2007 The IRS Research Bulletin

Publication 1500

The IRS Research Bulletin

Proceedings of the 2007 IRS Research Conference

Research
Analysis
Statistics

Lab Research on Customer Preferences and the Relationship Between Service and Compliance

Kathleen Holland and Howard Rasey, Internal Revenue Service

The relationship between the services IRS provides and compliance, although intuitive, is difficult to quantify. In 1989, Price Waterhouse conducted a study designed to measure this relationship by providing groups of participants with different types of service while they completed hypothetical tax situations.[1] The groups either had access to no extra service, telephone service, face-to-face service, or both telephone and face-to-face service. Their results indicate that those with assistance available had lower absolute error than those without assistance. This difference in absolute error existed regardless of the type of available service.

The current study was designed to partially replicate and expand on the Price Waterhouse 1989 study. In addition to testing telephone and walk-in assistance, IRS.gov Web assistance was also tested. This addition resulted in a five-condition experimental design: 1) forms and publications only; 2) forms, publications, and IRS.gov; 3) forms, publications, and face-to-face assistance; 4) forms, publications, and telephone assistance; and 5) forms, publications, IRS.gov, face-to-face, and telephone assistance. Another modification to the Price Waterhouse study was the decision to analyze data based on actual service use as well as analyze the data based on service availability. This second analysis was designed to examine if service availability alone impacts compliance or if the actual provision of service is necessary to have an effect.

In addition to providing further information on the relationship between service and compliance, this study was also designed to provide information about taxpayer use of different IRS channels. One group was given equal access to all three major types of IRS service. The behaviors of this group were expected to provide information on unqualified use of services by taxpayers. Because participants were able to use service more than once, this group could also provide information about patterns of service usage.

[1] Price Waterhouse, *Study of the Effect of Taxpayer Assistance on Voluntary Compliance, IRS Management Briefing and Executive Summary*, July 1989.

Research Methods

Participants

A contractor recruited 195 people to participate in the study. Of the 195 people scheduled to participate, 176 actually participated in the study. In order to be eligible, participants were required to be at least 18 years of age, have a 10th grade education or above, speak English as their primary language, and have completed their own tax returns without the use of tax preparation software at least once in the last 5 years. Screening also required that approximately half of the participants not have Internet access at home. Of the 176 participants, 52 percent did not have Internet access at home. Unfortunately, none of the participants was over the age of 65; therefore, our sample does not represent those over the age of 65.

Materials

For the current study, it was determined that certain tax situations would be included in the tax scenarios. These included itemizing deductions (completing a Schedule A), claiming Earned Income Tax Credit (EITC), and reporting taxable Social Security income. These tax topics or situations were chosen so that further analysis could be conducted on these topics. Schedule A was included because Price Waterhouse found significant errors with itemized deductions. Taxable Social Security was included because it is one of the top ten math error codes every year. Further, the IRS has found that this error is commonly made by both taxpayers and paid preparers. EITC was chosen in an attempt to further understand this special population.

Scenarios were matched to the participant's own tax situation so that no participant encountered anything completely new during the study. In total, 16 scenarios were created for the study. In order to ensure the completeness of the scenarios and to ensure accuracy for scoring the return, the scenarios were either adapted from Volunteer Income Tax Assistance (VITA) training materials or from scenarios utilized in a Government Accountability Office (GAO) study. The scenarios were also independently completed by Wage and Investment research analysts.

In addition to the tax scenarios, a debrief survey was also designed and given to the participants. The debrief survey asked about the participant's experience completing the tax scenario and also about any service he or she used while completing the tax scenario. The debrief also asked questions about completing their most recent personal tax returns and their service usage when preparing their most recent personal tax returns.

Procedures

A contractor, Development Associates, was hired to recruit participants for the current study. The contractor recruited through advertising on Craigslist.com and in small market local print media. The contractor screened interested individuals for eligibility and asked about their personal tax situations to assign them to the proper tax scenario. Those who had completed a Form1040EZ the previous tax year were asked if they would be comfortable completing a Form1040A. If they said no, they were excluded from the study. Interested, eligible, available participants were then scheduled to attend one of the test sessions at an Atlanta, Georgia Taxpayer Assistance Center (TAC). At the time of recruitment, participants were told participation would take 2 hours. No participant took the full 2 hours, and most were done in less than 1 and a half hours.

Scheduled participants were randomly assigned to one of five service groups with different options for service: 1) forms and publications only; 2) forms, publications, and IRS.gov; 3) forms, publications, and face-to-face assistance; 4) forms, publications, and telephone assistance; and 5) forms, publications, IRS.gov, face-to-face, and telephone assistance.

