culture or atmosphere of IRS implied that enforcement results affected performance evaluations; (3) higher-level managers pressured group managers to increase production or revenue yield; and (4) IRS' business plan, reports, and other documents emphasized enforcement results as goals. Front-line employees and managers also made positive and negative comments about using enforcement and workload measures to establish group goals or in performance evaluations. Generally, the number of negative comments made by front-line employees and managers exceeded the number of positive comments. Appendix VI contains examples of voluntary comments received from front-line employees and managers.

Employee Evaluations Contained Policy Violations and Other References to Enforcement and Productivity-Related Activities

Using IRS guidance for determining whether an evaluation contained a violation, we estimated that 9 percent of front-line employees received at least one written performance evaluation containing a violation in the two most recent annual evaluations. The potential effect of the violations on future employee performance could vary, however, depending on how the enforcement results were used in evaluations. Some violations could influence employees to be more aggressive in their dealings with taxpayers. For example, the following two comments taken from evaluations could create an incentive for employees to more aggressively pursue fraud cases, liens, and levies.

> "You had one possible fraud case during this period. Keep trying for more fraud cases."

> "You do not hesitate to lien or levy and have demonstrated innovative ideas in your efforts to comply with group goals. . . Your workload management produces a closure rate that demonstrates your commitment to group goals."

Both evaluations violations link a positive evaluation to achieving a higher number of enforcement results, rather than focusing on the appropriateness of the employees' action.

In other cases, however, the potential impact of the violations on employee performance was less clear. For example, many of the violations mentioned closing cases when summarizing employees' accomplishments, as shown in the following two examples.

> "The effectiveness of your planning and time utilization is demonstrated by the fact that you closed 38 key cases during the year."

"You have selected the proper technique to detect unreported income and applied your technique to detect and recognize indications of fraud. For example, [employee name], during this fiscal year you have submitted and had accepted six civil fraud referrals."

For most of the violations we identified, IRS officials said that the comments would not have been considered violations under the guidance in effect at the time the evaluations were prepared. However, the officials agreed that the comments would be considered inappropriate under IRS' revised guidance, which is discussed in the following section.

The incidence of violations was relatively low, given that 36 percent of front-line employees who believed that tax enforcement results affected their evaluations based their perception on their written evaluations. Our analysis of employee evaluations found that an estimated 69 percent of the evaluations contained two types of narrative that employees could have interpreted as inappropriate references to tax enforcement results but which did not violate IRS guidance. The first type of narrative involved the use of general or case-specific references to enforcement-related activities, such as the amounts of revenue collected or use of collection tools. These references, which are not based on statistics such as group goals, are not violations. We estimated that 50 percent of employees received evaluations that contained this type of narrative. For example, one evaluation included the statement "You had one case that the property was in foreclosure and you were looking at seizing or redeeming the property but would only get around $23,000. Through your working with a third-party lending institution, you were able to secure $30,000. . . ." Another included "[employee name] . . . have done a very good job in locating assets even in unusual situations. An example of this would be the research you completed on a large dollar tax protestor assigned to your inventory." Since the comments were based on specific cases reviewed, they are not considered tax enforcement results violations under IRS' guidelines. However, IRS guidance cautions managers that discussing a tax deficiency or referencing a dollar amount collected may give the employee an improper perception that only the size of the deficiency or the amount collected was the basis for the employee's evaluation.

The second type of narrative that could create misperceptions about the use of tax enforcement results was the use of the terms "overage" (the age of the cases in the employee's workload inventory) and "cycle time" (the number of cases worked within the established guidelines for closing cases). IRS views these statistics as measures of process time and responsiveness to taxpayers because they are indications of timely attention to taxpayer issues rather than measures of tax enforcement results. We estimate that 41 percent of employees received evaluations that included either general or case-specific references to overage cases and cycle time. For example, one evaluation included the statement “You are making excellent progress in closing out cases and specifically targeting overage and high cycle time to minimize these problems.” Another included the statement “You appear to be doing a better job of meeting the time frames for initial contact and follow-up action on your cases, but your high overage and potential overage rates still indicate a need to improve in this area.” Although neither cycle time nor overage cases are defined by IRS as enforcement results, IRS guidance cautions managers in their use of such terms because of the potential for either of the terms being interpreted as a prohibited statistic such as “hours per case.” Also, during our discussions with NTEU officials, they raised concerns about whether front-line employees and group managers clearly understood the distinction between the inappropriate use of productivity-related enforcement measures, such as hours per case, and the appropriate use of overage and cycle-time data. Appendix VII provides the detailed results for our review of employee evaluations.

IRS’ Proposed Revisions to the Certification Process Address Some Weaknesses

Because of Internal Audit reports, IRS has undertaken several steps to strengthen the certification process, including expanding the number of employees covered, instituting an independent review process of the certifications, revising guidance, and training managers.

According to IRS officials, IRS has expanded the certification process to include all appropriate enforcement supervisors, not just those working in the collection function. Employees who are to be covered by the certification include revenue agents, tax auditors, appeals officers, and criminal investigators. The IRS Restructuring and Reform Act of 1998 also mandates quarterly certification of all appropriate enforcement supervisors. IRS is also instituting an annual independent review to verify the accuracy of the certifications. The reviews are to be performed annually by cross-functional management teams and must include a review of Employee Performance Folders and employee evaluations, which may include documentation related to the evaluation process—such as award narratives and case reviews. The review team may also consider other sources, such as local memos, minutes of meetings, and grievances submitted by employees.

IRS has also implemented a new employee evaluation process effective September 1998 that includes a revised form to document a performance evaluation. The guidance on the new form sent to IRS managers in August 1998 states that by signing the form the rater and reviewer certify that enforcement statistics were not used in preparing the appraisal.

In addition, IRS has revised the handbook that provides guidance on the appropriate and inappropriate use of tax enforcement results. The handbook provides guidance on which employees are considered enforcement officers and examples of common scenarios involving the proper and improper use of tax enforcement results for each job classification. The handbook also includes a decision table that leads managers through a series of four questions to help them determine whether they are using tax enforcement results appropriately. This new guidance may help clarify some of the confusion managers had on what constitutes a tax enforcement result violation.

Furthermore, IRS trained managers on the new certification process and handbook. After providing background on the prohibitions, the training discussed the appropriate use of tax enforcement results and how the prohibitions affect supervisory activities, such as setting expectations as well as explaining the self-certification and independent review processes. To signal its commitment to addressing issues involving the use of enforcement statistics, IRS involved senior executives in the training. For example, IRS directed that district and service center directors would teach the section on self-certification and the independent review. IRS completed the training before the new certification process became effective in the last quarter of 1998.

These initiatives address existing weaknesses that we have identified by clarifying which employees are covered and the steps directors should take to independently validate certifications. However, managers may be confused about how to apply the decision table contained in the revised guidance to the preparation of written evaluations. The scenarios included in the revised guidance focus primarily on the prohibited use of statistical reports to imply group goals with relatively few examples relating to the appropriate and inappropriate uses of tax enforcement results in written evaluations. Providing such examples is particularly important because, as noted earlier, IRS officials said that most of the violations we identified would be considered violations under the revised guidance but not under the guidance in effect before 1998. IRS officials agreed that expanding the guidance to include additional examples could be beneficial, especially since managers raised many questions about the appropriate use of tax enforcement results in written evaluations during the training on the revised guidance.

