

GAO

Report to the Chairman, Subcommittee
on Commerce, Consumer, and Monetary
Affairs, Committee on Government
Operations, House of Representatives

January 1990

TAX
ADMINISTRATION

Monitoring the
Accuracy and
Administration of IRS'
1989 Test Call Survey



General Government Division

B-234202

January 4, 1990

The Honorable Doug Barnard, Jr.
Chairman, Subcommittee on Commerce,
Consumer, and Monetary Affairs,
Committee on Government Operations,
House of Representatives

Dear Mr. Chairman:

This report responds to your request that we evaluate the Internal Revenue Service's (IRS) administration of its Integrated Test Call Survey System (ITCSS) during the 1989 tax filing season. ITCSS was designed to measure the quality of service IRS provides through its toll-free telephone system—a nationwide system in which IRS assistors answer taxpayers' telephone inquiries. To accomplish this purpose, IRS designed a survey sample to produce statistical estimates on the accuracy of its telephone assistors in answering a set of 62 tax law test questions. These test questions were developed for tax law areas in which IRS determined that individual taxpayers commonly make inquiries when preparing their tax returns. IRS administered the test by placing anonymous calls to its telephone assistors and scoring their responses to the test questions.

You requested that we report to you on IRS' administration of its 1989 test call survey and on the validity of the statistical estimates produced during the test. To respond to your request, we monitored and independently scored a statistically valid random sample of IRS' test calls. As you know, we worked with IRS to develop ITCSS and mutually agreed in advance on what constituted a correct answer for each question. This report evaluates the validity of IRS' overall national accuracy rate. Appendix I provides selected ITCSS filing season results for IRS regions and call sites and by major tax law categories for individual taxpayers.

This report updates and supplements the preliminary results of our work, which we reported in testimony before your Subcommittee on March 15, 1989. We did our work from January 1989 to August 1989 using generally accepted government auditing standards.

Results in Brief

IRS' overall ITCSS results for the 1989 tax filing season showed that IRS telephone assistors responded correctly 62.8 percent of the time to the survey's tax law test questions. On the basis of our monitoring of a statistically valid random sample of test calls, we agree with the overall

telephone assistance accuracy rate IRS reported. Also, overall, IRS fairly administered its test call survey. With few exceptions, IRS test callers (Taxpayer Service employees responsible for making the test calls) asked tax law test questions in a fair manner and scored telephone assistors' responses objectively and accurately.

The test question scoring criteria for correct assistor responses we used in our assessment are those on which we and IRS mutually agreed. During the filing season, however, IRS reported a higher ITCSS accuracy rate that was based on more liberal scoring criteria with which we did not agree. IRS' deviation from the agreed-upon scoring criteria in reporting assistor accuracy was the subject of your March 15, 1989, hearing. In July 1989, the Assistant Commissioner (Taxpayer Service) said that for the 1990 filing season, IRS would only report accuracy rates that were based on scoring criteria mutually agreed upon by IRS and us.

Background

Because our tax laws are complicated, taxpayers often need assistance in understanding the tax laws and in preparing their tax returns. The principal vehicle IRS' Taxpayer Service Division uses to assist taxpayers is a toll-free telephone program. IRS has assisted taxpayers through this program for over two decades. Historically, IRS has considered telephone assistance to be the most efficient method of helping taxpayers. Accordingly, it has devoted substantial staff resources to telephone assistance and encourages taxpayers to use the telephone as a means of getting answers to their tax law questions.

During the 1989 tax filing season, IRS employed over 5,000 telephone assistors at 32 telephone sites. These assistors answered about 18.8 million taxpayer calls on individual and business tax law issues, procedural issues, and account-related matters. IRS' telephone assistors are primarily composed of two groups—frontline and backup assistors. Frontline assistors initially take taxpayers' calls and, if they are unable to answer the taxpayers' questions, refer them to backup assistors who usually have more experience and expertise.

Over the past 2 years, both Congress and the public have raised concerns over the quality of the responses assistors have provided to questions designed to test their tax law knowledge. For the 1988 filing season, we did our sixth survey of assistors' tax law knowledge and

reported that they provided correct responses to our questions 64 percent of the time and incorrect responses 36 percent of the time.¹ Also during the 1988 filing season, IRS implemented ITCSS and found that assistors correctly responded to its test questions over 70 percent of the time. IRS expressed its belief that both our test results and its own indicated an unacceptably low rate of assistor accuracy.

For the 1989 filing season, the Subcommittee indicated interest in our working with IRS to develop a reliable testing system that we could monitor and that would avoid the need for IRS and us to do separate tests. Accordingly, we worked with IRS to develop ITCSS and a monitoring system that would enable us to conclude whether or not IRS properly administered ITCSS and accurately reported on its assistors' performance.

The centerpiece of ITCSS is a 62-question test covering what IRS identified as the seven major individual tax law categories in which taxpayers ask questions. As shown in table 1, these 7 tax law categories contained 32 subcategories of tax law.

¹Tax Administration: Accessibility, Timeliness, and Accuracy of IRS' Telephone Assistance Program (GAO/GGD-89-30, Feb. 2, 1989).

Table 1: Individual Tax Law Categories and Subcategories Tested by IRS — 1989 IRS Test Call Survey

