Comparing Student and Non-Student Reporting Behavior in Tax Compliance Experiments

James Alm, Tulane University; Kim M. Bloomquist, Internal Revenue Service; and Michael McKee, Appalachian State University

Laboratory experiments have been increasingly used to examine various issues in tax compliance. The use of laboratory experiments in economics began in the early 1960s with the establishment of a well-defined framework for experimental work by Smith (1976, 1982), and laboratory methods are now widely accepted as a methodological approach in the analysis of theory and policy, especially of behavioral economics. For comprehensive surveys of experimental methods, see Davis and Holt (1993) and Kagel and Roth (1995).

Tax compliance is an area that seems especially amenable to laboratory experiments, given limitations in empirical approaches based on field data. As discussed in more detail later, theoretical models are not able to incorporate fully, appropriately, or tractably many factors deemed relevant to the individual compliance decision; also, these modeling efforts can benefit from the behavioral insights obtained through laboratory investigations that illuminate many of the factors relevant to the individual compliance decision. Empirical studies of tax compliance using field data are plagued by the absence of reliable information on individual compliance decisions: it is difficult to measure—and measure accurately—something that by its very nature people want to conceal, and it is difficult to control in econometric work for the many unobservable factors that may affect the compliance decision. In contrast, laboratory methods allow many factors suggested by theory to be introduced in experimental settings. Also, experiments generate precise data on individual compliance decisions, which allow econometric estimation of individual responses in ways that are simply not possible with field data. Indeed, laboratory methods have examined a wide range of factors in the compliance decision, factors that have not proven amenable to either theoretical analyses or empirical analyses with field data.

However, laboratory studies of compliance are also sometimes viewed with some skepticism. The most common criticism of experimental investigations of tax compliance behavior is that the student subjects typically used in experiments may not be representative of taxpayers. Undergraduates may have little experience with filing tax returns, and their economic and demographic backgrounds may differ from that of taxpayers. It is this issue that we examine here.

We present evidence that relates directly to the use of student subjects. In particular, we compare the experimental responses of *student subjects* to the experimental responses of *non-student subjects*, in *identical experiments*; that is, do students behave differently than non-students in identical experiments? These data are generated from a series of experiments conducted as part of an Internal Revenue Service (IRS) funded research program, in which student subjects and non-student subjects participated. We examine the mean levels of compliance, the frequency distribution of compliance rates, and the econometrically estimated behavioral responses. While we find that the mean *levels* of compliance of students are not always the same as non-students, the behavioral *responses* of students in laboratory experiments to policy innovations are largely the same as non-students in identical experiments. Also, the frequency distributions of individual compliance rates are virtually identical for students and non-students. (Note that in other work we compare experimental data versus non-experimental data from the National Research Program of the IRS, in an attempt to answer whether students in experiments behave differently than non-students in naturally occurring settings. These other results address explicitly the issue of generalizing from experiments to the naturally occurring world (e.g., the "external validity" of tax compliance experiments). These results are not presented here, but these comparisons also indicate largely similar patterns. See Alm, Bloomquist, and McKee (2011) for a detailed

analysis of all student versus non-student and experimental versus non-experimental comparisons.) Our results largely confirm that the observed behaviors of student and non-student subjects are qualitatively and quantitatively similar.

The Uses and Misuses of Experimental Economics

Experimental economics involves the creation of a real microeconomic system in the laboratory, one that parallels the naturally occurring world that is the subject of investigation and one in which subjects (usually students) make decisions that yield individual financial payoffs whose magnitude depends on their decisions. The essence of such a system is control over the environment, the institutions, the incentives, and the preferences that subjects face. Of these, control over preferences is particularly crucial. As emphasized by Smith (1976), "[s]uch control can be achieved by using a reward structure to induce prescribed monetary value on actions."