Furthermore, IRS has revised its certification form to reflect that more employees are covered by the certification process, but the form contains only a vague statement of conformance. As shown in appendix IX, the certification does not specifically state that first-level managers have not personally misused records of tax enforcement results in written performance evaluations they prepared or to set individual production goals or quotas during their verbal communications with employees. In the case of second-level managers, the certifications do not specifically require that managers certify that tax enforcement results were not used in written performance evaluations they prepared or reviewed or during their verbal communications with employees. As a result, the form does not clearly explain to managers what they are certifying to.

IRS has not informed managers about what specific sanctions can be imposed for misusing tax enforcement results or submitting misleading certifications. Neither the revised guidance nor the certification form refers to potential sanctions.

In addition to these possible sources of confusion, IRS' independent review process is to focus on reviewing documents and records to identify written violations. As a result, IRS does not have a mechanism, such as our employee survey, for monitoring employee perceptions of how often tax enforcement results are verbally communicated or whether their written evaluations are influenced by records of tax enforcement results. Consequently, without an employee survey, IRS cannot determine whether further clarifications of the guidance or additional training is needed.

Conclusions

IRS directors reported few violations through the quarterly certification process in fiscal years 1996 and 1997, and we found an estimated 9 percent of employees received evaluations that were in violation of IRS' revised guidance during our review of employee evaluation files. Nonetheless, our survey of IRS employees indicated a widespread perception that managers consider tax enforcement results when preparing performance evaluations. Most of the employees indicated that the violations occurred during verbal communications with their supervisors, such as staff meetings or performance feedback sessions, rather than in their written performance evaluations. Perceptions based on such verbal communications could encourage employees to focus on achieving statistical benchmarks rather than acting solely on the merits of the case. Furthermore, the use of overage and cycle-time data in evaluations may further employee perceptions that tax enforcement results affect their evaluations, because they may be misconstrued as an enforcement statistic, such as hours per case.

Although IRS is taking steps to strengthen its reporting of violations, weaknesses remain. IRS' revised guidance has few examples of the appropriate and inappropriate use of tax enforcement results in written evaluations. Furthermore, the certification form does not specifically require managers to certify that tax enforcement results were not used in written evaluations or used inappropriately during verbal communications with employees. IRS has not provided clear guidance on the sanctions for misusing tax enforcement results or submitting misleading certifications. Additionally, IRS does not have a mechanism for monitoring how often violations result from verbal communications or employee perceptions of whether records of tax enforcement results are influencing their evaluations.

Recommendations to the Commissioner of Internal Revenue

To better ensure managerial accountability for the proper use of tax enforcement results, we recommend that the Commissioner

- expand the guidance to include additional examples of the appropriate and inappropriate use of records of tax enforcement results in written evaluations,
- revise the quarterly certification form to specifically state that tax enforcement results were not used in any written employee evaluation prepared or reviewed, including appraisals, awards, or promotion justifications, and that the manager did not verbally communicate to employees that tax enforcement results affected their evaluations or were used to set individual production goals or quotas,
- revise the penalty guide to specifically list the disciplinary actions that can be taken for violations, and
- survey employees periodically to determine whether they perceive that tax enforcement results were used in written evaluations or verbally communicated by their supervisors and use the results to assess whether IRS needs to further clarify the guidance, provide additional training, or take any other appropriate action.

Also, to avoid the potential inappropriate use of overage and cycle-time data, the IRS Commissioner should either designate overage and cycle-time data as prohibited tax enforcement results or emphasize in official policies or procedures to front-line employees and managers how overage and cycle-time data may be used appropriately.

Agency Comments and Our Evaluation

We obtained written comments on a draft of this report from the Commissioner of Internal Revenue. The Commissioner agreed with all of our recommendations and described the actions IRS plans to take in response to our recommendations. Regarding our recommendation that IRS should either designate overage and cycle-time data as prohibited tax enforcement results or emphasize in official policies or procedures to front-line employees and managers how overage and cycle-time data may be used appropriately, the Commissioner said that IRS would revise the Internal Revenue Manual to include examples of the appropriate use of these data. If effectively implemented, the actions described should help to resolve the problems we identified.

In addition, the Commissioner pointed out that IRS has undertaken a number of initiatives to address the use of tax enforcement results in employee performance evaluations. These initiatives included (1) revising the Internal Revenue Manual provisions concerning the use of enforcement statistics; (2) providing training to all managers on the use of enforcement statistics, which included numerous examples of appropriate and inappropriate language for discussing tax enforcement results in evaluations; (3) providing an orientation course on the proper use of statistics for all IRS employees; (4) establishing a Support Panel to answer questions from managers or employees on the proper use of enforcement statistics; (5) establishing independent review teams that are obliged to review documentation in Employee Performance Files and Employee Evaluations and may, at the director's discretion, look to other sources of information to ensure potential violations are uncovered, including group discussions, union comments or reports, grievances, and minutes of managers' meetings; and (6) developing a balanced performance measure system to include measures of customer satisfaction, employee satisfaction, and business results. The business results measure is to focus on the quality and quantity of work done. The quantity measures are to consist exclusively of outcome-neutral production data, such as cases closed and time per closing.

We are sending copies of this report to the Ranking Minority Member of the House Ways and Means Committee; the Ranking Minority Member of the Subcommittee on Oversight, Committee on Ways and Means; various other congressional committees; the Secretary of the Treasury; the Commissioner of Internal Revenue; and other interested parties. We will also make copies available to others on request.

Please contact me at (202) 512-9110 if you or your staff have any questions. Major contributors to this report are listed in appendix X.

James R. White
Director, Tax Policy
and Administration

Contents

Tables

Abbreviations

IRS	Internal Revenue Service
NTEU	National Treasury Employees Union

Potential Violations Reported on Fiscal Years 1996 and 1997 Certifications

This appendix is a summary list of all potential violations that were reported on quarterly certifications for fiscal years 1996 and 1997. No violations were reported in the quarters ending June 30, 1996, September 30, 1996, and December 31, 1996. The Internal Revenue Service (IRS) determined that the reported incidents numbered 1, 2, 5, and 11 were actual violations.

1. Quarter Ending 12/31/95. A branch chief discussed enforcement statistics in a group manager's evaluation. According to a district office official, the violation, discovered during the next level review, was a statement to the effect that "Your group led the branch in the number of seizures." Corrective action included assigning the individual to another position outside the Collection Division, assigning an acting branch chief until a permanent replacement could be brought in, and reviewing guidelines.