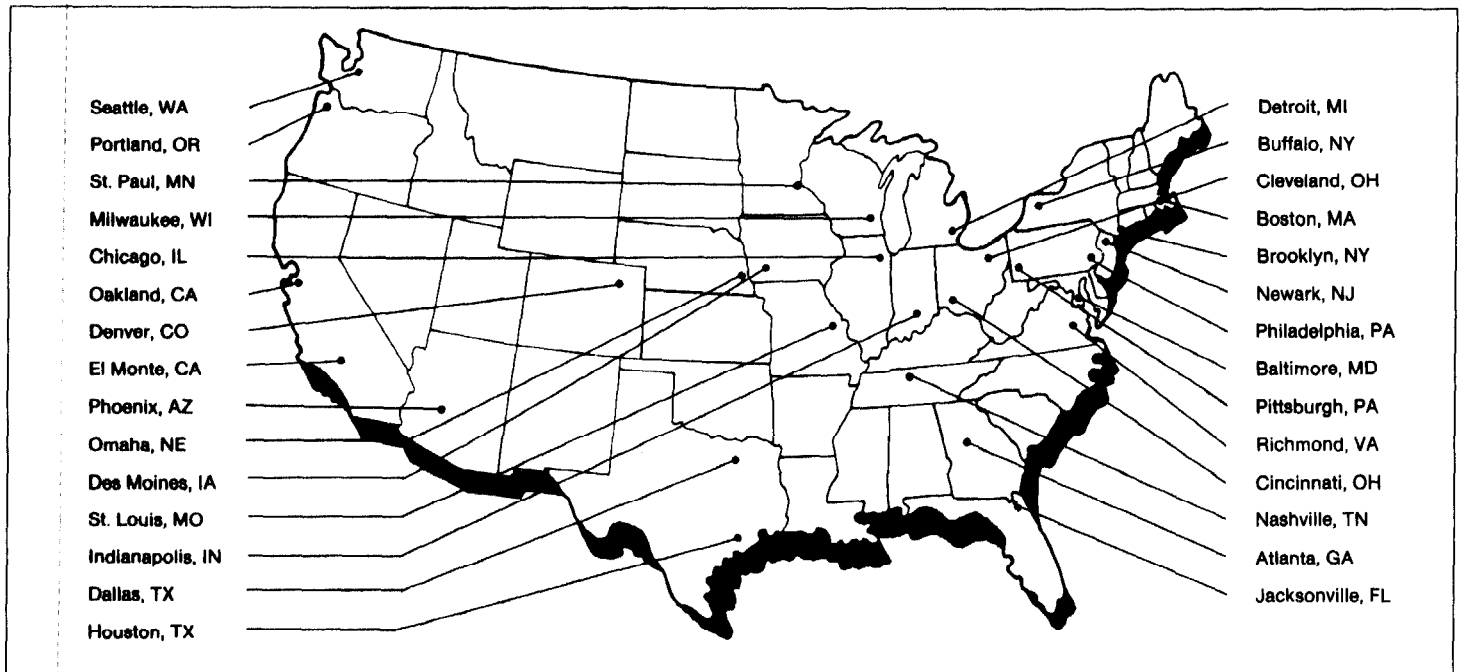
<p>Filing Information</p> <ul style="list-style-type: none"> • Filing Requirement • Estimated Tax <p>Dependents/Exemptions/Filing Status</p> <ul style="list-style-type: none"> • Dependents • Personal Exemptions • Filing Status - Head of Household • Filing Status - Other <p>Individual Income</p> <ul style="list-style-type: none"> • Wages, Alimony, & Unemployment Compensation • Interest & Dividend Income, Sch. B • Taxable Refunds & Other Income • Non-Taxable Income <p>Capital Gains & Losses</p> <ul style="list-style-type: none"> • Schedule D • Sale/Exchange of Residence • Other Gains/Losses 	<p>Pensions/Deferred Compensation</p> <ul style="list-style-type: none"> • Pension & Annuity Income • All IRA Inquiries • Other Retirement Plans • Taxation of Social Security Benefits • Lump Sum Distribution <p>Adjustments/Deductions</p> <ul style="list-style-type: none"> • Employee Business Expense • Other Adjustments to Income • Medical & Dental Deductions • Tax Deductions • Interest Deductions • Miscellaneous Deductions • Gifts to Charity <p>Tax Computation/Credits/Payments</p> <ul style="list-style-type: none"> • Standard Deduction • Itemized vs Standard Deduction • Child & Dependent Care Credit • Self-Employment Tax • Earned Income Credit • Other Credits/Taxes/Payments • Supplemental Medicare Premium
--	--

Source: Internal Revenue Service

We reached agreement on the 62 test questions that comprised the test and on two specific categories of correct responses: (1) correct and (2) correct and complete. A correct answer was the minimal standard IRS expected its telephone assistants to meet and we, therefore, focused our monitoring to determine whether ITCSS accurately measured assistants' responses against that standard. Answers that exceeded this standard would be classified as correct and complete, but they would also be considered as correct for monitoring and scoring purposes. It was agreed that all other answers would be scored as incorrect, meaning that the telephone assistant's answer could lead taxpayers to a wrong result on their tax return. Appendix III provides examples of selected ITCSS test questions and the responses required for both categories of correct responses.

IRS administered its test by having test callers (1) place anonymous calls to telephone assistors located at 29 telephone sites located within the continental United States and (2) score assistors' responses to the test questions. During an 11-week period beginning February 6, 1989, eight test callers completed and scored 14,876 rccss test calls. Figure 1 shows the geographic distribution of these 29 call sites. For various technical and administrative reasons, IRS did not include three telephone sites in its test—Alaska, Hawaii, and Puerto Rico.

Figure 1: Locations of Toll-Free Telephone Sites IRS Surveyed — 1989 IRS Test Call Survey



Source: Internal Revenue Service

Objectives, Scope, and Methodology

To monitor the validity of IRS' overall test and to comment on its accuracy results, we listened in on and independently scored 577 randomly selected rccss test calls during an 8-week period of the tax filing season and compared our scores for those questions and answers to IRS' scores. We based our scoring on the scoring criteria to which we had mutually agreed with IRS. Those criteria established specific acceptable combinations of required assistor probes for factual information and/or responses that needed to be present in the conversation for the response to be considered correct.

Our monitoring sample was randomly selected from IRS' test call survey plan. Overall, our sample called for us to monitor 830 test calls, covering all test callers, time periods, and test questions. We calculated that this sample size would allow us to report our accuracy results for the period at the 95-percent level of confidence with a sampling error of plus or minus 2.5 percent. However, we were unable to monitor and score 244 test calls primarily because of (1) deviations from IRS' calling schedule, (2) the inability of test callers to complete calls to the telephone sites, and (3) occasional problems with our monitoring equipment. In addition, we dropped nine calls from our sample because test callers deviated from the agreed-upon question scripts, thereby affecting the outcome of the call. Accordingly, the reduction in our sample size caused our sampling error to increase to plus or minus 4.4 percent at the 95-percent level of confidence.

We monitored how well IRS administered its test and discussed the development of and planning for the test with IRS officials in the Taxpayer Service and Statistics of Income Divisions and with the project manager of the contractor IRS selected to develop and implement ITCSS' computerized scoring response program. Our objectives, scope, and methodology are discussed in additional detail in appendix II.

ITCSS Produced a Valid Indicator of Overall Assistor Performance

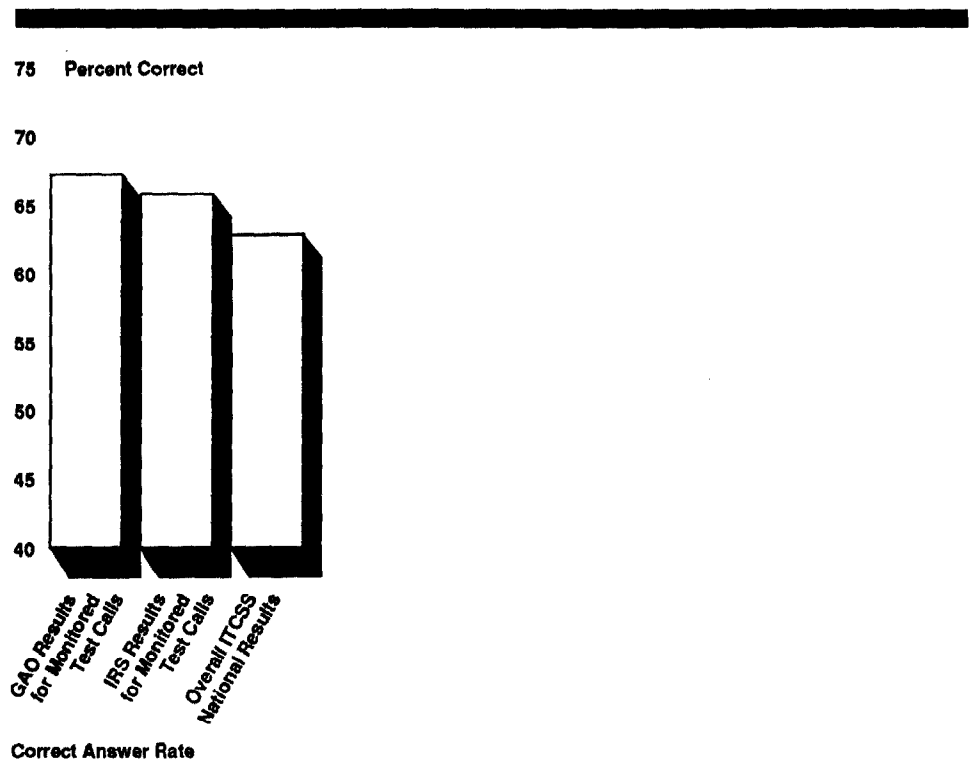
Our monitoring results showed that IRS telephone assistors correctly answered 391 of the 577 tax law test questions. For the same 577 test calls, IRS scored 377 of them as correct. Using the same method as IRS to statistically weight our scoring, our results show a 67.2-percent IRS telephone assistance accuracy rate compared to IRS' accuracy rate for the monitored calls of 65.8 percent. The difference in these rates is not statistically significant and, therefore, the overall 62.8-percent accuracy rate IRS reported for all ITCSS calls can be relied upon as a valid indicator of assistors' performance.

