A COMPARATIVE ANALYSIS OF REPORTING COMPLIANCE BEHAVIOR IN LABORATORY EXPERIMENTS AND RANDOM TAXPAYER AUDITS*

Kim M. Bloomquist, U.S. Internal Revenue Service

USING RANDOM AUDITS OF INDIVIDUAL TAX returns to develop estimates of reporting compliance is an expensive and labor intensive undertaking requiring the support of hundreds of tax professionals to perform a variety of tasks such as designing statistically reliable samples, auditing tax returns, creating databases and analyzing data. However, largely due to this substantial investment of effort and resources, there is broad agreement that random audit studies provide the most accurate measure of taxpayers' compliance behavior. Yet, despite this clear advantage, random audit data often cannot provide answers to important questions facing tax administrators. A typical such question might be: How much would voluntary reporting compliance improve if X more tax returns were audited? Using random audit data to try to answer such a question is technically challenging for at least two reasons. One, while U.S. audit rates have varied over time, the historical level of variation is arguably imperceptible to most taxpayers. For example, the U.S. Census Bureau's (2009) report states that the percentage of individual tax returns examined fell from 1.7 percent in 1995 to 0.5 percent in 2000 and recovered to 1.0 percent in 2007. Two, even if we assume taxpayers adjust their compliance behavior for minor year-to-year changes in audit coverage, such adjustments would be difficult if not impossible to detect in an environment where tax laws and economic conditions are in continuous flux.

The inability of random audit data to provide operational answers for specific tax administration issues has motivated research in the emerging field of experimental economics (Friedland, Maital, and Rutenberg, 1978; Becker, Buchner, and Sleeking, 1987; Webley, Robben, Elffers, and Hessing, 1991; Alm, Jackson, and McKee, 1992; Alm and McKee, 2006; Alm, Deskins and McKee, 2009).[1] A key advantage of laboratory experiments over other data collection methodologies (e.g., taxpayer random audits, field experiments) is the ability to control for exogenous influences (Smith, 1982). In the lab, the experimenter determines what information to make available to subjects, the size and frequency of monetary and other payoffs and the nature of the response space. Given this high degree of control, data obtained from well-conducted experiments may be as free from exogenous influences as is possible to achieve using human subjects.

Before we can use data from laboratory experiments to make general statements about human conduct it is necessary to assume that behavior observed in the lab is a close parallel (subject to sampling error) to behavior in the naturally occurring world. However, in a recent review article, List and Levitt (2007) caution researchers about making this assumption. Their chief concern is an excessive amount of "pro-social" behavior they claim human subjects display under laboratory conditions. List and Levitt (2007) identify several factors they believe cause subjects in laboratory experiments to behave in a more pro-social manner. These factors include: a higher level of scrutiny in the lab versus the real world, subjects' relative lack of anonymity, context and influences on decision making outside the control of experimenters, the size of stakes involved, differences between the typical experimental subject (e.g., college sophomore) and the broader population, artificial restrictions on choice sets, and a limited time horizon. While List and Levitt (2007) acknowledge experiments can be useful to obtain qualitative insights on individual preferences, they contend data collected in lab experiments should not be used to develop quantitative estimates for "deep" structural parameters.

This paper presents a comparative analysis of reporting compliance behavior in laboratory experiments and a sample of randomly audited tax returns of sole proprietors. The random audit data are from the Internal Revenue Service's (IRS) National Research Program (NRP) sample of individual taxpayers for tax year (TY) 2001. The analysis finds that the data from laboratory experiments contain features that are qualitatively *and* quantitatively similar to random audit data. These features include: a bimodal distributional shape, the presence of a small minority of individuals with

*The views expressed here are those of the author and should not be interpreted as those of the U.S. Internal Revenue Service (IRS).

100 percent reporting compliance and comparable individual mean reporting compliance rate.[2] These findings indicate, contrary to the views of List and Levitt (2007), that data from well-designed laboratory experiments can provide useful insights on tax reporting behavior with respect to changes in the level of audit probabilities.