Why use experimental methods? On some level, the use of experimental methods derives from a fundamental problem with economics. Like other sciences, economics is based on the development of theory and on the ability of that theory to explain observed activities. However, unlike some other sciences, especially the natural sciences, economics faces substantial difficulties in empirically testing the predictive power of its theories using data from the naturally occurring world. Given the dizzying array and complexity of forces that operate in market (and non-market) systems, economists can never be quite certain that they are "holding constant" the many factors that may be driving individual choices, so that they can focus on the "true" driving factors that are the object of empirical testing. Methods for achieving such identification have become increasingly sophisticated over time, especially with the use of so-called "natural experiments" and "controlled field experiments". Even so, there are few instances in which such identification is uncontroversial.

There are of course numerous avenues for testing the predictions of economics, aside from experimental methods, including the use of naturally occurring field data, hypothetical choices, natural experiments, and controlled field experiments. Even so, experimental methods have often been a common approach, and economists have increasingly begun to emulate the methods of natural scientists by conducting carefully controlled laboratory experiments.

Economics generally and public economics specifically have profited from the use of laboratory experiments, for several reasons. Econometric data on research questions obtained from the naturally occurring world can be unreliable, can fail to show the variation or distinctions of interest, or can fail to provide sufficient identification to discern "cause and effect". Indeed, in some cases data simply cannot be assembled outside the laboratory because the real world setting of relevance does not exist. For its part, theoretical analyses often cannot incorporate fully, appropriately, or tractably many relevant factors.

What can experiments do? Unlike standard theoretical work, experiments are not as constrained by the same degree of simplification required in analytical studies, which allows the impact of numerous factors not amenable to theoretical work to be examined precisely and unambiguously in a controlled environment. Unlike traditional empirical work based on naturally occurring data, experiments generate data under settings in which there is control over extraneous influences. Laboratory experiments also provide a controlled environment that allows one to examine the mechanisms of interest, as well as the changes in these environments and institutions, in isolation from each other.

Given the limitations of theoretical and econometric work, there are, we believe, compelling reasons for the use of experiments, as an *additional* (and not as the *only*) methodological tool, in large part because laboratory experiments give a researcher the twin advantages of *control* (including *data generation* and *replicability* from this control) and *flexibility*.

However, despite the demonstrated usefulness of experimental methods, there are sound reasons for caution in interpreting and generalizing experimental results. Perhaps the most common criticism of experimental economics is that the student subjects typically used may not be representative of taxpayers (Levitt and List, 2007). Although there is now much evidence that the experimental responses of students are seldom different than the responses of other subject pools (Plott, 1987), there are no comparisons of student versus non-student behavior in the specific context of tax compliance. It is this issue that we examine here.

Some Simple Comparisons of Students and Non-students in Identical Experiments

We consider the responses of student subjects versus non-student subjects in identical experiments. These comparisons are based upon results in Alm, Cherry, Jones, and McKee (2010, 2011) who conducted experiments designed to investigate the compliance behavior of individuals under various policy initiatives. In both studies the subject pool consisted of both students and non-students (e.g., university staff and faculty), and we report here a comparison of students and non-students in these different experimental settings. The first setting investigates the effects on tax reporting of the provision of information services by the tax agency in an environment in which subjects may not know with certainty their true tax liability. The second setting introduces positive inducements via social programs as an incentive to truthfully report tax liabilities; these social programs include an income tax credit (in which receipt requires that the subject must file a tax return) and unemployment benefits (in which benefits are a positive function of past reported taxes). We first discuss the experimental design of these experiments, and then we present the comparison of student versus non-student responses in order to answer the question of whether students behave differently than non-students in identical experiments.

The experimental setting implements the fundamental elements of the voluntary reporting system of the individual income tax in most countries. Participants earn income by performing a task, and they self report tax liability to the tax authority. At the time of reporting only the individual knows his or her true level of tax liability, and the subject can choose to report any amount from zero on up. An audit occurs with an announced probability, and any unreported taxes are discovered. If the participant has underreported the tax liability, then both the unpaid taxes and a penalty are collected. This process is repeated over a number of rounds each representing a tax period. Participants are informed that they will be paid their after-tax earnings at the end of the experiment, converted from lab dollars to U.S. dollars at a fixed and announced conversion rate.