The variance in our and IRS' scoring of test calls was due primarily to differences in interpretation as to the adequacy of assistors' probes and responses. Probing is important because taxpayers who call with questions usually are not sufficiently familiar with the tax laws to know what information assistors need to answer their questions. Without knowing certain facts about a taxpayer's situation or status, assistors cannot be certain that the response they give would actually apply to the taxpayer. Assistors, therefore, must elicit that information from the taxpayer or provide a conditional response.

Generally, assistor probes and responses clearly met or failed to meet the agreed-upon scoring criteria for a correct response. However, instances occurred where assistors' probes and responses varied somewhat from predetermined acceptable probes and responses; therefore, judgments had to be made on whether the responses expressed were acceptable. On 60 monitored calls, or about 10 percent of our sample, we disagreed with IRS test callers as to whether a given probe or response fully met the scoring criteria. For 38 of the 60 calls, we scored the responses as correct and IRS scored them as incorrect. For the other 22 calls, we scored the assistors' responses as incorrect, but IRS scored them as correct.

Figure 2 shows IRS' overall telephone assistor accuracy rate for ITCSS during the 1989 tax filing season compared to IRS' and our results for the sample of test calls we monitored.

Figure 2: IRS' Overall Test Results Compared to GAO- and IRS-Monitored Test Call Results for the 1989 Tax Filing Season



Our scoring of assistors' responses to the 577 monitored test calls was based on scoring criteria that we and IRS mutually devised. As discussed

in our March 1989 testimony before the Subcommittee,² IRS also reported a higher accuracy rate that was based on more liberal scoring criteria than those on which we had agreed. About 2 weeks after the start of the test, IRS determined that, for certain questions, assistors were providing answers that IRS believed were “not wrong” but failed to meet our agreed-upon standards for correct answers. IRS officials said that it would be unfair to imply to 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—that failed to meet minimal standards but which it proposed to add to the “correct and complete” and “correct” categories in reporting accuracy statistics.

We did not agree to IRS’ revision of the scoring criteria. In our opinion, the responses IRS categorized as “right” were incomplete and potentially misleading and would increase the likelihood that taxpayers following such advice would make errors. For example, to defer the capital gains tax on the sale of a principal residence, a taxpayer must replace and occupy a new residence within a specified time period. IRS considered an answer as “right” if only the time period for replacement was provided. We considered the answer as incomplete and potentially misleading because of the tax consequences that would result if the taxpayer did not meet the occupancy requirement. In July 1989, the Assistant Commissioner (Taxpayer Service) said that for the 1990 filing season IRS would only report assistor accuracy rates that were based on mutually agreed-upon scoring criteria.

Conclusion

For the 1989 tax filing season, IRS fairly administered its test call survey, and we agree that its reported overall 62.8-percent assistor accuracy rate is reliable. For the sample of ITCSS test calls that we monitored, the difference in the accuracy rates for correct answers between our scoring and IRS’ scoring of those calls was not statistically significant. Thus, we believe that with periodic oversight the test call system administered by IRS can be used as the principal monitor of its assistors’ performance.

Agency Comments and Our Evaluation

In providing comments to this report, the Commissioner of the Internal Revenue Service said that he was not satisfied with the telephone assistance accuracy rate that IRS achieved in 1989 and that one of his major

²IRS’ Telephone Assistance Program (GAO/T-GGD-89-13, Mar. 15, 1989).

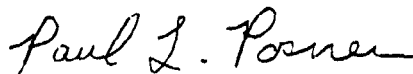
goals is to improve this accuracy rate in 1990. The Commissioner agreed with our findings but recommended the deletion of the table that presents accuracy rates for each IRS call site. He believes that because the sample size of the data pertaining to each call site is smaller than national or regional sample sizes, the confidence interval associated with any call site accuracy range is too wide to be meaningful (see app. IV).

As discussed on page 13, we agree that call site accuracy rate ranges are wider than for national or regional accuracy rate ranges. However, the potential variance in the accuracy of call site data varies from plus or minus 4.95 percent to plus or minus 6.4 percent—a range we believe useful for comparisons of call site performance. To mitigate the Commissioner's concerns and to permit reader perspective, we added to the call site data table the accuracy rate ranges for each call site.

As arranged with the Subcommittee, we are sending copies of this report to the Commissioner of Internal Revenue and other interested parties. We will make copies available to others upon request.

The major contributors to this report are listed in appendix V. Please contact me on 272-7904 if you or your staff have any questions concerning the report.

Sincerely yours,



Paul L. Posner
Associate Director, Tax Policy and
Administration Issues

Contents

Letter		1
Appendix I		12
Integrated Test Call	ITCSS Accuracy Rate Reflects Overall Quality of IRS	12
Survey System Results	Telephone Service Provided to Taxpayers	
for the 1989 Tax	Analysis of Selected Data	13
Filing Season by Tax		
Law Category, Region,		
and Call Site		
Appendix II		24
Objectives, Scope, and		
Methodology		
Appendix III		26
Selected ITCSS Test	Sample Question 1	26
Questions and	Sample Question 2	27
Required Responses		
Appendix IV		29
Comments From the		
Internal Revenue		
Service		
Appendix V		30
Major Contributors to		
This Report		
Related GAO Products		31
Tables	Table I.1: Estimated Regional Accuracy Rates and	15
	Accuracy Rate Ranges — 1989 IRS Test Call Survey	
	Results	

Contents

Table I.2: Estimated National Accuracy Rates and Accuracy Rate Ranges by Tax Law Category — 1989 IRS Test Call Survey Results	17
Table I.3: Estimated Regional Accuracy Rates and Accuracy Rate Ranges by Tax Law Category — 1989 IRS Test Call Survey Results	20
Table I.4: Accuracy Rate Ranges for IRS Toll-Free Telephone Sites — 1989 IRS Test Call Survey Results	22

Figures

Figure I.1: IRS' National Level Accuracy	12
Figure I.2: Distribution of 62 Tax Law Questions by Tax Law Category	14
Figure I.3: Estimated Regional Accuracy Rates	15
Figure I.4: Estimated National Accuracy Rates by Tax Law Category	16
Figure I.5: Estimated Regional Accuracy Rates by the Tax Law Categories of Filing Information, Exemptions, Individual Income, and Capital Gains	18
Figure I.6: Estimated Regional Accuracy Rates by the Tax Law Categories of Pensions, Adjustments to Income, and Tax Computation	19

Abbreviations

IRS	Internal Revenue Service
ITCSS	Integrated Test Call Survey System

Integrated Test Call Survey System Results for the 1989 Tax Filing Season by Tax Law Category, Region, and Call Site