The paper is organized as follows. The next section describes the data from laboratory experiments and taxpayer random audits selected for comparison. This is followed by a comparative analysis of the reporting compliance rates from the two data sets. The fourth section critiques the List and Levitt (2007) argument against using data from laboratory experiments to generalize about real-world behavior in light of the findings reported in this paper. The final section concludes.

DATA

Taxpayer Data

The taxpayer data are a subset of the IRS's National Research Program (NRP) random audit study for tax year (TY) 2001 (Bennett, 2005). The full NRP sample consists of 44,768 cases representing approximately 126 million taxpayers who filed timely tax returns in TY 2001. The subsample of the NRP data set selected for this study consists of individuals whose sole source of income (pre- and post-audit) is from a Schedule C sole proprietorship. This group consists of 1,673 records representing 1,101,977 taxpayers. The restriction on source of income was made to remove any indirect influence on reporting behavior from sources of income covered by third-party information reporting (e.g., Form W-2 for wage and salary income). The subsample was further restricted to Schedule C filers having positive taxable income as determined by the examiner. This additional limitation was imposed so that taxpayers and experimental subjects share similar decision making circumstances; that is, both groups of individuals must decide how much of an initial positive amount of income to report to tax authorities. The final taxpayer sample has 1,101 cases representing 559,555 individuals. Within this data set there are 29 cases where reported taxable income exceeds the amount of taxable income following examination. These cases (representing 13,131 taxpayers) were assumed to have 100 percent reporting compliance.

Table 1 displays summary statistics for the taxpayer sample. The figures in the two rightmost columns are the mean of the individual reporting compliance rates and the overall mean reporting compliance rate. The latter is simply the sum of reported taxable income divided by the sum of taxable income per exam. The range of taxable income per exam for this sample spans five orders of magnitude from less than $40 to more than $4 million. The probability of audit for individual taxpayers as a whole in calendar year 2002 was 0.57 percent and 1.72 percent for all Schedule C filers (U.S. Department of the Treasury, 2003).

Figure 1 displays two histograms of the distribution of individual reporting compliance rates (unweighted and weighted) for the taxpayer sample. Both histograms evidence a bimodal shape with the largest mode at 0 percent compliance. The smaller mode at 100 percent compliance supports an earlier observation by Erard and Ho (2003) using TY 1988 Taxpayer Compliance Measurement Program (TCMP) data of a group of "pathologically honest" individuals.

Experimental Data

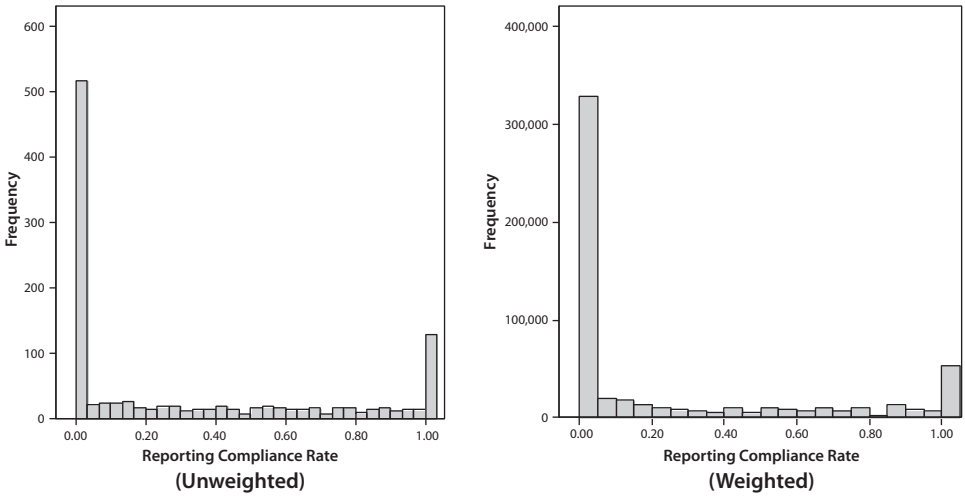
The experimental data used in this study were collected from student subjects in undergraduate classes in economics and business as part of an

Table 1
Summary Statistics for Taxpayer Sample

	Taxable Income as Reported				*Taxable Income per Exam*			*Mean Reporting Compliance Rate*	
	N	*Mean*	*Std Dev*	*Sum ($M)*	*Mean*	*Std Dev*	*Sum ($M)*	*Mean of Individuals*	*Overall Mean*
Unweighted	1,101	$5,461	$12,081	$6.1	$25,277	$132,064	$27.8	0.313	0.216
Weighted	559,555	$3,708	$9,854	$2,075.0	$16,054	$78,165	$8,983.3	0.242	0.231

Source: See text.

