

GAO

Testimony

For Release
on Delivery
Expected at
10:00 a.m. EST
Wednesday
March 15, 1989

IRS' Telephone Assistance Program

Statement of
Jennie S. Stathis, Director
Tax Policy and Administration Issues

Before the
Subcommittee on Commerce,
Consumer, and Monetary Affairs
Committee on Government Operations
House of Representatives



0-4-89: 138177

Mr. Chairman and Members of the Subcommittee:

We are pleased to be here today to discuss IRS' Integrated Test Call Survey System and the work we are doing at your request to monitor this system. The Test Call System is designed to measure the accuracy of answers IRS gives taxpayers who call one of IRS' toll-free telephone numbers. In particular you asked that we comment today on the results IRS publicly reported last week on the accuracy of its test calls to date.

At a March 10 press conference, IRS reported an error rate of 30.8 percent for toll-free telephone assistance. IRS said that this meant that 30.8 percent of its answers would lead taxpayers to a wrong result on their returns. Conversely, it meant that 69.2 percent were receiving right answers. IRS indicated that it was not satisfied with this performance and that clearly it needed to improve its responses to taxpayers.

As you know we have worked with IRS over the past year to develop a testing system designed to produce a valid measure of the accuracy of IRS' responses to taxpayers' questions in the area of individual tax law. A major part of our effort has been spent in developing a reasonable set of questions in various frequently asked tax law categories and reaching agreement with IRS on the answers that would warrant a correct score for each question. After considerable debate, we reached agreement on 62 questions

and specific responses that would be categorized as correct or correct and complete. A correct answer was to be the minimal standard IRS expected its assistors to meet. Answers which exceeded this standard would be classified as correct and complete. It was agreed that all other answers would be scored incorrect.

The figures that IRS reported last week were not consistent with the agreements we reached in three respects. First, they covered only the last 2 weeks of what was at that time a 4-week test. Second, the results were based on only 29 of the 62 test questions. Finally, IRS used a more liberal scoring of responses to certain questions than we had agreed on. Using the agreed on scoring for the same time period and the same set of questions would show that assistors correctly responded 60.4 percent rather than 69.2 percent of the time.

About two weeks after the start of the test, IRS determined that for certain questions assistors were providing answers that IRS believed were "not wrong" but nevertheless failed to meet the standard for a correct answer. Early IRS results indicated that scores for these questions were very poor. IRS officials said that it would be unfair to imply to the Congress or the public that assistors were providing wrong answers if that advice would not necessarily lead taxpayers to file inaccurate tax returns. Thus, IRS devised another category of response -- "right" answers

-- which failed to meet minimal standards but which they proposed to add to "correct and complete" and "correct" categories in reporting accuracy statistics. In computing the accuracy rates reported on March 10, IRS added this new category of response -- "right" answers -- to the three categories that we had mutually agreed on.

We did not agree to revise the scoring. The proper scoring of answers to these questions had been exhaustively discussed in prior meetings in which many of the same considerations surfaced; yet agreements had been reached on what would be correct answers for each question. Subsequently, we had agreed to clarify some questions and revise some scoring based on initially low accuracy rates which appeared to be caused by factors other than assistor errors. IRS identified these situations from calls made prior to the official start of the test.

We are against further revisions for the following reasons.

-- Many of the answers IRS wants to score as "right" are incomplete and potentially misleading. Such answers would increase the likelihood of taxpayers making errors.

-- Once a test is underway, it is important for statistical validity that the standards remain as

consistent as possible.

-- It is important to maintain consistent scoring for questions used both last year and this year to determine whether or not accuracy improves.

-- It is important to avoid a repeat of last year's experience when IRS, well into its test, dropped some questions and revised the scoring for some others. This raised an issue regarding the integrity of the test in that it created the perception that IRS was manipulating the statistics.

Mr. Chairman, you asked us to monitor the administration of IRS' test and to report to you on the validity of the statistics IRS produces. Overall, we believe that the 62 questions and the scoring standards that we agreed to, if properly implemented, would produce a reasonable measure of the accuracy of assistors' answers in the area of individual tax law. The test as designed requires a limited amount of probing by assistors to ferret out essential facts before answering most questions; but it does not require knowledge of complex or obscure areas of individual tax law. However, IRS' tact of changing the agreed upon scoring methodology this far into the test raises questions about the validity of its reported results and whether it is making progress toward improving the accuracy of its responses.

That concludes my prepared statement, Mr. Chairman. We will be pleased to respond to any questions.