ITCSS Accuracy Rate Reflects Overall Quality of IRS Telephone Service Provided to Taxpayers

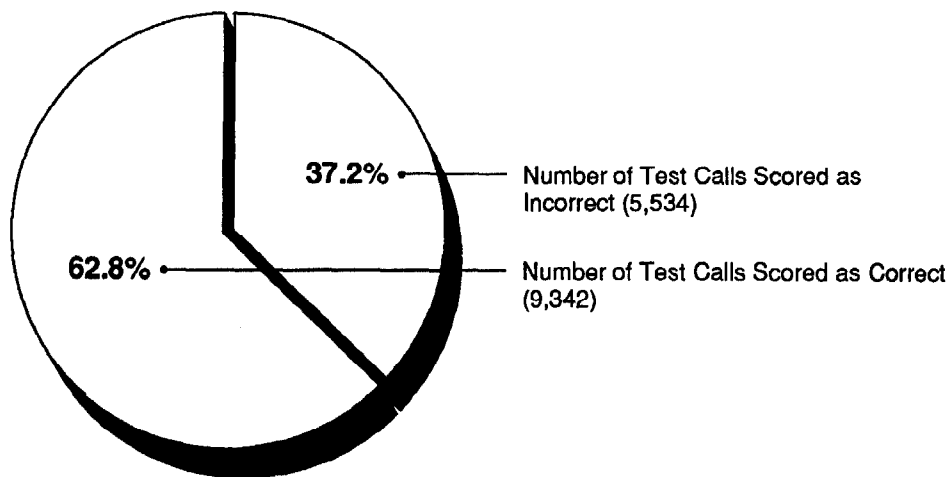
The Integrated Test Call Survey System was developed by IRS to more accurately measure the accuracy of IRS responses to taxpayer telephone inquiries. Accuracy measurement is important because IRS believes that the higher the telephone assistance accuracy rate the better the quality of service it provides to the public.

The 1989 test call survey system was designed by IRS to place 1,488 test calls per week for 11 weeks to 29 of its 32 call sites throughout the United States. Each test call came from a group of 62 questions dealing with tax law issues pertaining to individuals. All test questions were derived from tax law categories in which IRS believes most taxpayers ask questions. To be credited with a correct response, ITCSS implementation guidance directed that each IRS telephone assistant in the 29 toll-free telephone assistance call sites nationwide (1) obtain relevant facts from the taxpayer as necessary before attempting to give an answer and (2) ensure that an answer was tailored to satisfy the taxpayer's needs.

Overall National Level ITCSS Accuracy Results

From February 6, 1989, to April 21, 1989, IRS National Office test callers completed and scored 14,876 test calls. Figure I.1 shows the national level accuracy results for these test calls.

Figure I.1: IRS' National Level Accuracy



Analysis of Selected Data

The data presented in this section represent selected results obtained by IRS during its 1989 tax filing season test call survey sample. We should point out that our monitoring sample was designed to evaluate the validity of IRS' overall national accuracy rate, not the accuracy of IRS' statistical results at the tax law category, region, or call site levels. It should be expected that ITCSS results by categories, regions, and call sites have larger sampling errors than the overall ITCSS results because it is a common statistical property that a subsection of a sample has more variability than the whole sample.

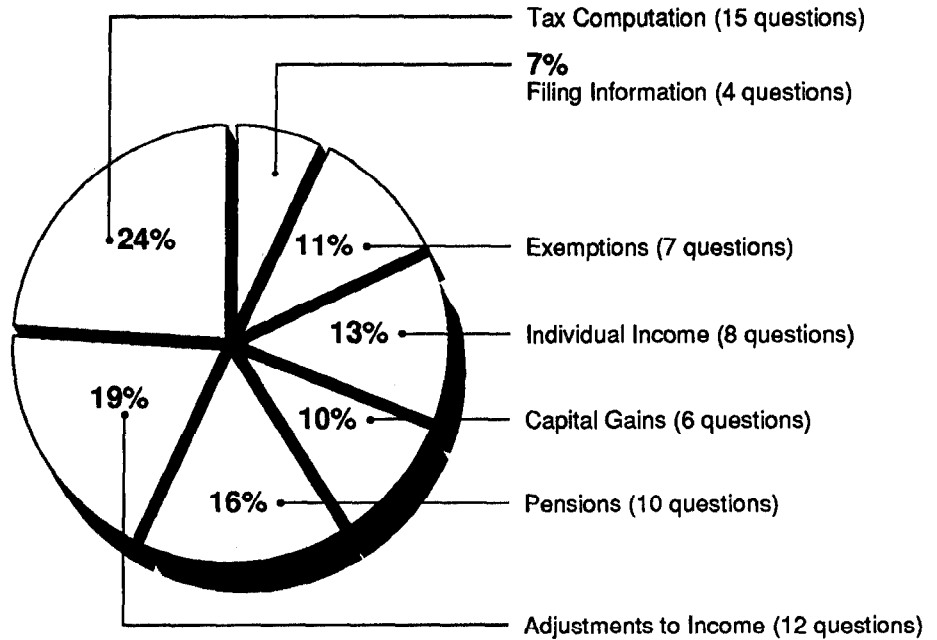
Accuracy rates are estimated because they are drawn from a statistical sample of test calls. Each estimate has a range of precision, or confidence interval, associated with it. The size of this accuracy rate range varies by the size of the test call sample used to produce the confidence interval. Therefore, the variability of accuracy rate ranges and estimated accuracy rates relating to the tax law category, regional, and call site data tables and figures that follow is the result of differing sample sizes associated with each level of data. For example, tax law category and regional data were based on larger sample sizes than the call site sample sizes and, therefore, produced narrower confidence intervals. The narrower the interval the higher the probability that the estimated accuracy rate approximates the actual accuracy rate. All data ranges shown in this section have been calculated to express the results at the 95-percent level of confidence.

Percentage of ITCSS Questions by Tax Law Category

The 62 tax law questions comprising IRS' test call survey covered seven tax law categories in which IRS determined that taxpayers commonly made telephone inquiries. Figure I.2 shows the distribution of the test call questions across the seven tax law categories.

**Appendix I
Integrated Test Call Survey System Results
for the 1989 Tax Filing Season by Tax Law
Category, Region, and Call Site**

Figure I.2: Distribution of 62 Tax Law Questions by Tax Law Category

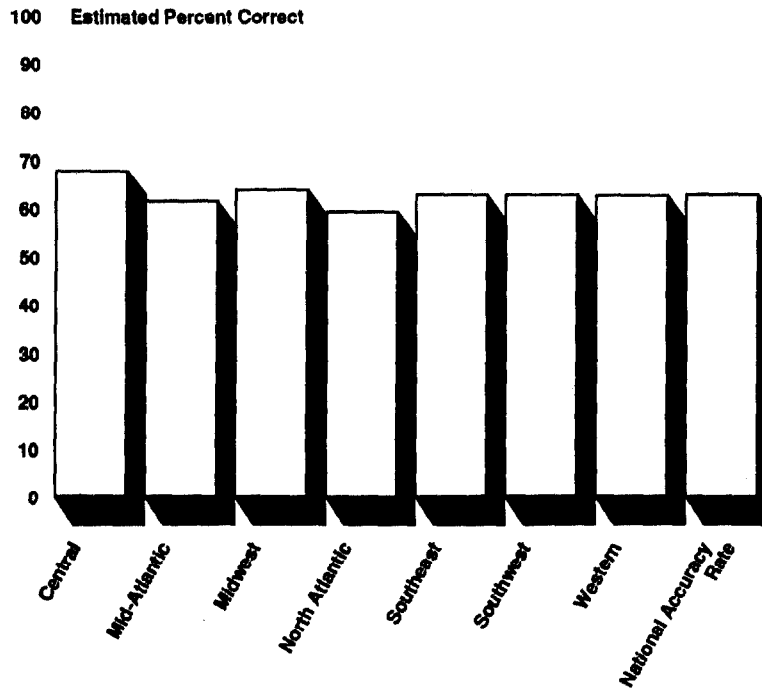


Estimated Accuracy Rates for IRS Regions

Figure I.3 shows the estimated accuracy rates achieved by each IRS region, and table I.1 shows the specific accuracy rate range associated with each region's estimate. These data indicate that the Central Region accuracy rate clearly exceeded both the North Atlantic and Mid-Atlantic Regions' accuracy rates.