2. Quarter Ending 03/31/96. A proposal to redesign the revenue officer position under an IRS test project recommended that individual groups negotiate their own production goals. One group established a goal to increase the amount of dollars collected by 10 percent in the next year. The National Office Coordinator certified that this goal was not a policy violation because (1) it was established as a part of a test, and (2) awards or individual evaluations would not be based on the achievement of the goal. District officials were still uncertain about whether this goal was in conflict with policy and requested an opinion from the Office of Assistant Commissioner (Collection), who in turn requested legal advice from IRS' Chief Counsel. The Chief Counsel ruled that the group's goal was a violation because it was a production quota or goal and corrective action needed to be taken to be in compliance with policy. The Chief Counsel also cautioned that IRS managers must have the authority to void goals negotiated under the test project if the goals violated law or policy.

3. Quarter Ending 03/31/97. A local chapter of the National Treasury Employees Union (NTEU) raised a concern about a branch chief who discussed the number of seizures reflected on a monthly activity report with his group managers. Although this instance was reported on the quarterly certification, the incident was not considered a violation because, under the policy, managers can discuss recorded statistics with employees as long as they do not imply that goals or quotas are being established. Because the comments were open to misinterpretation, district management said they reinforced the policy guidelines.

4. Quarter Ending 03/31/97. An Automated Call Site manager discussed the low number of calls made during 1 week with his employees and requested them to quadruple the number of calls. The manager also commended the group on the number of closures they made and identified the number of closures he was "going for." Although this instance was reported, it was not considered a violation because (1) Automated Call Site calls were not considered an enforcement statistic and (2) individual goals were not established. Nonetheless, the manager was counseled and requested to immediately meet with his team again to emphasize that it was the quality of the case work that resulted in the closures.

5. Quarter Ending 06/30/97. A group manager prepared a positive employee evaluation that was based on the amount of dollars collected. According to a district office official, this violation was discovered during a branch manager's quarterly policy review. The group manager was counseled and the evaluation was revised.

6. Quarter Ending 09/30/97. A district reported a pending allegation in which an employee claimed he was forced to make seizures. According to a district official, based on subsequent district and regional offices' investigations, the allegation was found to be unwarranted.

7. Quarter Ending 09/30/97. A district identified, during a division and branch chiefs' sample review of employees' performance binders, a few instances in which performance evaluations and documents contained enforcement results that could easily be misconstrued as violations (e.g., "His collection actions resulted in both the generation of substantial revenue and in the proper disposition of numerous cases"; "You have processed a number of Offer In Compromise recommendations"; etc.). Corrective actions included emphasizing the policy guidelines on the use of tax enforcement results and requiring more thorough reviews to assess adherence with policy and the law.

8. Quarter Ending 09/30/97. An acting branch manager distributed a memorandum to group managers indicating dollars collected were low and needed to be improved. Although not a violation, the acting branch manager was advised by the division chief that his memorandum was inappropriate and could be misconstrued. He was also instructed to schedule a meeting with the group managers on the real intent of his message to ensure they understood the focus was to identify reasons for the current level of performance and find ways to improve it.

9. Quarter Ending 09/30/97. According to the NTEU, a group manager shared with his group a report containing revenue officers' statistical accomplishments, including a recap of the dollars collected over a 3-year period. As a corrective action, managers were given instructions that this information should not be shared at the group level and that branch statistics, comparing groups within a branch, should not be shared below the branch level. While this was not classified as a violation at that time, district management officials said they would consider it a violation under the new draft certification procedures.

10. Quarter Ending 09/30/97. The NTEU reported an allegation that, in the presentation of a manager's award, a group manager mentioned that the revenue officer being honored had made the most seizures in the group. The manager denied making that statement, and the incident was not considered a violation. District officials said that, although not directly related to this incident, several actions were taken, including holding meetings and briefings to discuss the policy and the discontinuation of the distribution of statistics.

11. Quarter Ending 09/30/97. The NTEU and an outside party reported that a group manager placed a routing slip on employees' desks indicating the amount of dollars that needed to be collected by each revenue officer each day, week, and month. The group manager was temporarily reassigned until an investigation and administrative determination was completed. Instructions were given to branch chiefs not to provide statistical information below their branch level. As a general corrective action applicable to these three incidents, a meeting was held to discuss the importance of adhering to the policy and law, and guidance was distributed to managers.

GAO Sampling Methodology

This appendix discusses the sampling methodology we used to determine IRS employees' perceptions of the use of tax enforcement results and how tax enforcement results were used in employee evaluations.

Sample Designs

To determine IRS employees' perceptions of the use of tax enforcement results, we conducted statistically representative surveys of five groups of IRS employees: (1) revenue agents assigned to district office Examination Divisions and Employee Plans and Exempt Organizations Divisions, (2) tax auditors assigned to Examination Divisions, (3) revenue officers assigned to Collection Divisions, (4) group managers assigned to Examination Divisions and Employee Plans and Exempt Organizations Divisions, and (5) group managers assigned to Collection Divisions. To determine how tax enforcement results were used in employee evaluations, we reviewed performance evaluations for a separately drawn sample of revenue agents assigned to district office Examination Divisions and Employee Plans and Exempt Organizations Divisions, tax auditors assigned to Examination Divisions, and revenue officers assigned to Collection Divisions.

Both samples were drawn from IRS' personnel database as of January 3, 1998. Because the questionnaires for the employee surveys were not mailed until March 1998, the results do not include the opinions of any IRS employees who were in the groups of interest as of January 3, 1998, but were not in these groups at the time the questionnaire was sent. We assumed that the opinions of these employees, if they would have been included, would have been the same in the aggregate as those of employees who were included in the surveys. We sampled a total of 1,104 IRS employees from the 5 employee groups of interest from the population of 20,974 employees in the January 3, 1998, IRS personnel database. We allocated proportionately more of our sample to Collection Division revenue officers and managers, since they are directly engaged in collection activities.

For the objective of determining how tax enforcement results were used in employee evaluations, we reviewed the evaluations of a sample of 300 IRS employees from the 3 employee groups of interest. By design, none of the 300 employees selected were the same as the 1,104 employees sampled for the first objective. IRS managers are not required to write performance narratives for every rating dimension for every employee each year. To review more narratives, the two most recent ratings were requested from IRS for each employee in the sample. The results presented in the report only reflect employees for whom two performance evaluations were received.

Sample Disposition

For our survey of employees' perceptions of the use of tax enforcement results, we received usable responses (those that indicated the employee was a member of one of the five target groups at the time the questionnaire was received and responded to at least some of the questions) from 814 employees, for a response rate of approximately 74 percent. The disposition of the sampled cases for our employee survey and review of performance evaluations is shown in table II.1 and table II.2, respectively.

Table II.1: Disposition of Sample Cases for Our Survey of the Extent to Which Employees Perceived Their Ratings Were Based on Enforcement Results

		Initial selection			Nonrespondents	
Strata ID	Definition of strata	Number in universe	Number of selections	Eligible respondents	Known ineligible	Unknown eligibility
1	Examination group managers	1,335	125	100	3	22
2	Collection group managers	543	175	123	13	39
3	Tax auditors	1,867	125	84	6	35
4	Revenue agents	11,796	320	250	11	59
5	Revenue officers	5,433	359	257	8	94
	Total	**20,974**	**1,104**	**814**	**41**	**249**

Source: IRS data and GAO sample.