The study was primarily conducted in a large conference room at an Atlanta TAC. On arrival, participants were given an orientation by the contractor. During orientation, participants were given an ID badge (with participant number and assigned service group) and a packet with instructions and mock tax scenarios. Participants were also read the instructions during the orientation. Orientation took place in a small room outside the TAC. After orientation, participants were taken to the large conference room. All participants were provided with a calculator and pencils. Eligible participants who elected to use service signaled an usher who then took them to the service location. Service was provided at the TAC for face-to-face service, in a small room with a phone for IRS toll-free telephone assistance, and at a bank of three computers in the large conference room for IRS.gov assistance. After completing mock tax scenarios, participants were given a debrief questionnaire. Before leaving the facility, participants were paid an honorarium of $75 for their time.

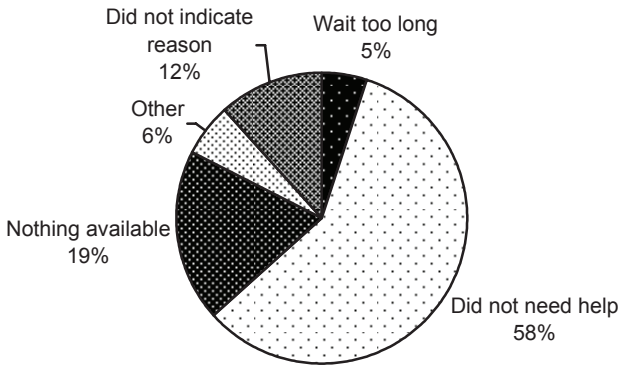
Research Findings

Of the 176 participants, 7 indicated that they were not able to complete the tax scenarios and were dropped from further analysis. Only 24 percent of the 136 with access to service other than forms and publications actually used service. Of the 32 with access to IRS.gov, 13 percent used it. Of the

30 with access to face-to-face assistance, 33 percent used it. Of the 35 with access to the telephone, 14 percent used it. Of the 38 who had access to all three forms of service, 13 percent used IRS.gov, 16 percent used face-to-face assistance, and 5 percent used telephone assistance.

On completing the mock tax scenario, participants completed a short questionnaire regarding their experiences with the scenario. One of the questions asked if participants used service and, if not, why not. As shown in Figure 1, of those who indicated they did not use service, 58 percent indicated it was because they did not need it.

Figure 1. Reasons Indicated for Not Using Service



Analysis of accuracy across the different service channels is not statistically valid due to the small number of people who used service. Therefore, service usage was collapsed across groups. Accuracy was measured as the absolute value of the error for four lines of the 2005 Form 1040: adjusted gross income (line 37), taxable income (line 43), tax (line 44) and total tax (line 63). The accuracy of participants who received service was then compared with those who did not receive service. Results are presented in Figure 2.

No significant differences at the p=.05 level were found for any of the four line item accuracy measures. The lack of significant differences is consistent with the large variability in errors as evidenced by the large standard deviations (standard deviations are presented in parentheses in Figure 2). Although the difference was not statistically significant, those who received service did have a lower average absolute error or higher accuracy on all four line items than those who did not receive service.

Figure 2. Comparison of Accuracy as Measured by the Absolute Value of the Error

	Did not use service		Used service	
	Average Absolute Error (Standard Deviation)	n	Average Absolute Error (Standard Deviation)	n
Adjusted Gross Income	3,917.78 (11,707.53)	132	3,388.06 (9,355.70)	31
Taxable Income	7,723.43 (10,246.41)	124	7,452.84 (11,763.28)	31
Tax	1,579.24 (2,208.61)	104	1,342.23 (1,806.89)	21
Total Tax	4,766.77 (11,683.25)	118	2,036.19 (2,518.11)	27

In addition to the average absolute error, error rates were also calculated for the four line items. Error rates are the percentages of participants who incorrectly completed that line. Failure to enter anything on a specific line was coded as an error. No corrections were made to the tax forms; therefore, an error on an earlier line should cause an error on subsequent lines, and the error rate should increase over the Form 1040. As shown in Figure 3, the error rate of all participants for adjustable gross income is 51 percent and increases to 82 percent for taxable income.

Figure 3. Percentage and Confidence Intervals of Errors on Select Lines

	Overall (n=169)		Did not use service (n=137)		Used service (n=32)	
	Point Estimate	Confidence Interval	Point Estimate	Confidence Interval	Point Estimate	Confidence Interval
Adjusted Gross Income	51%	44%-59%	49%	41%-57%	63%	46%-80%
Taxable Income	82%	76%-88%	83%	77%-89%	78%	64%-92%
Tax	81%	75%-87%	83%	77%-89%	75%	60%-90%
Total Tax	80%	74%-86%	80%	73%-87%	78%	64%-92%

Error rates and confidence intervals were computed for all participants, as well as for those who used service and those who did not. The results are presented in Figure 3. Due to the low number of people who used service and the resulting wide confidence intervals, lack of significant differences

cannot be meaningfully interpreted. Also due to the small sample size, these numbers cannot be inferred to the population.