Figure 1: **Distribution of Reporting Compliance Rate for Taxpayer Sample**



IRS-funded laboratory experiment described in McKee, Alm, and Jackson (2005), Alm and McKee (2006), and Alm, Deskins, and McKee (2009). The following summary of the experimental design and data collection methodology draws heavily from those sources and from McKee, Alm and Jackson (2005).

The tax compliance experiment captures the essential features of the voluntary income reporting and tax assessment system used in the United States and many other countries. Human subjects in a controlled laboratory environment earn income through performance in a task, where the actual income earned is determined by the (relative) performance in the task.[3] Once the earning task is completed, the subjects are informed via the computer of their income for the round and presented with a screen that resembles a tax form in which they report their income. This screen informs the subjects of the current tax rate, the current probability of an audit, and the penalty rate applied to non-disclosed income. The language in the experiment uses tax lexicon, unlike some other experimental work.

The subjects must decide how much of this income to report to a tax agency. Taxes are paid on reported income, and no taxes are paid on unreported income. However, unreported income may be discovered via an audit and the subject must then pay a fine on the unpaid taxes. This reporting, audit, and penalty process is repeated for a given number of rounds that each represent a tax period, and is replicated with different sets of subjects. At the completion of the experiment, each subject is paid an amount that depends upon his or her performance during the experiment. Subject earnings ranged from $19 to $37, depending upon his or her performance during the experiment. Subjects were paid in private at the end of the session. Subjects were told that all responses are anonymous, and that the only record of participation that contains their name is the receipt signed when they receive their payments.

The currency used in the experiment is called "lab dollars," and subjects are told that all lab dollars they earn during the experiment will be redeemed for cash at the end of the experiment at a fixed conversion rate of 90 lab dollars per 1 U.S. dollar.

The process for determining who is audited is randomly generated by the computer. Where the audit probabilities are announced in advance, the subjects observe a computerized draw of a colored ball from a bucket on the subject's computer screen. The relative number of colored balls in the cage shows the actual probability. All audits investigate only the current period disclosure.

The full data set obtained from this experiment contains 16,560 observations on 1,072 individual subjects. In addition to "base case" or "no treatment" sessions, data were collected on several treatment scenarios including the existence of a

public good, unofficial communication among participants, and "official" communication from the tax authority. For this paper only, data from the base case sessions are used. This restriction was made to preclude the influence of the various treatments on reporting compliance. The selected subset has 3,780 observations from 252 individuals. The income earned by each subject in each round ranged from a low of 20 to a high of 100 lab dollars.

Descriptive statistics for the full data set and the selected data set are shown in Table 2. Five audit probabilities were used in the base case sessions. One session with 16 subjects was conducted with zero probability of audit. As mentioned previously, subjects were informed of the audit probability (including an audit probability of zero) prior to each round of reporting. As was the case in Table 1, the figures in the two rightmost columns of Table 2 are the mean of the individual reporting compliance rates and the overall mean reporting compliance rate.

Figures 2a and 2b display histograms of the individual reporting compliance rates for the lab subjects. The histogram for audit probability of 0.40 shows a continuation of the trend of increasing numbers of subjects reporting at 100 percent compliance but is omitted here for brevity.

COMPARATIVE ANALYSIS

In this section the two data sets just described are compared to identify similarities and differences. Of primary interest is the mean reporting compliance rate, the distributional form and the relative level of fully compliant reporting behavior.

Mean Reporting Compliance Rate

In Table 2 the mean individual reporting compliance rate for the selected sample of experimental subjects is 0.288 when the audit rate is zero and 0.404 when the audit rate is 0.05. These two values bracket the unweighted mean individual compliance rate of 0.313 for the taxpayer sample (Table 1) for which the actual audit probability was 0.0172. However, the weighted mean individual compliance rate of 0.242 is considerably below the experimental results. Arguably, the appropriate comparison in this case is between the unweighted mean values since weights are not available for the experimental data. Thus, we can tentatively conclude that the mean individual compliance rates for the laboratory experiments and actual taxpayers are comparably close in value. In the next section, we investigate the weighted mean compliance rate in the taxpayer data to see if size of stakes or other factors led to the lower value.