Table II.2: Disposition of Sample Cases for Our Review of the Extent to Which Enforcement Results Were Discussed in Employees' Evaluations

		Initial selection			Nonrespondents		
Strata ID	Definition of strata	Number in universe	Number of selections	Eligible respondents	Eligible	Known ineligible	Unknown eligibility
1	Tax auditors	1,867	30	28	1		1
2	Revenue agents	11,796	135	127	2		6
3	Revenue officers	5,433	135	120	6		9
	Total	**19,096**	**300**	**275**	**9**		**16**

Source: IRS data and GAO sample.

Calculation of Sample Estimates

After weighting the responses to the employee survey to account for selection probabilities and nonresponse, we were able to estimate the percentage of IRS employees who perceived that certain tax enforcement results were or were not used in their evaluations. We also were able to estimate the percentage of IRS employees for whom tax enforcement results were used in at least one of their two most recent employee evaluations.

Sampling Error

Because we reviewed a statistical sample of employees, each estimate developed from the samples had a measurable precision or sampling error. The sampling error is the maximum amount by which the estimate obtained from a statistical sample can be expected to differ from the true population value being estimated. Sampling errors are stated at a certain confidence level—in this case, 95 percent. This means that the chances are 19 out of 20 that, if we surveyed all IRS employees in the groups of interest, the true value obtained for a question on these surveys would differ from the estimate obtained from our sample by less than the sampling error for that question.

The sampling errors for our survey of the extent to which employees perceived that their ratings were based on enforcement results are less than plus or minus 10 percentage points, unless otherwise reported. The sampling errors for our sample of the extent to which enforcement results were discussed on employee evaluations are no more than plus or minus 8 percentage points.

Nonsampling Error

In addition to the reported sampling errors, the practical difficulties of conducting any survey may introduce other types of “nonsampling” errors. For example, differences in how a particular question is interpreted or the types of individuals that do not respond can introduce unwanted variability into the survey results.

Survey Results on Employee Perceptions of the Effect of Tax Enforcement Results on Evaluations

Our survey asked a representative sample of IRS employees the question "To what extent, if at all, do you perceive that each of the following factors was considered by your supervisor or managers in preparing your most recent performance evaluation?" The "factors" shown for this question were the tax enforcement results associated with each employee category. The responses are shown in tables III.1 through III.6.

To compare the perceptions of front-line employees and managers, we consolidated the responses of front-line employees (revenue agents, tax auditors, and revenue officers) and managers (Examination Division and Collection Division group managers) as shown in table III.1.

Table III.1: Estimated Percent of Front-Line Employees and Managers Who Perceived That Tax Enforcement Results Were Considered by Their Supervisor or Manager in Preparing Their Most Recent Performance Evaluation

	Percent	
Employee group	**One or more factors considered**	**Factor not considered/ Do not know if factor(s) considered**
Front-line employees	75	25
Managers	81	19

Source: GAO analysis of employee survey results.

Table III.2: Estimated Percent of Revenue Agents Who Perceived Tax Enforcement Results Were Considered by Their Supervisor or Manager in Preparing Their Most Recent Performance Evaluation

	Percent		
Tax Enforcement Result	To no extent/ do not know[a]	To some/ moderate extent	To great/ very great extent
Average hours spent per return	27	30	43
Additional dollars proposed per return	39	24	37
Additional dollars proposed per hour	40	25	36
Number of criminal and civil fraud referrals	57[c]	28	15
No-change rate for examined returns	53	26	21
Number of work units closed	41	32	27
Number of first contact closures[b]	80	12	8

Note: Percents may not add to 100 percent due to rounding.

[a]We assumed the behavior of employees who responded “To no extent” to be the same as that of employees who responded “Do not know” because neither category would be encouraged to take inappropriate case action to receive a positive evaluation.

[b]The category “Number of first contact closures” refers to the number of times the IRS employee was able to close the case based on information provided during the first contact with the taxpayer.

[c]Sampling error is plus or minus 10.15 percent.

Source: GAO analysis of employee survey results.

Table III.3: Estimated Percent of Tax Auditors Who Perceived Tax Enforcement Results Were Considered by Their Supervisor or Manager in Preparing Their Most Recent Performance Evaluation

	Percent		
Tax enforcement result	To no extent/ do not know[a]	To some/ moderate extent	To great/ very great extent
Average hours spent per return	22	36[d]	42[e]
Additional dollars proposed per return	38[c]	31	31
Additional dollars proposed per hour	37[c]	25	38[e]
Number of criminal and civil fraud referrals	33	42[d]	25
No-change rate for examined returns	43[c]	38[d]	19
Number of work units closed	31	36[d]	33
Number of first contact closures[b]	33	40[d]	26

Note: Percents may not add to 100 percent because not all respondents completed each tax enforcement result category.

[a]We assumed the behavior of employees who responded “To no extent” to be the same as that of employees who responded “Do not know” because neither category would be encouraged to take inappropriate case action to receive a positive evaluation.

[b]The category “Number of first contact closures” refers to the number of times the IRS employee was able to close the case based on information provided during the first contact with the taxpayer.

[c]Sampling errors range from plus or minus 10.15 percent to plus or minus 10.40 percent.

[d]Sampling errors range from plus or minus 10.07 percent to plus or minus 10.37 percent.

[e]Sampling errors range from plus or minus 10.21 percent to plus or minus 10.44 percent.

Source: GAO analysis of employee survey results.

Table III.4: Estimated Percent of Revenue Officers Who Perceived Tax Enforcement Results Were Considered by Their Supervisor or Manager in Preparing Their Most Recent Performance Evaluation

	Percent		
Tax enforcement result	To no extent/ do not know[a]	To some/ moderate extent	To great/ very great extent
Average dollars collected per return	71	18	11
Number or returns secured	67	23	10
Number of installment agreements obtained	71	23	6
Number of installment agreements defaulted	80	15	5
Number of levies filed	67	15	17
Number of liens filed	66	17	18
Number of seizures	60	16	24
Number of cases closed	41	21	37
Number of cases closed as “currently not collectable”	64	30	6
Number of fraud referrals	74	18	8

Note: Percents may not add to 100 percent because not all respondents completed each tax enforcement result category or due to rounding.

[a]We assumed the behavior of employees who responded “To no extent” to be the same as that of employees who responded “Do not know” because neither category would be encouraged to take inappropriate case action to receive a positive evaluation.

Source: GAO analysis of employee survey results.

Table III.5: Estimated Percent of Examination Division Group Managers Who Perceived Tax Enforcement Results Were Considered by Their Supervisor or Manager in Preparing Their Most Recent Performance Evaluation

	Percent		
Tax enforcement result	To no extent/ do not know[a]	To some/ moderate extent	To great/ very great extent
Average hours spent per return	26	33	41
Additional dollars proposed per return	28	32	40
Additional dollars proposed per hour	24	22	54
Number of criminal and civil fraud referrals	37	37	26
No-change rate for examined returns	38	43	19
Number of work units closed	42	30	28
Number of first contact closures[b]	77	13	10

[a]We assumed the behavior of employees who responded “To no extent” to be the same as that of employees who responded “Do not know” because neither category would be encouraged to take inappropriate case action to receive a positive evaluation.