**Appendix I
Integrated Test Call Survey System Results
for the 1989 Tax Filing Season by Tax Law
Category, Region, and Call Site**

Figure I.3: Estimated Regional Accuracy Rates



IRS Regions and National Accuracy Rates

Table I.1: Estimated Regional Accuracy Rates and Accuracy Rate Ranges — 1989 IRS Test Call Survey Results

Figures in percent

IRS region	Estimated accuracy rate	Accuracy rate range
Central	67.7	64.9 - 70.6
Mid-Atlantic	61.4	58.6 - 64.2
Midwest	63.7	61.3 - 66.1
North Atlantic	59.2	55.7 - 62.6
Southeast	62.7	59.7 - 65.8
Southwest	62.7	59.8 - 65.7
Western	62.6	59.8 - 65.5

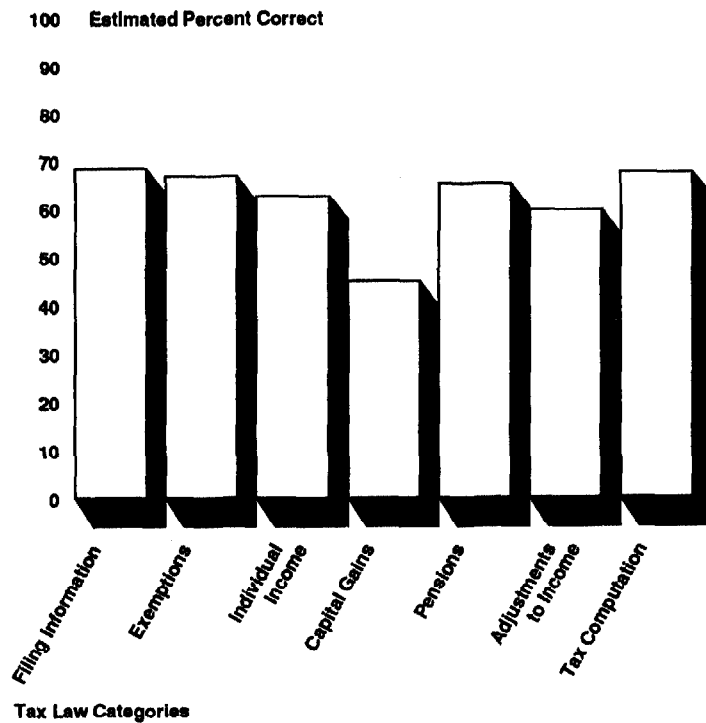
Source: Internal Revenue Service

**Appendix I
Integrated Test Call Survey System Results
for the 1989 Tax Filing Season by Tax Law
Category, Region, and Call Site**

**Estimated National and
Regional Accuracy Rates
by Tax Law Category**

Figure I.4 shows the estimated accuracy rates achieved by IRS telephone assistants within each tax law category, and table I.2 shows the accuracy rate range data associated with these estimates. These data illustrate that telephone assistants clearly had the most difficulty providing correct responses to questions dealing with capital gains.

Figure I.4: Estimated National Accuracy Rates by Tax Law Category



**Appendix I
Integrated Test Call Survey System Results
for the 1989 Tax Filing Season by Tax Law
Category, Region, and Call Site**

Table I.2: Estimated National Accuracy Rates and Accuracy Rate Ranges by Tax Law Category — 1989 IRS Test Call Survey Results

Figures in percent

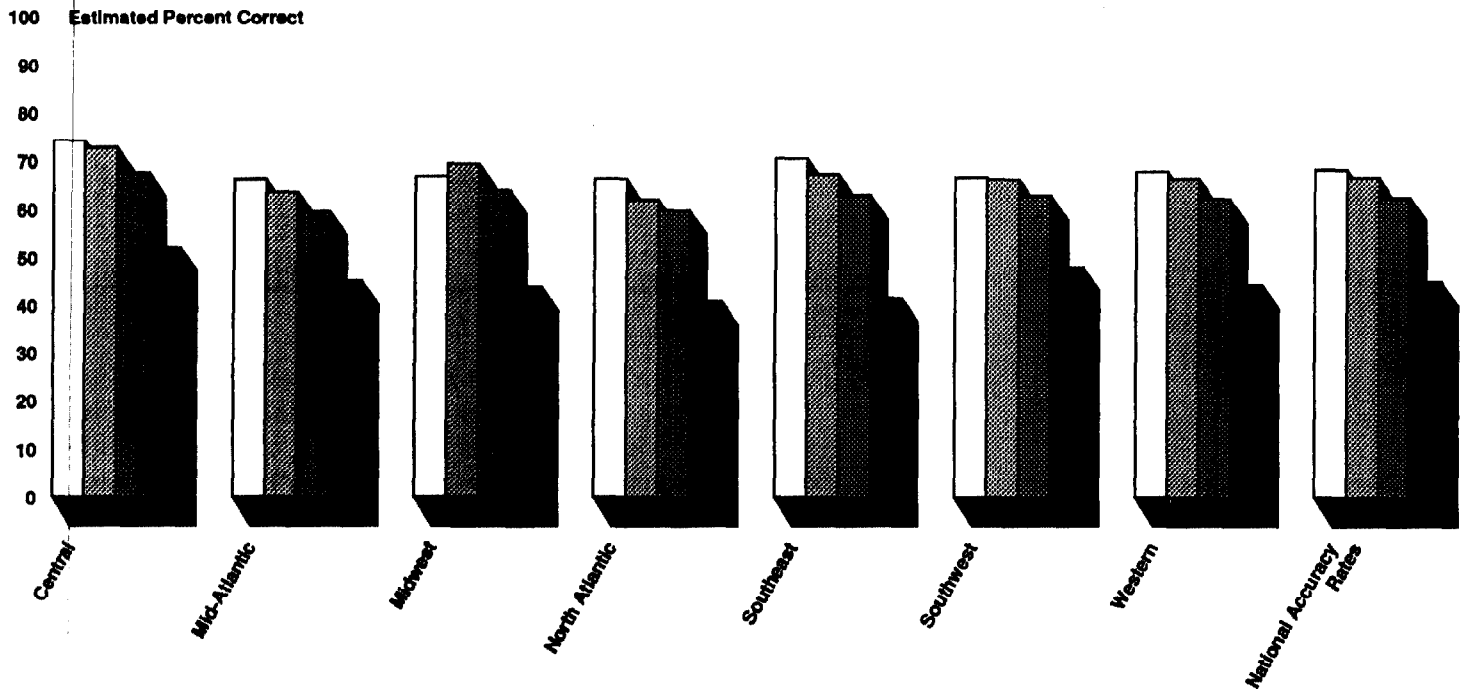
Tax law category	Estimated accuracy rate	Accuracy rate range
Filing Information	68.3	65.3 - 71.3
Exemptions	66.7	63.6 - 69.8
Individual Income	62.7	59.6 - 65.8
Capital Gains	44.9	41.3 - 48.5
Pensions	65.3	62.4 - 68.2
Adjustments to Income	59.8	56.9 - 62.7
Tax Computation	67.7	65.3 - 70.1

Source: Internal Revenue Service

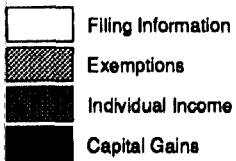
Figures I.5 and I.6 show the estimated accuracy rates, and table I.3 shows the corresponding accuracy rate ranges for each IRS region in each tax law category. For purposes of comparison, we have included national accuracy rates and ranges for the same tax law categories.