Distributional Form

A comparison of the histogram of unweighted data in Figure 1 to Figures 2a and 2b reveals the following similarities. First, both data sets have bimodal distributions with modes at 0 and 100 percent reporting compliance. Second, the mode at 0 is larger than the mode at 100 percent (except when the audit rate = 0.30 in the experimental sample). Lastly, there is what appears to be a random scattering of observations between the two modes. In

Table 2
Summary Statistics for Experiment Sample

	Full Sample			*Selected Sample*			
						Mean Reporting Compliance Rate	
Audit Probability	*N of Subjects*	*N of Obs*	*Mean Reporting Compliance Rate*	*N of Subjects*	*N of Obs*	*Mean of Individuals*	*Overall Mean*
0.00	16	240	0.288	16	240	0.288	0.286
0.05	180	2,700	0.413	48	720	0.404	0.368
0.10	356	5,580	0.544	78	1,170	0.475	0.476
0.30	298	4,710	0.590	32	480	0.558	0.536
0.40	222	3,330	0.638	78	1,170	0.672	0.668
Total	1,072	16,560	0.551	252	3,780	0.521	0.517

Source: See text.

Figure 2a: **Distribution of Reporting Compliance Rates for Lab Subjects**

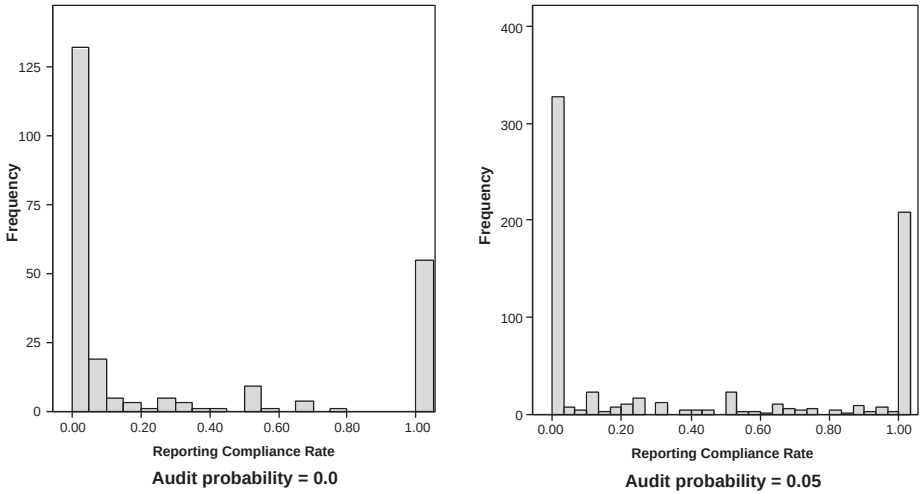
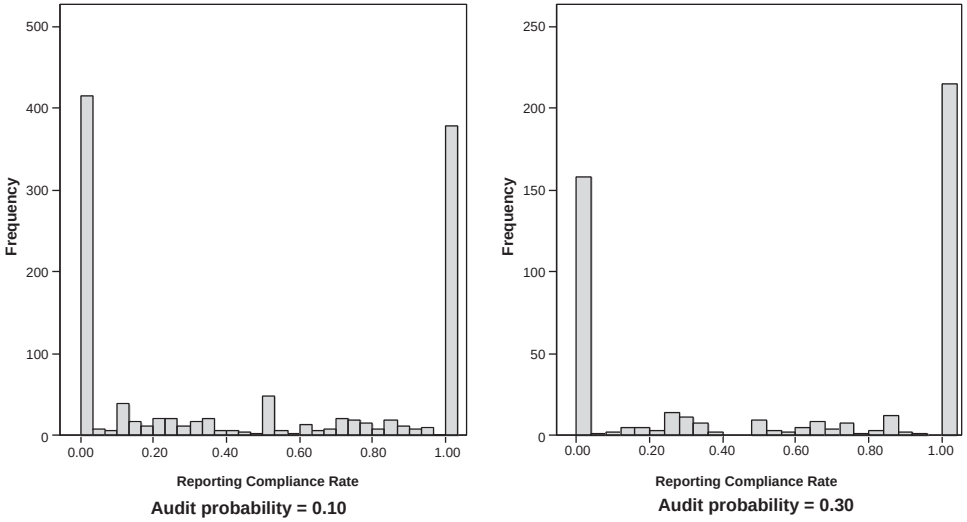