[b]The category “Number of first contact closures” refers to the number of times the IRS employee was able to close the case based on information provided during the first contact with the taxpayer.

Source: GAO analysis of employee survey results.

Table III.6: Estimated Percent of Collection Division Group Managers Who Perceived Tax Enforcement Results Were Considered by Their Supervisor or Manager in Preparing Their Most Recent Performance Evaluation

	Percent		
Tax enforcement result	To no extent/ do not know[a]	To some/ moderate extent	To great/ very great extent
Average dollars collected per return	61	20	19
Number or returns secured	61	25	14
Number of installment agreements obtained	61	32	7
Number of installment agreements defaulted	84	11	5
Number of levies filed	71	18	11
Number of liens filed	76	17	7
Number of seizures	57	25	18
Number of cases closed	23	35	42
Number of cases closed as “currently not collectable”	55	34	11
Number of fraud referrals	61	29	9

Note: Percents may not add to 100 percent because not all respondents completed each tax enforcement result category.

[a]We assumed the behavior of employees who responded “Do not know” to be the same as that of employees who responded “To no extent.”

Source: GAO analysis of employee survey results.

Survey Results on the Basis for Employee Perceptions That Enforcement Results Affect Evaluations

For survey respondents who perceived that their supervisor or manager considered one or more tax enforcement results in preparing their most recent performance evaluation, we asked “Why do you believe that one or more of the factors for which you checked box 2 through 5 [To some extent, To a moderate extent, To a great extent, To a very great extent] in [the prior question] influenced your rating?” They were asked to check all the reasons that applied. The responses are shown in tables IV.1 through IV.8.

To compare the perceptions of front-line employees and managers, we consolidated the responses of front-line employees (revenue agents, tax auditors, and revenue officers) and managers (Examination Division and Collection Division group managers) as shown in table IV.1. To compare the importance of verbal and written communications, we also consolidated the responses indicating verbal communications (feedback and meetings) and written communications (performance expectations, performance evaluations, and award justifications). Two other categories—promotion patterns and a general “other” category—are reported separately.

Table IV.1: Estimated Percent of Front-Line Employees and Managers Whose Perception Was Based on Verbal and/or Written Communication

	Percent	
Employee category	Verbal communication	Written communication
Front-line employees	70	49
Managers	83	51

Source: GAO analysis of employee survey results.

The responses of front-line employees and managers on the specific basis for their perception that tax enforcement results affected their ratings are given in tables IV.2 through IV.4.

Table IV.2: Estimated Percent of Front-Line Employees and Managers Who Had Various Bases for Their Perception That Tax Enforcement Results Influenced Their Evaluation

Basis for employees' and managers' perception that tax enforcement results influenced evaluation	Front-line employees (percent)	Managers (percent)
One or more factors were mentioned in my performance expectations	29	27
One or more factors were mentioned in my performance evaluation	36	41
One or more factors were mentioned in feedback I received from my group/branch manager	52	69
One or more factors were mentioned by my group/branch manager in group meetings	56	70
One or more factors were mentioned in an award justification	13	18
One or more factors were mentioned in a promotion justification	6	2
Promotion patterns suggest that management considers one or more factors	37	19
Other - Please describe[a]	30	31

[a]The category "other" is discussed in greater detail in appendix VI.

Source: GAO analysis of employee survey results.

Table IV.3: Estimated Percent of Revenue Agents, Tax Auditors, and Revenue Officers Who Had Various Bases for Their Perception That Tax Enforcement Results Influenced Their Evaluation

Basis for employees' perception that tax enforcement results influenced evaluation	Revenue agents (percent)	Tax auditor (percent)	Revenue officer (percent)
One or more factors were mentioned in my performance expectations	27	30[b]	34
One or more factors were mentioned in my performance evaluation	35	29[b]	41
One or more factors were mentioned in feedback I received from my group manager	50	48[b]	56
One or more factors were mentioned by my group manager in group meetings	55	74	50
One or more factors were mentioned in an award justification	13	14	11
One or more factors were mentioned in a promotion justification	5	1	9
Promotion patterns suggest that management considers one or more factors	36	23	43
Other - Please describe[a]	29	23	35

[a]The category "other" is discussed in greater detail in appendix VI.

[b]Sampling errors range from plus or minus 10.21 percent to plus or minus 10.35 percent.

Source: GAO analysis of employee survey results.

Table IV.4: Estimated Percent of Examination and Collection Division Group Managers Who Had Various Bases for Their Perception That Tax Enforcement Results Influenced Their Evaluation

Basis for managers' perception that tax enforcement results influenced evaluation	Examination managers (percent)	Collection managers (percent)
One or more factors were mentioned in my performance plan (Form 9688)	29	23
One or more factors were mentioned in my performance evaluation	44[b]	36
One or more factors were mentioned in feedback I received from my branch chief	76	58
One or more factors were mentioned by my division chief or branch chief in group meetings	71	68
One or more factors were mentioned in an award justification	17	21
One or more factors were mentioned in a promotion justification	3	2
Promotion patterns suggest that management considers one or more factors	21	16
Other - Please describe[a]	27	38

[a]The category "other" is discussed in greater detail in appendix VI.

[b]Sampling error is plus or minus 10.64 percent.

Source: GAO analysis of employee survey results.

Survey Results on Inclusion of Enforcement Results in Group Goals

Our survey asked a representative sample of IRS employees "To your knowledge, were performance goals established for your group in calendar year 1997?"

To compare the perceptions of front-line employees and managers, we consolidated the responses of front-line employees (revenue agents, tax auditors, and revenue officers) and managers (Examination Division and Collection Division group managers) as shown in table V.1.

Table V.1: Estimated Percent of Front-Line Employees and Managers Who Had Performance Goals Established for Their Group in Calendar Year 1997

Employee category	Percent
Front-line employees	44
Managers	49
Revenue agents	44
Tax auditors	46[a]
Revenue officers	46
Examination Division group managers	45
Collection Division group managers	54

[a]Sampling error is plus or minus 10.54 percent.

Source: GAO analysis of employee survey results.

For those employees who had group performance goals, we asked "Did the performance goals established for your group in calendar year 1997 include any reference to any of the following factors?" The point estimates and confidence intervals for each employee group are shown in tables V.2 and V.3.

Table V.2 Estimated Percent of Revenue Agents, Tax Auditors, and Examination Group Managers Who Had Performance Goals Established for Their Group That Included at Least One Tax Enforcement Result

Tax enforcement result	Revenue agent (percent)	Tax auditor (percent)	Examination manager (percent)
Average hours spent per return	77	89	60[d]
Additional dollars proposed per return	54[b]	55[c]	53[d]
Additional dollars proposed per hour	76	66[c]	70[d]
Number of criminal and civil fraud referrals	38	59[c]	47[d]
No-change rate for examined returns	50[b]	51[c]	23[d]
Number of units closed	49[b]	74[c]	28[d]
Number of first contact closures[a]	11	57[c]	7

[a]The category “number of first contact closures” refers to the number of times the IRS employee was able to close the case based on information provided during the first contact with the taxpayer.