Appendix I
 Integrated Test Call Survey System Results
 for the 1989 Tax Filing Season by Tax Law
 Category, Region, and Call Site

Figure I.5: Estimated Regional Accuracy Rates by the Tax Law Categories of Filing Information, Exemptions, Individual Income, and Capital Gains

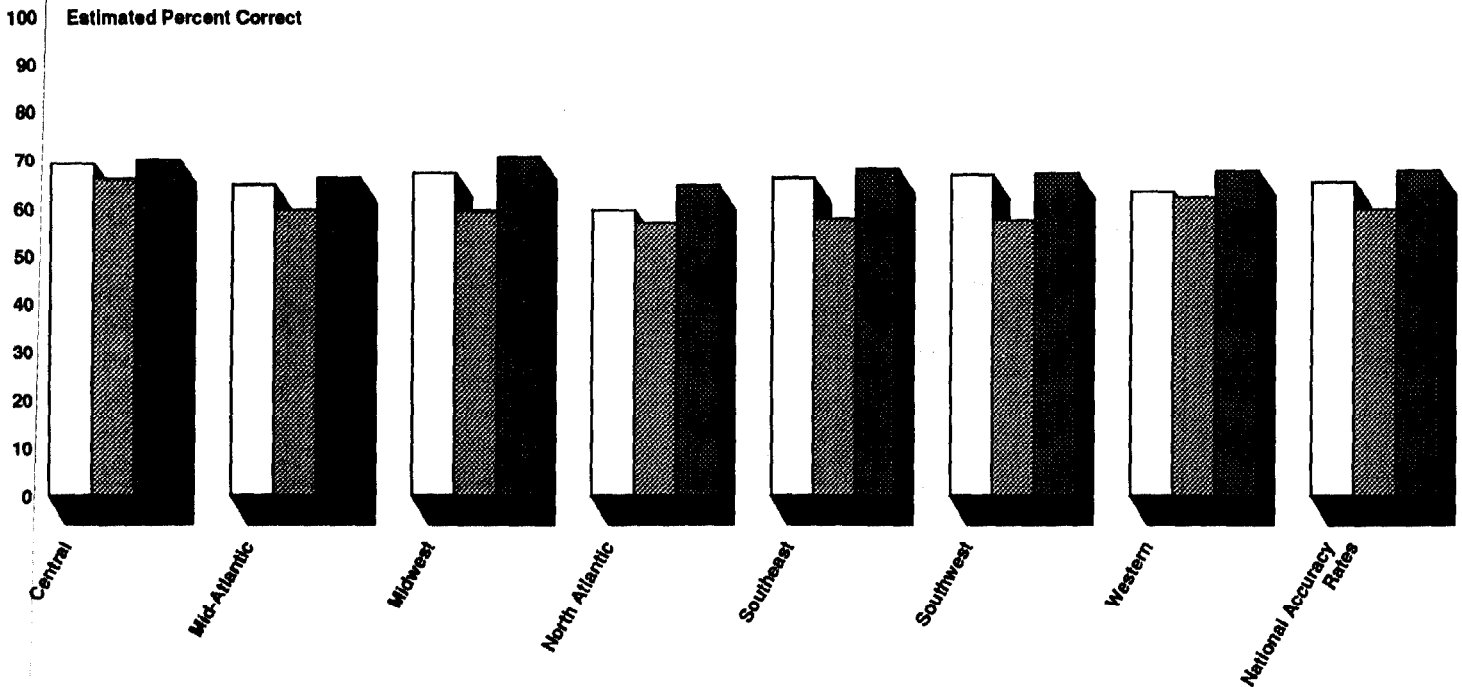


IRS Regions and National Accuracy Rates

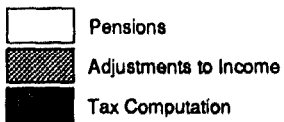


Appendix I
 Integrated Test Call Survey System Results
 for the 1989 Tax Filing Season by Tax Law
 Category, Region, and Call Site

Figure I.6: Estimated Regional Accuracy Rates by the Tax Law Categories of Pensions, Adjustments to Income, and Tax Computation



IRS Regions and National Accuracy Rates



**Appendix I
Integrated Test Call Survey System Results
for the 1989 Tax Filing Season by Tax Law
Category, Region, and Call Site**

Table I.3: Estimated Regional Accuracy Rates and Accuracy Rate Ranges by Tax Law Category — 1989 IRS Test Call Survey Results

Figures in percent

Tax law category	IRS region	Estimated accuracy rate	Accuracy rate range
Filing information	Central	74.1	66.7 - 81.5
	Mid-Atlantic	66.1	58.6 - 73.6
	Midwest	66.6	60.0 - 73.2
	North Atlantic	66.2	57.1 - 75.3
	Southeast	70.5	62.6 - 78.4
	Southwest	66.7	58.7 - 74.7
	Western	67.9	60.3 - 75.5
	(National average)	(68.3)	(65.3) - (71.3)
Exemptions	Central	72.7	65.0 - 80.4
	Mid-Atlantic	63.6	55.8 - 71.4
	Midwest	69.3	62.7 - 75.9
	North Atlantic	62.0	52.4 - 71.6
	Southeast	67.2	58.9 - 75.5
	Southwest	66.2	58.0 - 74.4
	Western	66.4	58.6 - 74.2
	(National average)	(66.7)	(63.6) - (69.8)
Individual income	Central	67.5	59.6 - 75.4
	Mid-Atlantic	59.5	51.7 - 67.3
	Midwest	63.9	57.2 - 70.6
	North Atlantic	59.7	50.3 - 69.1
	Southeast	63.1	54.8 - 71.4
	Southwest	62.9	54.7 - 71.1
	Western	62.2	54.4 - 70.0
	(National average)	(62.7)	(59.6) - (65.8)
Capital gains	Central	51.9	42.5 - 61.3
	Mid-Atlantic	44.9	36.1 - 53.7
	Midwest	43.7	35.9 - 51.5
	North Atlantic	40.8	30.2 - 51.4
	Southeast	41.5	32.0 - 51.0
	Southwest	48.0	38.6 - 57.4
	Western	44.3	35.3 - 53.3
	(National average)	(44.9)	(41.3) - (48.5)

(continued)

**Appendix I
Integrated Test Call Survey System Results
for the 1989 Tax Filing Season by Tax Law
Category, Region, and Call Site**