Participants are informed, with certainty, of the audit probability, the penalty rate, and the tax rate. The tax rate is set at 35 percent for all sessions; the penalty rate is also fixed for all sessions at 150 percent (i.e., the participants must pay unpaid taxes plus a penalty of 50 percent of unpaid taxes if audited). The audit probability is varied once within the session, and the participants are told that there is zero probability of audit if no tax form is filed. There is no public good financed by the tax payments in order to focus subject attention on the tax setting.

Into this setting, various policy innovations are introduced in the different studies of Alm, Cherry, Jones, and McKee (2010, 2011). A first set of experiments investigates the effects of taxpayer information services on compliance decisions (Alm, Cherry, Jones, and McKee, 2010). Here the tax reporting decision is "complicated" through the introduction of uncertainty regarding the true tax liability, and then information services are provided by the "tax administration" that partially or fully resolve the uncertainty, thereby allowing subjects to compute more easily their tax liabilities. Complicating factors include both a tax deduction (a reduction in taxable income) and a tax credit (a reduction in tax owed, comparable to the U.S. Earned Income Tax Credit), each of which is conditional upon filing. The tax deduction is set at 15 percent of income, and tax credit is introduced in which the credit starts at a given level and declines at a stated rate as income increases. These factors complicate the tax reporting decision. Also, the exact levels of the deduction and the credit are, in some settings, uncertain to the taxpayer at the time of filing, and this uncertainty adds an additional level of complications. These uncertainties on the credit and the deduction are implemented via mean-preserving spreads (with a uniform distribution) in each, where the participants are informed of the means of the allowed credit and deduction and the ranges for each. When information services are provided, information is complete, accurate, and costless to the participant.

A more direct set of positive inducements is investigated in Alm, Cherry, Jones, and McKee (2011). Here positive inducements for filing are introduced in several alternative treatments. In one treatment tax credits are introduced that are available to participants but only to those who file a tax return. In a second treatment a "social safety net" (e.g., unemployment replacement income) is present in which individuals face some probability of unemployment but also in which unemployment replacement income may be provided, with any benefits again conditional upon past filing behavior. There is a known probability of unemployment,

and, if the individual becomes unemployed and earns no income, then they are unemployed for two periods. Unemployment replacement income is received only if the individual has filed a tax return in each of the two previous periods. Other features of the experimental design are identical to Alm, Cherry, Jones, and McKee (2010).

Table 1 summarizes the experimental design of Alm, Cherry, Jones, and McKee (2010, 2011), with the top panel showing the information services design and the bottom panel showing the positive inducements design. Treatment T1 provides a baseline setting that entails no uncertainty and no tax authority information. The second treatment (T2) introduces tax liability uncertainty, in which participants face uncertainty regarding their allowed deduction and tax credit. The third treatment (T3) entails the same uncertainty as in the second treatment, but introduces the option of resolving the uncertainty by receiving information from the tax authority; that is, participants in this treatment are able to click on a button to reveal the true levels of the deduction and the tax credit. In Table 2, treatment T4 establishes a baseline with no positive inducements. In T5 a tax credit is introduced, and in T6 an individual may be able to collect income benefits while unemployed. The unemployment benefits are based on the average filed earnings for the previous two periods. Thus, an individual who chooses to not file can earn no benefits. See Alm, Cherry, Jones, and McKee (2010, 2011) for a detailed discussion of the experimental designs.

TABLE 1. Experimental Treatments

Information Services

	Information Service Provided?	
Tax Liability Uncertain?	**Yes**	**No**
No	T1	—
Yes	T2	T3

Positive Inducement via Social Programs

Positive Inducements Provided?		
No	**Yes, via Tax Credit**	**Yes, via Unemployment Benefits**
T4	T5	T6