After completing the tax scenarios, participants completed a debrief questionnaire. The debrief questionnaire asked participants to rate their confidence in the accuracy of the tax return they had just completed using an eight-point scale. Of the 137 who gave a confidence rating for their accuracy, 64 percent rated their accuracy at six or above. The relationship between self-reported confidence in accuracy and the four line item measures of accuracy was analyzed. Three of the line item accuracy measures indicate a significant negative correlation between self-reported accuracy and the average absolute error for adjusted gross income ($r = -.22$, $p < .01$), tax ($r = -.27$, $p < .01$), and total tax ($r = -.26$, $p.<01$). This negative correlation indicates that, as self-reported confidence increases, the average absolute error decreases. The correlation was not significant between self-reported confidence and the absolute value of the error for taxable income ($r = -.13$, $p. = .10$).

The accuracy of the telephone service and face-to-face service was defined as follows. It was considered accurate if the assistor completely answered the participant's question, whether proper procedures were followed or not. For the telephone, it was expected that, with the time information, the name of the assistor, and the assistor's badge number, W&I Research would be able to identify and score the telephone assistance. Unfortunately, there was some technical difficulty in identifying the correct telephone calls, mostly due to transfers, and assessment of the telephone assistance was not possible.

For face-to-face service, recordings were taken on hand-held digital recorders and scored for accuracy by a member of W&I Research. Scores were then reviewed by a second member of W&I Research. It was determined that 100 percent of participants who received face-to-face service had their tax questions answered accurately and completely. As with telephone service, the accuracy scores do not indicate if proper procedures were followed or if the service provided was in scope. Accuracy for this study indicates that tax law questions were answered accurately and completely.

Conclusions

This study resulted in several unexpected findings. These include the low incidence of service usage, 24 percent, and the low accuracy rate, about 20 percent. The results of the debrief questionnaire indicate that the low service usage rate was mostly the result of people thinking they did not need assistance, and the low service usage rate did not allow for accuracy comparisons between service channels. The accuracy rate is better, but not significantly,

for those who used service than for those who did not. The lack of significant differences in accuracy and the low service usage rate together prevent conclusions about the relationship between service and compliance.

The service usage of the different channels by those who had equal access to all three channels indicates the possibility of an interesting pattern, relatively equal opting for IRS.gov and face-to-face assistance and less for telephone assistance. If this pattern were to hold up in a larger sample with a higher service usage rate, it would provide useful operational information. Currently, we do not know and cannot tell if this pattern is representative of the population or is sample-specific.

Due to low service usage and technical difficulties in retrieving information for telephone calls, the meaningfulness of accuracy rates for assistors is questionable. Future research will need to address technical issues encountered with the phone service. It is also important to note that face-to-face assistors are aware of the research project and who is in the research project, whereas telephone assistors are not. Any differences in accuracy rates between telephone and face-to-face assistors may be an artifact of the research project.

The results of this study offer some suggestions for improving future research in this area. One suggestion for increasing the service usage rate is to make a portion of the honorarium based on accuracy. This should increase the participant's motivation and more closely approximate real-life motivation levels. It is hoped that this incentive structure will discourage participants who are simply participating for the money and increase the incidence of participants who are more intrinsically motivated, thereby increasing service usage and overall accuracy.

Another suggestion is to improve the quality of recruitment efforts. There is some evidence that a few of the participants in the current study were not recruited properly and failed to meet recruitment requirements. Specifically, some participants were not screened properly and we strongly suspect that others were aware of the recruitment requirements and lied in order to participate. While participants who were known to have been improperly recruited were dropped from the analyses, we do not know how many other participants were improperly recruited or who manipulated the recruitment system to participate when they were not truly qualified. Therefore, recruitment errors may have contributed to the overall low accuracy rate in the current study. For future studies, recruitment screening should be designed to decrease the ability of recruits to determine the reasons for termination. Also, quality assurance screening should be conducted on recruiters to ensure that proper recruitment procedures are being followed.

Another suggestion for future studies is to change the tax preparation requirement. For the current study, participants were required to have completed their own tax returns without the use of a tax professional or tax preparation software at least once in the last 5 years. This qualification was instituted because people process completely new information differently than they process familiar information. This is why participants were given tax scenarios similar to their own tax situations. How people use service when encountering a new tax situation is a topic for future research. Based on the low overall accuracy rate, as well as questions and statements made by participants while receiving service, and informal conversations between some participants and researchers, the once-in-five-year qualification may not have been stringent enough to ensure that people were familiar with completing their tax returns and the 1040 tax form. It is suggested that future research change the requirement to once in the last 3 years to increase familiarity with tax forms.