Tax law category	IRS region	Estimated accuracy rate	Accuracy rate range
Pensions	Central	69.1	61.7 - 76.5
	Mid-Atlantic	64.6	57.4 - 71.8
	Midwest	67.2	60.9 - 73.5
	North Atlantic	59.6	50.6 - 68.6
	Southeast	66.1	58.3 - 73.9
	Southwest	66.8	59.2 - 74.4
	Western	63.4	56.0 - 70.8
	(National average)	(65.3)	(62.4) - (68.2)
Adjustments to income	Central	65.8	58.5 - 73.1
	Mid-Atlantic	59.6	52.5 - 66.7
	Midwest	59.5	53.2 - 65.8
	North Atlantic	57.1	48.4 - 65.8
	Southeast	57.9	50.1 - 65.7
	Southwest	57.6	50.0 - 65.2
	Western	62.2	55.0 - 69.4
	(National average)	(59.8)	(56.9) - (62.7)
Tax computation	Central	69.9	63.7 - 76.1
	Mid-Atlantic	66.1	60.1 - 72.1
	Midwest	70.6	65.5 - 75.7
	North Atlantic	64.6	57.3 - 71.9
	Southeast	68.0	61.6 - 74.4
	Southwest	67.0	60.7 - 73.3
	Western	67.7	61.7 - 73.7
	(National average)	(67.7)	(65.3) - (70.1)

Source: Internal Revenue Service

**Appendix I
Integrated Test Call Survey System Results
for the 1989 Tax Filing Season by Tax Law
Category, Region, and Call Site**

**Estimated Accuracy Rates
for IRS Toll-Free
Telephone Assistance Call
Sites**

Table I.4 below shows for the 1989 tax filing season the variations in accuracy rate ranges for the 29 telephone sites tested by IRS.

**Table I.4: Accuracy Rate Ranges for IRS
Toll-Free Telephone Sites — 1989 IRS
Test Call Survey Results**

Figures in percent		
IRS telephone sites by region	Estimated accuracy rate	Accuracy rate range
Central Region		
Cincinnati	65.6	59.5 - 71.7
Cleveland	69.8	63.8 - 75.7
Detroit	65.2	59.1 - 71.3
Indianapolis	70.6	65.7 - 75.6
Mid-Atlantic Region		
Baltimore	64.1	57.9 - 70.2
Newark	52.0	45.6 - 58.4
Philadelphia	54.9	48.5 - 61.3
Pittsburgh	72.6	66.8 - 78.3
Richmond	61.1	54.9 - 67.4
Midwest Region		
Chicago	58.2	52.8 - 63.5
Des Moines	69.7	63.8 - 75.7
Milwaukee	65.0	58.8 - 71.1
Omaha	73.0	67.3 - 78.7
St. Louis	62.5	56.2 - 68.7
St. Paul	69.1	63.2 - 75.1
North Atlantic Region		
Boston	67.7	62.7 - 72.8
Brooklyn	52.1	45.7 - 58.5
Buffalo	67.7	61.7 - 73.7
Southeast Region		
Atlanta	57.2	51.9 - 62.6
Jacksonville	66.0	60.9 - 71.2
Nashville	67.4	62.3 - 72.5

(continued)

**Appendix I
 Integrated Test Call Survey System Results
 for the 1989 Tax Filing Season by Tax Law
 Category, Region, and Call Site**

IRS telephone sites by region	Estimated accuracy rate	Accuracy rate range
Southwest Region		
Dallas	59.6	54.3 - 64.9
Denver	64.6	58.4 - 70.7
Houston	65.5	59.4 - 71.6
Phoenix	62.0	55.8 - 68.2
Western Region		
El Monte	56.0	49.6 - 62.4
Oakland	65.1	59.9 - 70.3
Portland	71.3	65.5 - 77.1
Seattle	67.5	62.5 - 72.6

Source: Internal Revenue Service

Objectives, Scope, and Methodology

Our objectives were to report on IRS' administration of its 1989 test call survey and on the validity of the statistical estimates produced during this test. To evaluate how well IRS administered ITCSS, we interviewed Taxpayer Service Division officials, reviewed IRS planning documents and managerial records, and monitored a randomly selected sample of test calls. We also interviewed officials and reviewed documents from IRS' Statistics of Income Division and Mathematica Policy Research, Inc., (the contractor IRS selected to develop and implement the computerized response scoring program) for information pertaining to ITCSS' design and implementation. Finally, IRS' internal audit reviewed ITCSS procedures and results, and we interviewed the IRS auditors and reviewed their evaluation documentation. We did our work at the IRS National Office in Washington, D.C., from January 1989 to August 1989.

To evaluate the validity of the statistical estimates produced by ITCSS, we monitored and scored a statistically valid random sample of survey test calls and compared our scoring of those calls with documentation showing how IRS scored the same calls. IRS devised its test call survey sample plan to produce statistical estimates of the accuracy of its telephone assistants in answering scripted test questions involving tax law for individuals. ITCSS design methodology called for eight IRS test callers at the National Office to place a total of 16,368 randomly selected test calls over an 11-week period (Feb. 6, 1989, through Apr. 21, 1989) to 29 toll-free telephone assistance sites throughout the United States (see fig. 1). During the test period, IRS test callers actually completed and scored 14,876 test calls.

Each test caller was scheduled to make 186 test calls per week to various call sites and at various times specified in the test call sample. The test call sample assigned to each test caller was randomly selected from a pool of 62 tax law questions representing the seven major tax law categories in which IRS determined that individual taxpayers commonly ask questions. These seven tax law categories contained 32 subcategories of tax law, as shown in table 1. The ITCSS design methodology was developed to produce results that would have a sampling error of plus or minus 2 percent at the 95-percent level of confidence.

We began monitoring IRS' test on Wednesday, February 22, 1989, about 2 weeks after IRS started its test call sample, and the first workday that telephone monitoring equipment supplied by IRS was operable. We continued our monitoring until Friday, April 14, 1989, a total of 38 test days. In order to comment on IRS' accuracy results, we developed a sampling plan that called for us to listen to and score a randomly selected

sample of 830 scheduled test calls that covered all test callers, daily time periods, and test questions. We calculated that this sample size would allow us to report our accuracy results for the period at the 95-percent level of confidence with a sampling error of plus or minus 2.5 percent.

To accomplish our monitoring, we developed monitoring records that incorporated the scripted test questions, probes, and responses used by IRS' test callers. We used an individual monitoring record to document the scoring of each test call and to note any test caller deviations from the scripted test calls or assistor variations from acceptable probes. At the end of each day, we provided IRS with a listing of the calls we monitored, and IRS later provided us with documentation showing the RTCSS test callers' scoring of the same test calls. We compared our scoring with IRS' scoring for each monitored test call and documented the results.

We evaluated test caller deviations from the scripts to determine whether they could have had a material effect on the assistants' responses. We determined that nine deviations were material (e.g., inappropriate information provided by the test caller either led to or preempted an assistor's response), and we deleted those calls from our monitoring sample.

In addition to the nine calls deleted because of test caller script deviations, we were unable to monitor and score 244 test calls primarily due to (1) deviations from IRS' calling schedule because of test caller absences and IRS staff meetings; (2) IRS' inability to complete test calls as scheduled due to heavy call volumes at the sites called; and (3) occasional monitoring equipment problems, which impaired our ability to clearly hear the assistants' responses. However, anticipating such problems, we purposely oversampled to accommodate lost calls. Although we oversampled, the number of lost calls exceeded our estimates and caused our sampling error to increase. Accordingly, the 577 test calls we monitored and scored are a statistically valid sample size that allows us to report our results with 95-percent confidence that our sampling error is no greater than plus or minus 4.4 percent.