[b]Sampling errors range from plus or minus 10.35 percent to plus or minus 10.90 percent.

[c]Sampling errors range from plus or minus 13.80 percent to plus or minus 15.90 percent.

[d]Sampling errors range from plus or minus 12.21 percent to plus or minus 14.41 percent.

Source: GAO analysis of employee surveys.

Table V.3: Estimated Percent of Revenue Officers and Collection Group Managers Who Had Performance Goals Established for Their Group That Included at Least One Tax Enforcement Result

Tax enforcement result	Revenue officer (percent)	Collection manager (percent)
Average dollars collected per return	40	43[a]
Number of returns secured	33	30[a]
Number of installment agreements obtained	11	11
Number of installment agreements defaulted	6	2
Number of levies filed	23	8
Number of liens filed	16	6
Number of seizures	32	15
Number of cases closed	70	70
Number of cases closed as “currently not collectable”	24	15
Number of fraud referrals	13	8

[a]Sampling errors range from plus or minus 10.01 percent to plus or minus 10.63 percent.

Source: GAO analysis of employee survey results.

Examples of Voluntary Employee Comments Included on Surveys of IRS Employees

This appendix gives some examples of voluntary written comments by front-line employees and group managers to our questionnaires on why they believe enforcement results influenced their performance evaluations.[2] The comments generally involved the following scenarios: (1) supervisors or managers implied that an employee's performance evaluation was influenced by tax enforcement results; (2) the general culture or atmosphere of IRS implied that enforcement results affected performance evaluations; (3) higher level managers pressured group managers to increase production or revenue yield; and (4) IRS' business plan, reports, and other documents emphasized enforcement results as goals. Front-line employees and managers also wrote about positive and negative effects of using enforcement and workload measures to establish group goals or in performance evaluations. Generally, the number of negative comments made by front-line employees and managers exceeded the number of positive comments.

Comments Indicating That Supervisors or Managers Implied That an Employee's Performance Evaluation Was Influenced by Enforcement Results

- "The district office provided monthly statistics to each branch, rating branches and groups based on 'dollars per hour,' 'dollars per return,' 'no-change rate,' [and] 'collections percent by exam division.' . . . Even though performance goals were not established formally based on 'dollars per return,' etc., the statistics were mentioned at each group meeting and pressure was felt by each agent to assess the most tax at the least amount of time spent on the case. It was always felt that our evaluations were based indirectly on the factors mentioned in Item 19 (productivity and enforcement statistics)."
- "Our emphasis used to be on quality and customer service but when our district went under Boston, the emphasis changed to 'dollars per hour.' At each group meeting, we would hear about 'stats' and 'dollars per hour.' This kind of pressure does tend to carry over into the work product."
- "Although numbers [were] never mentioned, [my] group manager discussed closure of cases, fraud referrals, and defaulted installment agreements in meetings and in appraisals so it gave the appearance that [is] what he expected."
- "We were told by management in group meetings and in town hall meetings that goals in the form of yield (dollars per hour) would no longer be set. Yet, at group meetings, our manager reads our yield from Table 37 [a management information report] and comments on whether it is good, bad, or fair. Our yield is obviously still very important to management."

[2]Of the 814 surveys received, a total of 402 front-line employees and 151 group managers added voluntary comments.

Comments Indicating That the General Culture or Atmosphere of IRS Implied That Enforcement Results Affected Performance Evaluations

- "There is one important goal in any revenue agent's life—dollars per hour. From the day I started with the Service until the present, that has been the only constant in the organization. Quality programs have come and gone, emphasis on taxpayer service ebbs and flows, lip service to auditing standards is periodically given. Through it all, any successful revenue agent knows that low time and high dollars will result in recognition, promotion, and awards."
- "It is well known that these factors are considered by upper management to be important, although not all are mentioned in evaluations."
- "Management has never been specific relative to the connection between promotion and enforcement. However, the perception is here that the more enforcement minded a revenue officer is, the greater the chance of promotion."
- "You are expected (unsaid) to do a seizure every quarter."

Comments Indicating That Higher Level Managers Pressured Group Managers to Increase Production or Yield

- ". . . [My present manager] was required to post statistics on most of these items on our library room wall up to just a few months ago. When Congress started seriously making noises about statistics, they were taken down. . . . [My manager] would make reference to these statistics in a general way as pressure was put on him through his chain of command. The pressure relative to improving these statistics came primarily by passing word down from the chief of audit and his assistant chief that our statistics were not measuring up. . . ."
- "The manager only works toward faster closures and higher yield because it has been stressed by the district management."

Comments Indicating That IRS' Business Plan, Reports, and Other Documents Emphasized Enforcement Results

- "Time and dollars per hour, although less so within the past few years, has always been a major consideration in 'informal' discussions and meetings. Tables 36 and 37 [management information reports] were always utilized and disbursed to the group and discussed with agents, although seldom in writing."
- "Statistics on dollars collected, number of returns [secured], number of levies and liens were reported by group on a monthly basis to the division via the branch chief."
- "Monthly statistics . . . were distributed regarding levies, dollars collected, liens, seizures, fraud referrals, currently not collectible cases, and cases closed."

Positive Comments on the Use of Enforcement Results

- “If a revenue agent is doing a good job, these factors (adjustments, etc.) are present. The revenue agent will have good adjustments that are technically correct and will also have no-changes. That is part of the job.”
- “Both time spent on a return and deficiencies are indicators of effectiveness and efficiency in the audit process. Excessive time may indicate complexity of issues or inefficiency and floundering in the audit. The deficiency amount will indicate effectiveness of audit as compared to similar audits or, potentially, a lack of issue identification. When used as indicators of efficiency or effectiveness, these can help direct employees toward improvement. It’s not all bad.”
- “Our job is to collect (hence ‘collection’) dollars and returns. It is ludicrous to think that a revenue officer who collected no dollars or returns would get a good performance evaluation. After all—that is our job!”
- “I definitely feel pressure to be productive in terms of dollars per hour, but realistically, that makes sense to me. We should spend time on issues which will most likely produce revenue and work them as quickly as possible.”

Negative Comments on the Use of Enforcement Results

- “Statistics regarding ‘dollars per hour’ are disseminated at group meetings. Our group is compared to the branch and the district. . . . Ten years ago these kinds of statistics were never discussed at the group level. Due to the constant referring to these numbers, it is obvious that ‘dollars per hour’ is the most important concept to management. Consequently, agents will always strive to get the highest yield per case. This means that you will disregard adjustments in the taxpayers favor so as not to reduce the tax yield. Individual statistics have not been discussed. However, the constant reminder of the yield causes agents to do anything they can to get good yield on their cases.”
- “Even though it is made very clear in the . . . region that there are no individual or group enforcement goals, statistics are kept and very well known throughout the region. Specifically, our ‘mission’ is very goal oriented and very driven by the goals of the Chief of Collections. Very clearly those groups that have good statistics—specifically the number of seizures, cases closed, fraud referrals, etc.—receive special recognition. In some cases, managers receive performance rewards. The revenue officers who specifically take enforcement action to resolve cases are rewarded with promotion even though the enforcement action is not mentioned in the performance evaluation. Those revenue officers who are most aggressive using enforcement are those rewarded most quickly with promotion.”