Figure 2b: **Distribution of Reporting Compliance Rates for Lab Subjects**



summary, the distributions of individual compliance rates for the two data sets are qualitatively similar with regard to their overall form and mass, particularly at low audit probabilities.

Relative Share of Fully Compliant Reporting Behavior

In the experimental data for sessions with audit probability = 0.05 there were 8 of 48 subjects (16.7 percent) with 100 percent compliance in all 15 rounds. Similarly, in the experimental data for sessions with audit probability = 0.10, there were 13 of 78 subjects (16.7 percent) with 100 percent compliance in all 15 rounds.[4] In the taxpayer sample, 129 of 1,101 individuals (12 percent) were found to be fully compliant and 152 (14 percent) had a compliance rate of 95 percent or higher.[5] Although it would be premature to draw any final

conclusions from these few observations, the fact that the share of fully compliant individuals in both data sets is relatively close should call into question the view of List and Levitt (2007) and many others that, at least with respect to tax reporting behavior, the behavior of subjects in laboratory experiments is characteristically different from behavior in the real world.

DISCUSSION

In contrast to the relatively similar behavior of real-world taxpayers and student subjects in tax compliance laboratory experiments just described, List and Levitt (2007) assert that conditions in a laboratory environment differ from the naturally occurring world in ways that cause individuals to exhibit more pro-social behavior. First, according to List and Levitt (2007), subjects in laboratory experiments are likely to perceive they are under greater scrutiny causing them to behave less selfishly. The authors cite the example of a lab experiment involving baseball card sellers who were willing to sell high-quality cards at a lower price in the lab than the same individuals were willing to offer at a sports card show when they were unaware they were being observed (List, 2006).

However, in the context of taxpayer compliance, it isn't clear that subjects perceive a heightened sense of scrutiny in the lab versus outside the lab. In fact, it is probably fair to say most taxpayers believe they are under some degree of scrutiny by tax authorities. Taxpayers in the United States are informed that income from their employer, their bank, and their mutual funds is being reported to the IRS. Moreover, most people are aware that the IRS audits some tax returns every year. Thus, unlike in the case of baseball card vendors engaged in private transactions, it is unclear whether the level of scrutiny perceived by subjects in a tax compliance laboratory experiment is much different that what taxpayers perceive in the real world.

Moreover, the existence of "excess" pro-social behavior in the lab claimed by List and Levitt (2007) could be due to initial uncertainty about the "rules of the game." Such uncertainty is often overcome with repeated rounds of play. This is seen clearly from the experimental data shown in Figure 3. Figure 3 plots the mean reporting compliance rate by audit probability and round for the experimental data described in the previous section. The highest reporting compliance occurs in the first few rounds, falls gradually as play proceeds and levels off about midway through the experiment.

The gradual reduction in reporting compliance over time in the lab occurs mainly as more participants report zero dollars as shown in Table 3. Such behavior would seem to indicate a process of

Figure 3: **Mean Individual Reporting Compliance Rate by Round: Experimental Sample**