Selected ITCSS Test Questions and Required Responses

This appendix presents two ITCSS test questions that were used in the 1989 test call survey. IRS and we agreed that these questions would not be used in the 1990 survey and, thus, we believe that they will provide readers of this report with concrete examples of the types of questions that comprise the test call survey.

To score ITCSS test questions, IRS and we agreed on the specific responses that would be categorized as (1) correct or (2) correct and complete. A correct answer was the minimal standard IRS expected its telephone assistors to meet. Answers that exceeded this standard would be classified as correct and complete. It was further agreed that all other answers would be scored as incorrect—meaning that the telephone assistor's answer could lead taxpayers to a wrong result on their tax return. IRS' reported 62.8-percent national accuracy rate for the 1989 tax filing season and our monitoring of how well IRS administered its test call survey were based on the agreed-upon scoring criteria for "correct" responses. Answers that met the correct and complete standard were also considered as correct for monitoring and scoring purposes.

For 48 of the 62 ITCSS test questions, scoring criteria required that assistors probe callers to obtain information that would be needed to answer their questions with a correct response. Of the 48 questions that required assistors to probe, 29 questions required 1 probe, 16 questions required 2 probes, and 3 questions required 3 probes. Probing is important because taxpayers who call with questions usually are not familiar with the tax laws and frequently do not know what information assistors need to answer their questions correctly. Without knowing certain facts about a taxpayer's situation or status, assistors cannot be certain that the response they give would actually apply to the taxpayer. Assistors, therefore, must elicit that information from the taxpayer or provide a conditional response.

The two sample questions that follow illustrate test questions that require no probing and questions that require multiple probing. To assist the caller in judging whether assistors covered the required probes and gave the correct responses, the required probing and response points were enumerated individually.

Sample Question 1

Tax law category: Capital gains and losses.

Subcategory: Sale/Exchange of residence.

Question: My husband (wife) and I have been working for a major corporation in Germany and have decided to sell our home in the United States. We were told that we only have 2 years in which to replace the property. Doing that will be a real burden on us since we'll still be out of the United States. Is there a way around that 2-year requirement for replacement? We are not eligible for the one-time exclusion for people 55 or older.

Background:

- Caller and spouse have been overseas for 6 months.
- Caller and spouse have not rented their U.S. home.
- Caller and spouse will be abroad about 3 years.
- Caller's tax home is outside of United States.

Probing points: None.

Response points:

R1: The replacement for your main home is extended to 4 years from the date of sale of your old home.

R2: You must occupy the new home within the 4-year period.

R3: Refer to Publication 523, Tax Information on Selling Your Home.

Scoring:

Correct: R1 and R2.

Correct and complete: R1 and R2 and R3.

Sample Question 2

Tax law category: Individual income.

Subcategory: Wages, alimony, and unemployment compensation.

Question: My father was unemployed part of last year. He only made \$3,500 before he went on unemployment. Does he have to file a return?

Background:

**Appendix III
Selected ITCSS Test Questions and
Required Responses**

- Father received \$1,600 in unemployment compensation from the state and he made no contributions to the plan.
- Father received no interest or other income.
- Father is 61 and not blind.
- Father is single with no dependents.
- Father cannot be claimed as a dependent on caller's (or anyone else's) return.

Probing points:

P1: How much unemployment compensation did your father receive?

P2: How old is your father? Or is your father 65 or older?

P3: What is your father's filing status? Or is your father married?

Response Points:

R1: Yes, he must file a return.

R2: His unemployment benefits are taxable.

R3: His total income exceeds the threshold for filing; Or his total income exceeds \$4,950; Or his total income exceeds his standard deduction and personal exemption.

Scoring:

Correct: P1 and P2 and P3 and R1 and (R2 or R3).

Correct and complete: P1 and P2 and P3 and R1 and R2 and R3.

Comments From the Internal Revenue Service



COMMISSIONER

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

NOV 22 1989

Mr. Richard L. Fogel
Assistant Comptroller General
United States General Accounting Office
Washington, DC 20548

Dear Mr. Fogel:

We have reviewed your recent draft report entitled, "Tax Administration: Monitoring the Accuracy and Administration of IRS' 1989 Test Call Survey," which was produced at the request of the Chairman, Subcommittee on Commerce, Consumer, and Monetary Affairs, House Committee on Government Operations.

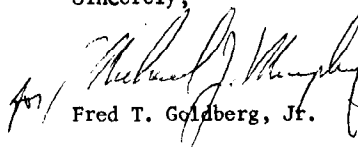
We generally agree with the report's findings which validate the design and our administration of the Integrated Test Call Survey System (ITCSS). However, we recommend deletion of the tables in Appendix I that present accuracy rates for each telephone site. Because of the smaller sample pertaining to each site, it is not possible to achieve statistical validity without presenting data in ranges that are too wide to be meaningful. For this reason, it has been IRS policy not to release individual call site data. We have no objection to publishing call site data at the end of the next filing season if we can work with GAO to assure statistically valid data.

We would also like to work with GAO to provide any data that is necessary to release the report for the 1990 filing season as soon as possible after the end of the filing season. We believe that earlier release of the report would avoid public confusion and the consequent increase in the volume of calls when the report is released at the beginning of the next filing season. It would also help us focus on remedial actions in planning for the next filing season.

The IRS is not satisfied with the accuracy rate that we achieved last year. One of my major goals is to improve the taxpayer service accuracy rate for 1990 and we are taking steps to achieve this improvement. For example, a test site in Boston provides IRS telephone assistors with a computerized data system designed to ensure that taxpayers are asked all necessary questions and correct answers are provided by telephone assistors. Taxpayer Service staff throughout the country have been provided with written desk guides that use these techniques to teach assistors to fully and accurately respond to taxpayer inquiries. We have also used the test call data from the 1989 filing season to modify our training of telephone assistors to improve weak areas. These and other actions lead us to believe that the 1990 filing season will see substantial improvements in our telephone tax assistance.

Best regards.

Sincerely,


Fred T. Goldberg, Jr.

Major Contributors to This Report

**General Government
Division, Washington,
D.C.**

Larry H. Endy, Assistant Director, Tax Policy and Administration Issues
Robert P. Glick, Assignment Manager
Martin S. Morris, Tax Attorney
William F. Bley, Evaluator-in-Charge
Susan Ragland, Evaluator
Maria Z. Oliver, Evaluator

**Program Evaluation
and Methodology
Division, Washington,
D.C.**

Harry M. Conley III, Statistician

Related GAO Products

Accessibility, Timeliness, and Accuracy of IRS' Telephone Assistance Program (GAO/GGD-89-30, Feb. 2, 1989).

Requests for copies of GAO reports should be sent to:

**U.S. General Accounting Office
Post Office Box 6015
Gaithersburg, Maryland 20877**

Telephone 202-275-6241

The first five copies of each report are free. Additional copies are \$2.00 each.

There is a 25% discount on orders for 100 or more copies mailed to a single address.

Orders must be prepaid by cash or by check or money order made out to the Superintendent of Documents.

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

Official Business
Penalty for Private Use \$300

First-Class Mail
Postage & Fees Paid
GAO
Permit No. G100