- “National and regional management have used dollar goals and quotas for as long as I can remember. It is common knowledge among employees and first-line managers that ‘increase voluntary compliance’ equals ‘dollars per hour.’ You get what you measure.”
- “Within the last couple of years, ‘dollars collected’ was being conveyed from the division to the group level. This caused group comparisons and competition. While never conveyed to an individual within my group that dollars needed to be increased, it certainly was implied. While this didn’t change my case decisions, I cannot speak for everyone.”

Results of Review of Employee Evaluations

Table VII.1 provides the results of our review of employee evaluations in terms of whether the evaluations contained law or policy violations and case or general discussions of enforcement-related activity.

Table VII.1: Estimated Percent of Employees Whose Performance Evaluations Contained References to Tax Enforcement Results

	Percent of employees whose evaluations included tax enforcement results references		
Types of tax enforcement results references	**Collection employees**	**Examination employees**	**Total**
Law or policy violations	1	12	**9**
Case-specific discussion of:			
1. Overage cases	2	12	**6**
2. Cycle time	1	2	**2**
3. Overage cases or cycle time	2	9	**7**
4. Assessment/Collection measures	33	31	**32**
A. Amount of revenue collected	14	N/A	**N/A**
B. Use of collection tools	33	N/A	**N/A**
C. Amount of revenue assessed	N/A	25	**N/A**
D. Amount of dollars collected on agreed cases	N/A	12	**N/A**
Any specific discussion of the above	33	38	**37**
General discussion of:			
1. Overage cases	31	33	**32**
2. Cycle time	9	29	**16**
3. Overage cases or cycle time	32	34	**33**
4. Assessment/collection measures	75	20	**36**
A. Amount of revenue collected	9	N/A	**N/A**
B. Use of collection tools	74	N/A	**N/A**
C. Amount of revenue assessed	N/A	17	**N/A**
D. Amount of dollars collected on agreed cases	N/A	5	**N/A**
Any general discussion of the above	78	54	**61**
Any discussion of overage cases or cycle time	38	43	**41**
Any discussion of assessment/collection measures	75	40	**50**
Any discussion of any factors above	78	65	**69**

Note: N/A means not applicable to this employee category.

Source: GAO analysis of IRS employees' performance evaluations.

Comments From the Internal Revenue Service

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

COMMISSIONER

November 16, 1998

Mr. James R. White
Director, Tax Policy and Administration Issues
United States General Accounting Office
Washington, D.C. 20548

Dear Mr. White:

Thank you for the opportunity to review and comment on your recent draft report entitled "IRS Personnel Administration: Use of Enforcement Statistics in Employee Evaluations."

While this review was being conducted, the Internal Revenue Service (IRS) was undertaking a substantial effort to review all instructions it had issued on the use of enforcement statistics and to draft revised guidance that was uniform among the various offices within the IRS and could be easily understood by both managers and non-supervisory employees. A task force composed of representatives from throughout IRS completely revised the Internal Revenue Manual (IRM) provisions concerning the use of enforcement statistics and those revised IRM provisions have now been issued. In addition, the IRS conducted training for all IRS managers on the use of enforcement statistics and has provided an orientation course regarding the proper use of enforcement statistics to all IRS employees.

To assist IRS employees better understand the rules, IRS also established a Support Panel that is responsible for answering all questions it receives from either managers or employees on the proper use of enforcement statistics. This Support Panel thus provides a direct source of information for anyone that does not fully understand the provisions of P-1-20 and the TBOR. The panel is made up of functional representatives and Chief Counsel. The panel can be contacted by electronic mail or facsimile. Final answers to questions presented to the special Panel are posted on the IRS Intranet and are thus available to all IRS personnel.

Finally, IRS has also established an independent review process to monitor compliance with our restrictions on the use of enforcement statistics. As a part of this process, the Independent Review Team that will conduct this review is obliged to review written documentation in Employee Performance Files and Employee Evaluations to ensure compliance with IRS policies. The Independent Review Team, at the district director's discretion, may also look to other sources of information, including group discussions, National Treasury Employees Union comments or reports, grievances, minutes of manager's meetings, etc., to ensure that any potential violations are uncovered. This peer review process will provide a mechanism to uncover any misuse of enforcement statistics in evaluating managers and employees. The peer review team will visit each of the 33 districts within a 3-year cycle. Accordingly, we have provided the directors with the flexibility they will need to conduct as extensive a review as they deem appropriate.

While we have accomplished quite a lot since the use of enforcement statistics received attention during the Senate Finance Committee hearings, we appreciate having your views as to additional actions we might pursue. The report prepared by your office recommends that guidance be expanded to include additional examples of the appropriate and inappropriate use of records of tax enforcement results (ROTERs) in written evaluations. Many steps are already underway to support this recommendation. The use of ROTERs in written evaluations was covered extensively in the "Managing Statistics for IRS Managers" training course. The course contains a discussion on the use of ROTERs in evaluations and numerous examples of appropriate and inappropriate language. This training was mandatory for all managers of enforcement officers.

Nonetheless, we agree that more guidance would better ensure managerial accountability. Our IRM will be further revised to include additional examples of appropriate and inappropriate use of ROTERs in written evaluations. This material will also be incorporated into the training provided to new managers.

As you have recommended, we will revise the Quarterly Certifications to specifically state that tax enforcement results were not used in any written performance evaluation prepared or reviewed and that the manager did not verbally communicate to employees that tax enforcement results affected their evaluations or were used to set individual production goals.

We also agree with recommendations three and four as stated in the report. The Penalty Guide will be revised to specifically list the disciplinary actions that can be taken for violations (recommendation three). Further, we will survey employees periodically to determine whether they perceive that tax enforcement results were used in written evaluations or verbally communicated by their supervisors and use the results to assess whether IRS needs to further clarify the guidance, provide additional training, or take any other appropriate action (recommendation four).

3

Finally, we agree with recommendation five that we need to clarify for front-line managers and employees how all outcome neutral performance data, such as overage and cycle-time data, may be used appropriately. The training presented to managers included instructions on the proper treatment of cycle time and overage statistics. Additionally, we will revise our Internal Revenue Manual instructions to include examples of how overage and cycle-time data may be used appropriately.

In closing, let me mention one other broader initiative that we currently are undertaking. As you know, the IRS Restructuring and Reform Act of 1998 (RRA) was enacted in July of this year. Sections 1201 and 1204 of RRA represent the most recent legislative action regarding performance measures used by the IRS. Section 1201 directs the IRS to establish a performance management system that will establish goals or objectives for individual, group, or organizational performance. The IRS is directed to use this performance system in the evaluation of employees or groups of employees, in determining salary adjustments and awards, and in other personnel matters. As mentioned in your draft report, section 1204 of RRA repealed the Taxpayer Bill of Rights (TBOR) 1 prohibitions on using tax enforcement results to evaluate or impose or suggest goals for collection employees and instead expanded the prohibitions to cover all IRS employees.