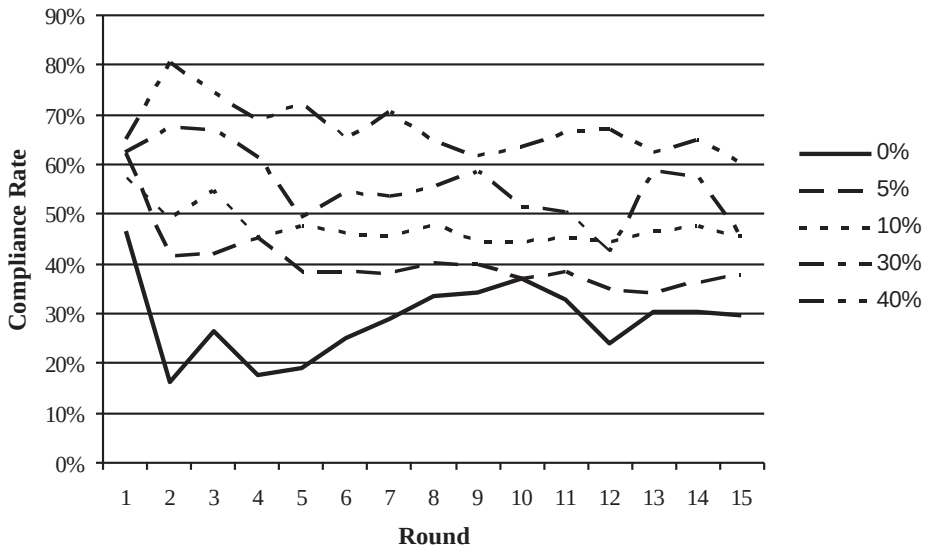


Table 3
Compliance Behavior of Lab Subjects in the First and Last Rounds

Audit Rate	*Total Number of Subjects*	*Number of Subjects Reporting Zero Taxable Income*		*Mean Reporting Compliance Rate*	
		Round 1	*Round 15*	*Round 1*	*Round 15*
0.00	16	4	7	0.47	0.30
0.05	48	5	23	0.62	0.38
0.10	78	19	26	0.57	0.46
0.30	32	5	14	0.62	0.45
0.40	78	14	19	0.65	0.60

Source: Author's calculations.

learning through repeated play. Over time, subjects gradually gain confidence in their ability to evade and not get caught. The last two columns of Table 3 compare the mean reporting compliance rate in the first and last rounds of the experiments.

A second issue cited by List and Levitt (2007) is the degree of anonymity provided to participants in laboratory experiments. The authors claim that when experimenters fail to maintain the anonymity of subjects' responses (e.g., through the use of a "double-blind" approach), this may cause some individuals to behave less selfishly than they would if their identities remained anonymous. They cite a public goods experiment by List, Berrens, Bohara and Kervleit (2004) that found as decisions became less anonymous, a greater number of subjects contributed to the public good in a one-shot game. However, as this example shows, the degree of anonymity is something that can be controlled by experimenters. Moreover, most tax compliance experiments last for multiple rounds (typically 15 or more) and only optionally do subjects' payoffs depend on their contribution to a public good (Alm, Jackson, and McKee, 1992; Becker et al., 1987).

Context is a third area where lab experiments differ from the real world, according to List and Levitt (2007). Context relates to the complex set of social norms, perceptions, and past experiences built up by individuals over time that determine our behavior in new or unfamiliar situations. For example, taxpayers in the real world can be given prison time for failing to pay their taxes. Consequences this severe cannot be replicated in a laboratory setting (one hopes). The issue of how context manifests itself in different behavior in the real world versus the lab remains an open empirical question, but one that can be investigated using careful and creative experimental design.

List and Levitt (2007) cite the difference in size of stakes involved in lab experiments versus the real world as a fourth area of concern. They offer examples from trust and gift exchange games indicating that subjects become less generous as the size of stakes increase. Clearly, the size of stakes involved in many real-world transactions greatly exceeds what can be reasonably simulated in an experimental setting. However, since experiments have been used to demonstrate the inverse correlation between size of stakes and pro-social behavior, this suggests that laboratory experiments are able to provide important insights on human behavior despite experimenters' inability to offer the full range of real-world incentives. Like other laboratory-based investigations, the stakes offered in tax compliance experiments typically are smaller than in the real world. The question is how much more self-interested do taxpayers become when presented with larger stakes?

In an attempt to answer this question, I examine compliance rates by amount of per exam income in the taxpayer sample to determine if reporting compliance varies with level of income. If self-interested behavior increases with the size of stakes, then we should expect to see less compliance as income increases. The results are shown in Table 4 along with results of a difference of means *t*-test where the mean reporting compliance rate of each income category was compared to the mean reporting compliance rate of taxpayers in all other income categories.