TABLE 2. Aggregate Results by Experimental Treatment by Subject Pool

	Mean Reporting Compliance Rate		
Treatment	**All**	**Staff**	**Students**
Information Services			
No Uncertainty (T1)	0.673	0.795	0.618
Uncertainty—No Information (T2)	0.621	0.571	0.689
Uncertainty—Information (T3)	0.704	0.657	0.768
Positive Inducements via Social Programs			
No Positive Inducements (T4)	0.483	0.444	0.504
Tax Credit (T5)	0.599	0.678	0.487
Unemployment Benefits (T6)	0.681	0.709	0.655

The dedicated experimental lab consists of 25 networked computers, a server, and software designed for this series of experiments. Sessions were conducted at a major state university, using both students and staff as participants. Recruiting was conducted using the Online Recruiting System for Experimental Economics (ORSEE). The participant database was built using announcements sent via email to all students and staff. Participants were invited to a session via email, and were permitted to participate in only one tax experiment, although other experimental projects are ongoing at the time and participants may have participated in other types of experiments. Only participants recruited specifically for a session were allowed to participate, and no participant had prior experience in this experimental setting. Methods adhere to all guidelines concerning the ethical treatment of human subjects.

Of most importance, participants included both students and non-students, thereby allowing one aspect of the external validity of experiments to be examined: do students behave differently than non-students in identical experiments? Students were recruited from the pool of undergraduate students at a U.S. public university. Non-students were recruited from faculty and staff at this same university. The student portion of the subject pool covered a very broad range of year in studies and major, and no single major exceeded 8 percent of the pool. The staff pool was similarly diverse, covering all levels of support staff, non-academic professional staff, and faculty. Also, the compensation varied for students versus non-students. For students the rate was 80 lab dollars to 1 U.S. dollar; faculty and staff participants received a higher exchange rate to reflect their higher outside earnings, with a conversion rate of 50 lab dollars to 1 U.S. dollar. Earnings averaged $18 for student subjects. The average payoffs for faculty and staff were $28.

Four hundred forty nine individuals participated in a session in one of the two sets of experiments. In the sessions designed to investigate the role of tax information services on reporting (T1 through T3), there were 131 subjects, 54 percent of whom were students. In the sessions designed to investigate the effects of positive inducements (T4 through T6), there were 318 subjects (68 percent were students). Table 2 reports the aggregate figures for reporting behavior only by treatment and by subject group.

Treatments T1 – T3 concern the taxpayer information services design. These aggregate numbers indicate that uncertainty concerning tax liability results in lower reporting compliance rates but that providing information that resolves the uncertainty increases reporting. With tax liability uncertainty, the overall reporting compliance rate is 0.621 (T2), which is statistically lower than the 0.673 rate without uncertainty (T1). Further, reporting compliance significantly increases when information services are provided in the uncertain setting (T3), or 0.704 versus 0.621.

When these aggregate levels of compliance are broken down by subject type, we see some differences by subject type in reporting compliance rates. In T1, the mean reporting compliance rate of staff is higher (0.795) than the reporting compliance rate of students (0.618), a difference that is statistically different. Similarly, the mean compliance rates of staff versus students are also different for the other two treatments, T2 and T3, although in these latter two treatments students report at higher rates than staff.

However, despite the somewhat different *levels* of compliance of the two subject pools, the *changes* in compliance rates in response to the treatment effects are similar in both pools of subjects. Comparing T2 and T3, we observe that the changes in compliance rates in response to the information services treatments are similar in magnitude and sign. For both subject pools, the provision of information that resolves tax liability uncertainty leads to an increase in the mean reporting rate, by 8.6 percentage points for staff and by 7.9 percentage points for students (or T3 versus T2). The introduction of uncertainty has different effects on these responses by subject pool (or T2 versus T1), reducing reported income for staff and increasing reported income for students; even so, the change in mean compliance rates for students is not statistically significant in these two treatments.

The positive inducements treatments demonstrate a similar pattern. Treatment T4 provides the baseline setting for the investigation of the tax reporting effects of positive inducements. When looking at all subjects combined, the provision of the tax credit (T5) and unemployment benefits (T6) leads in both cases to higher tax reporting relative to T4. Also, both the student subjects and the staff subjects respond positively to the positive inducements. From the results in Table 2, it appears the staff responses to these inducements are greater than for the student subjects, but the qualitative treatment effects results are similar.