We are taking steps to develop and establish a balanced performance measurement system and to provide guidance for implementing the restrictions on the use of records of tax enforcement results. At this time, we are developing a balanced measurement system that will contain three elements: (1) customer satisfaction measures; (2) employee satisfaction measures; and (3) business results measures. The business results measures will employ two parallel avenues to measure business results. The first will focus on the quality of the work done and the second will focus on the quantity of work done. Data used in this quantity measure will consist exclusively of outcome-neutral production data which do not reflect the outcome produced by any IRS official's exercise of judgment in determining liability for tax or the appropriate collection mechanism to be employed. Thus, the quantity measure can only be based on data such as cases closed, time per closing, or cycle time, rather than on data reflecting enforcement outcomes, e.g., examination adjustment proposed, penalties imposed, dollars collected, number of liens filed and number of levies served.

We believe the new balanced measurement system will channel IRS efforts in positive directions by providing guidance on what should be measured rather than what should

4

not be measured. Even more, the existence of balanced measures will, we hope, eliminate even the perception that IRS is driven by enforcement goals. We will be seeking public comment on our new system before it becomes operational.

Sincerely,

Charles Rossotti

Charles O. Rossotti

Enclosure

Response to Recommendations from GAO
"IRS Personnel Administration: Use of Enforcement Statistics in Employee Evaluations"

RECOMMENDATION #1: To better ensure managerial accountability for the proper use of tax enforcement results, we recommend that the Commissioner expand the guidance to include additional examples of the appropriate and inappropriate use of records of tax enforcement results in written evaluations.

Response: We agree with the recommendation. We will revise our Internal Revenue Manual instructions to include additional examples of appropriate and inappropriate use of ROTERs in written evaluations. These examples will also be incorporated into the training provided to new managers.

RECOMMENDATION #2: Revise the quarterly certification form to specifically state that tax enforcement results were not used in any written performance evaluation prepared or reviewed, including appraisals, awards, or promotion justifications, and that the manager did not verbally communicate to employees that tax enforcement results affected their evaluations or were used to set individual production goals or quotas.

Response: We agree with the recommendation. The Quarterly Certifications will be revised to include the language in the recommendation or similar language.

RECOMMENDATION #3: Revise the Penalty Guide to specifically list the disciplinary actions that can be taken for violations.

Response: We agree with the recommendation. The IRS' Office of Management and Finance has agreed to revise the Penalty Guide to include the recommended list of disciplinary actions for violations of the use of statistics.

RECOMMENDATION #4: Survey employees periodically to determine whether they perceive that tax enforcement results were used in written evaluations or verbally communicated by the supervisors and use the results to assess whether IRS needs to further clarify the guidance, provide additional training, or take any other appropriate action.

Response: We agree with the recommendation. The IRS' Office of Management and Finance will survey employees periodically to determine whether they perceive that tax enforcement results were used in written evaluations or verbally communicated by the supervisors. The results will be provided to the appropriate organization to assess whether IRS needs to further clarify the guidance, provide additional training, or take any other appropriate action. The survey instrument will be the IRS Corporate Climate Survey, to be administered March 1999, and annually thereafter.

2

RECOMMENDATION #5: To avoid the potential inappropriate use of overage and cycle-time data, the IRS Commissioner should either designate overage and cycle-time data as prohibited tax enforcement results or emphasize in official policies or procedures to front-line employees and managers how overage and cycle-time data may be used appropriately.

Response: We agree that the IRS Commissioner should emphasize through procedures to front-line employees and managers how overage and cycle-time may be used appropriately. We do not agree that overage and cycle-time should be identified as records of tax enforcement results because such data is not an outcome produced by an employee exercising judgment with regard to determining tax liability or the ability to pay. Rather, they are process measures which help ensure prompt customer service. We will revise our Internal Revenue Manual instructions to include examples of how overage and cycle-time data may be used appropriately.

Examples of Certification Forms

Exhibit 105.4.10-1
Taxpayer Bill of Rights (TBOR) Section 6231 Certification

Taxpayer Bill of Rights (TBOR), Section 6231 Certification

I certify that I have not used records of tax enforcement results in a manner prohibited by TBOR during the quarter ending ________ except as noted below.

☐ **I certify that, to the best of my knowledge, the certifications attached hereto from my subordinate managers are true and correct.**

Violations:

Corrective Actions:

Print Name	Title and Symbols
Signature	Date

Managing Statistics 105.4 page 10-9

Exhibit 105.4.10-2
P-1-20 Certification

P-1-20 Certification

I certify I have taken no action in violation of P-1-20 during the quarter ending ______ except as noted below, or on the attached TBOR Certification sheet.

☐ **I certify that, to the best of my knowledge, the certifications attached hereto from my subordinate managers are true and correct.**

Violations:

Corrective Actions:

Print Name ____________________ Title and Symbols ____________________

Signature ____________________ Date ____________________

IRM Handbook MT 105.4-XX (**/**/**) 10.5.4

Major Contributors to This Report

General Government Division

James A. Bell, Senior Statistician
James M. Fields, Senior Statistician
Stuart M. Kaufman, Senior Social Science Analyst

San Francisco Office

Ralph T. Block, Assistant Director
Jonda Van Pelt, Evaluator-in-Charge
Shari Caporale, Senior Evaluator
Suzy Foster, Senior Evaluator
Sam Scrutchins, Senior Evaluator
John N. Zugar, Senior Evaluator

Atlanta Office

Robert V. Arcenia, Senior Evaluator
Ronald J. Heisterkamp, Evaluator

Kansas City Office

Benjamin Douglas, Senior Evaluator

Ordering Information

The first copy of each GAO report and testimony is free. Additional copies are $2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. VISA and MasterCard credit cards are accepted, also. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

Orders by mail:

U.S. General Accounting Office
P.O. Box 37050
Washington, DC 20013

or visit:

Room 1100
700 4th St. NW (corner of 4th and G Sts. NW)
U.S. General Accounting Office
Washington, DC

Orders may also be placed by calling (202) 512-6000
or by using fax number (202) 512-6061, or TDD (202) 512-2537.

Each day, GAO issues a list of newly available reports and testimony. To receive facsimile copies of the daily list or any list from the past 30 days, please call (202) 512-6000 using a touchtone phone. A recorded menu will provide information on how to obtain these lists.

For information on how to access GAO reports on the INTERNET, send an e-mail message with "info" in the body to:

info@www.gao.gov

or visit GAO's World Wide Web Home Page at:

http://www.gao.gov

United States
General Accounting Office
Washington, D.C. 20548-0001

Bulk Rate
Postage & Fees Paid
GAO
Permit No. G100

Official Business
Penalty for Private Use $300

Address Correction Requested