Table 4 shows that reporting compliance falls in both directions away from incomes in the $30,000 to $40,000 range (weighted cases) except for the

Table 4
Mean Individual Reporting Compliance Rate by Income Category

Per Exam Income ($K)	*Unweighted*		*Weighted*	
	Mean Reporting Compliance Rate	*N*	*Mean Reporting Compliance Rate*	*Weighted N*
<1	0.237	43	0.248[a]	36,083
1-5	0.318	241	0.19[a]	147,022
5-10	0.363	242	0.225[a]	142,509
10-20	0.329	246	0.286[a]	121,441
20-30	0.319	120	0.327[a]	50,468
30-40	0.326	70	0.357[a]	23,619
40-50	0.233	40	0.186[a]	14,287
>50	0.184[a]	99	0.160[a]	24,125
Total	0.313	1,101	0.242	559,555

Source: See text.
[a]*t*-test $p < 0.01$

lowest income category. The apparent nonlinear relationship between reporting compliance and income would seem to indicate that size of stakes alone does not account for the pattern of compliance behavior among actual taxpayers.

The other interesting point to note about the data in Table 4 is the large difference between unweighted and weighted mean compliance rates for filers having per exam income between $1,000 and $10,000. Clearly, there is something about individuals in this income range that influences the group compliance rate when weights are applied. For all other income categories, the difference between the weighted and unweighted compliance rate is much smaller.

A fifth concern raised by List and Levitt (2007) concerns the fact that laboratory subjects often are drawn from student populations. The fact that many college-age students have had little or no prior experience filing and paying income taxes might suggest that their behavior would differ from that of experienced taxpayers. However, the results reported in this paper show there is little apparent difference between student subjects in laboratory experiments and taxpayers who face similar reporting decisions. The impact of taxpaying experience on compliance behavior in laboratory experiments is a topic of current IRS-funded research.

Finally, List and Levitt (2007) note that, compared to situations in the naturally occurring world, participants in lab experiments are constrained in the range of choices they can make and the time in which to make them. They suggest these artificial constraints imposed by laboratory conditions produce excessive pro-social behavior. In the real world, individuals have the option of filing or not filing a tax return and, conditional on filing, how much to report. These important features are present in the software used in tax compliance experiments funded by IRS. While it is undoubtedly true that taxpayers spend more time on tax filing activities in the real world than in the lab, it is unclear having less time leads to more (or less) pro-social behavior. This topic also can be explored using variations on the basic experimental design.

CONCLUSION

This paper compares the reporting compliance behavior of Schedule C filers subjected to IRS random audits to college-age participants in a tax compliance laboratory experiment and finds key similarities using both qualitative *and* quantitative criteria. These similarities include: (1) mean individual reporting compliance rate of 0.313 for unweighted NRP cases (implied audit probability of 0.0172) versus 0.288 and 0.404 for the experimental data for audit probability of 0.00 and 0.05, respectively; (2) a bi-modal distribution of reporting compliance rates with the largest mode at a reporting compliance rate of 0 percent and the second mode at 100 percent reporting compliance;

and (3) approximately equal relative shares of fully compliant individuals. While further research is needed to confirm these findings, the differences in size of stakes and age distribution of individuals in the two groups do not appear to result in markedly different reporting compliance behavior.

In their critique, List and Levitt (2007) caution that subjects' perception of a heightened sense of scrutiny in the lab leads to excess pro-social behavior. However, in the context of tax compliance experiments, this criticism may not be salient as subjects are well aware that tax returns are scrutinized by tax authorities. Finally, thoughtful and creative experimental design, clear instructions, and multiple rounds of interaction can improve parallelism with the naturally occurring world.

Notes

[1] Two other sources of data on taxpayer behavior are field experiments and surveys. While some researchers see field experiments as an alternative mid-way between lab experiments and real-world observations (List and Levitt, 2007); relatively few tax evasion field experiments have been conducted. The study by Slemrod, Blumenthal, and Christian (2001) is the most referenced example. Field experiments typically are more costly to conduct than lab experiments and the results observed may be influenced by exogenous factors outside the control of the experimenters. Taxpayer surveys are the primary data collection instrument used to measure customer satisfaction with respect to various services. However, taxpayer surveys are considered to be a less reliable source of evasion behavior due to the understandable reluctance of some survey participants to provide candid responses.