Also, the frequency distributions of individual compliance rates are virtually identical for students and non-students in the uncertainty sessions and in the positive inducements sessions.

It should be noted that there are several significant differences between the subject pools, and a simple comparison of means does not hold these differences constant. For example, the different pools are compensated at different rates, with staff compensated at roughly twice the rate of student subjects. Also, the average ages of the two groups are different, with the average age of staff exceeding 30 years and the average age of students slightly exceeding 20 years. There are other differences between the pools as well.

To hold constant these differences, we use a conditional analysis at the individual level to re-examine the initial impressions from the aggregate data. For the entire sample, staff only, and students only, we estimate the effects of various subject features and design parameters on reporting behavior, while holding other factors constant. Our basic specifications estimate individual reporting of each subject by round as a function of subject demographic characteristics (e.g., subject age, subject sex, subject own preparation of tax returns, subject claimed as a dependent on parental tax returns), subject variables that change by round (e.g., income, accumulated earnings, audit probability), and session characteristics (e.g., indicator variables that signify the presence of uncertainty about tax features, of agency-provided information, of a tax credit that the subject can claim on filing a tax report, or of a safety net that (partially) makes up for income lost due to unemployment). We estimate these specifications using Tobit estimation procedures. These detailed econometric results are not presented here, but they confirm even more strongly our initial impressions from the aggregate data. In particular, we see that the *changes* in compliance behavior of the two groups are quite similar, as measured by the estimated coefficients on the policy innovations, even if the average *levels* of compliance differ as in Table 2.

Conclusions

Our results indicate that the experimental behavior of students is often—although not always—similar to the experimental behavior of other subject pools. Although the *levels* of compliance may differ between student and non-student subjects in identical experiments, the *changes* in compliance behavior of these pools largely parallel each other, especially when potentially confounding influences are controlled in a regression framework. Further, the frequency distributions of individual compliance rates for students versus staff are very much the same. Other results that compare experimental data versus non-experimental data from the National Research Program of the IRS also indicate largely similar patterns. Again, see Alm, Bloomquist, and McKee (2011) for complete discussion and analysis of all results.

In sum, our results are consistent with other experimental studies that demonstrate that student and non-student subjects behave and, especially, respond similarly. Indeed, our evidence is consistent with the broader notion that there is also no reason to believe that the cognitive processes of students are different from those of "real" people, at least in the context of tax compliance and in the comparison of changes in behavior.

References

Alm, James, Kim M. Bloomquist, and Michael McKee (2011). "On the External Validity of Tax Compliance Experiments." Tulane University Working Paper. New Orleans, LA.

Alm, James, Todd Cherry, Michael Jones, Michael McKee (2010). "Taxpayer Information Assistance Services and Tax Reporting Behavior." *Journal of Economic Psychology* 31 (4): 577-586.

Alm, James, Todd Cherry, Michael Jones, Michael McKee (2011). "Social Programs as Positive Inducements for Tax Participation." Tulane University Working Paper. New Orleans, LA

Davis, Douglas D. and Charles A. Holt (1993). *Experimental Economics*. Princeton, NJ: Princeton University Press.

Kagel, John H. and Alvin E. Roth, eds. (1995). *The Handbook of Experimental Economics*. Princeton, NJ: Princeton University Press.

Levitt, Steven D. and John A. List (2007). "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives* 21 (2): 153-174.

Plott, Charles R. (1987). "Dimensions of Parallelism: Some Policy Applications of Experimental Methods." In *Laboratory Experimentation in Economics: Six Points of View*, edited by Alvin E. Roth, 193-229. New York, NY: Cambridge University Press.

Smith, Vernon L. (1976). "Experimental Economics: Induced Value Theory." *The American Economic Review Papers and Proceedings* 66 (2): 274-279.

Smith, Vernon L. (1982). "Microeconomic Systems as an Experimental Science." *The American Economic Review* 72 (5): 923–955.