[2] Mean individual reporting compliance rate is defined as the mean of all individuals' reporting compliance rates. Each individual's reporting compliance rate is calculated as reported taxable income divided by examiner-determined taxable income.

[3] The earnings task requires the subjects to sort the digits 1 through 9 into the correct order from a randomized order presented in a 3-by-3 matrix. They do this by pointing the computer mouse at the numbers in the correct sequence and "clicking" on the numbers. Actual income is determined by the relative speed of performance, with the fastest performer receiving the highest income and the slowest performer receiving the lowest income.

[4] In the experimental sample with zero audit probability (16 individuals in a single session), 1 subject was compliant in all 15 rounds (6.25 percent) and 1 subject was compliant in 14 of 15 rounds (reporting no income in the second round but otherwise 100 percent of income in the other 14 rounds). On average, there were 3.67 subjects with 100 percent compliance per round in the session with zero audit probability.

[5] On a weighted basis, 9.5 percent of taxpayers in the NRP sample were fully compliant.

References

Alm, James, John Deskins, and Michael McKee. Do Individuals Comply on Income Not Reported by Their Employer? *Public Finance Review* 37 (March 2009): 120-141.

Alm, James, Betty R. Jackson, and Michael McKee. Estimating the Determinants of Taxpayer Compliance with Experimental Data. *National Tax Journal* (March 1992): 107–114.

Alm, James and Michael McKee. Audit Certainty, Audit Productivity, and Taxpayer Compliance. *National Tax Journal* 59 (December 2006): 801-816.

Becker, Werner, Hans Buchner, and Simon Sleeking. The Impact of Public Transfer Expenditures on Tax Evasion: An Experimental Approach. *Journal of Public Economics* 34 (November 1987): 243–252.

Bennett, Charles. Preliminary Results of the National Research Program's Reporting Compliance Study of Tax Year 2001 Individual Returns. Paper presented at the Annual IRS Research Conference, Washington, D.C., 2005. http://www.irs.gov/pub/irs-soi/05bennett.pdf

Erard, Brian and Chih-Chin Ho. Mapping the Compliance Continuum: From Pathologically Honest to Flagrantly Defiant. Paper presented at the conference on "The Hard to Tax: International Perspective, Andrew Young School of Policy Studies, Georgia State University, 2003.

Friedland, Nehemiah, Shlomo Maital, and Aryeh Rutenberg. A Simulation Study of Income Tax Evasion. *Journal of Public Economics* 10 (October 1978): 107-116.

List, John A. The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions. *Journal of Political Economy* 114 (February 2006): 1-37.

List, John A., Robert Berrens, Alok Bohara, and Joe Kervliet. Examining the Role of Social Isolation on Stated Preferences. American Economic Review 94 (June 2004): 741-752.

List, John A. and Steven D. Levitt. What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives* 21 (Spring 2007): 153-174.

McKee, Michael, James Alm, and Betty Jackson. Final Report for TIRNO-03-C-00076, Estimating the Indirect Effects of Taxpayer Audits. 2005.

Slemrod, Joel, Marsha Blumenthal, and Charles Christian. Taxpayer Response to an Increased Probability of an Audit: Evidence from a Controlled Experiment in Minnesota. *Journal of Public Economics* 79 (March 2001): 455–483.

Smith, Vernon L. Microeconomic Systems as an Experimental Science. *American Economic Review* 72 (December 1982): 923–955.

U.S. Census Bureau. 2009 Statistical Abstract, Table 463. Individual Income Tax Returns Filed – Examination Coverage. Washington, D.C., 2009.

U.S.Department of the Treasury. Internal Revenue Service. IRS Data Book 2002. Publication 55B. Washington, D.C. 2003. http://www.irs.gov/pub/irs-soi/02databk.pdf

Webley, Paul, Henry Robben, Henk Elffers, and Dick Hessing. *Tax Evasion: An Experimental Approach*. Cambridge: Cambridge University Press, 1991.