

**An Experimental Evaluation of Tax-Reporting Schedules:
A Case of Evidence-Based Tax Administration**

Michael Wenzel

Australian National University, Canberra, Australia

Natalie Taylor

Australian Institute of Criminology, Canberra, Australia

2004, *Journal of Public Economics*, 88, 2785-2799.

Abstract

In an experiment in collaboration with the Australian Taxation Office (ATO), Rental Property Schedules (RPS) were sent to taxpayers for them to itemize their deductions and return the RPS to the ATO. The intervention significantly reduced deductions compared to taxpayers who did not have to return the RPS, taxpayers who received only an information letter, and a no-contact control group – regardless of a cooperative versus deterring letter tone and first-time versus second-time participation of taxpayers. The study sheds light on the mechanisms underlying the effects of tax-reporting schedules and highlights the importance of systematic experimental fieldwork in evaluating regulatory strategies.

Keywords: taxation; compliance; experiment

JEL classification: H26; C93

Corresponding author: Michael Wenzel, Research School of Social Sciences, Australian National University, Canberra ACT 0200, Australia. Phone: +61 2 612 50537, fax: +61 2 612 58503. Email: Michael.Wenzel@anu.edu.au.

An Experimental Evaluation of Tax-Reporting Schedules:

A Case of Evidence-Based Tax Administration

1. Introduction

Lack of compliance with tax laws jeopardizes revenue funds necessary for government services and undermines efficiency and fairness of the tax system. Tax authorities such as the Internal Revenue Service (IRS) in the US and the Australian Taxation Office (ATO) in Australia have the responsibility to collect taxes as specified in the law and to combat noncompliance. One strategy that the ATO uses on a larger scale to increase taxpayers' accountability and thus compliance are tax-reporting schedules. These forms ask taxpayers to report in greater detail on the specific sources of their taxable income, or the specific items of expenditures they claim as offsets against that income. For instance, Rental Property Schedules (RPS) itemize the expenses related to rental property that the law accepts as tax deductions. The ATO may send these forms to taxpayers (usually on the basis of certain risk criteria determined by the ATO) for them to report in detail on the rental deductions claimed in their tax return, usually combined with the warning that failure to provide truthful details will incur penalties. The idea is that schedules make taxpayers account in detail for their deduction claims, increase the correctness of their claims, and thus (assuming that overclaiming of deductions exceeds underclaiming) increase tax revenue. However, does this approach work? If so, how does it work, and is there an optimal way of using RPS which maximizes effectiveness while minimizing the cost of the strategy for the ATO?

To answer such questions and to select measures of enforcement and regulation tax administrations, like other regulators, often – too often – rely on unsystematically gathered

intelligence and untested implicit assumptions about the underlying dynamics of the behavior they regulate. In addition, there is usually no rigorous evaluation of strategies that administrations employ to increase compliance. Hence it is often difficult to say with any certainty whether and why strategies are or are not effective. In this paper we argue that the use of carefully designed controlled experiments (specifically, if used with a cumulative research strategy) can provide invaluable information about the effectiveness of regulatory measures and the conditions of their effectiveness, as well as more general insights about the regulated behavior through testing theoretically derived predictions. We adopted such an approach for questions around the use of tax-reporting schedules, demonstrating the usefulness of the methodology on the one hand, and reporting on the usefulness of schedules, and equivalent measures, as regulatory tools on the other hand.

2. Experimental evaluations in tax administration

The defining principles of experimentation are (1) controlled manipulation of independent variables or treatments and (2) randomized assignment of participants to those treatments (Judd, Smith, & Kidder, 1991). Together, these principles ensure that the various experimental groups differ in the treatment they receive, but are equivalent on all other potentially influential variables because these are randomly and non-systematically distributed across the groups. As a consequence, any differences between experimental groups observed on the dependent variables can be unambiguously and causally attributed to the differences in treatment. This advantage sets experiments clearly apart from other empirical methods. Accordingly, Roth, Scholz, and Witte (1989) expressly recommended randomized experiments also for the evaluation of innovations and interventions in tax administration (see also Boruch, 1989).

However, while laboratory experiments and experimental simulation studies are frequently used in tax compliance research, experimental evaluations of regulatory interventions under real-life conditions, using actual tax return data, are rather scarce. In a pioneering study, Schwartz and Orleans (1967) used a field-experimental design to test the effects of an (implicit) moral appeal that made salient ethical reasons for truthfully paying one's taxes, and sanction threats that made salient the severity of sanctions against tax offenders. Compared to control groups that either received a neutral message or no message at all, the moral appeal increased the amount of actual taxable income reported. In a conceptual replication of the study, McGraw and Scholz (1991) used videotaped messages about the moral implications of tax evasion versus the personal profitability of aggressive tax planning, but they did not find any effects on actual or self-reported taxpaying behavior. Both studies, however, were burdened by low sample sizes. This is particularly problematic given that monetary tax return data as dependent variables are generally characterized by huge variation. Further, both studies tested the effects of interventions applied by researchers outside the tax administration but they did not evaluate regulatory measures used, or to be used, by tax authorities themselves.

More recently, the Minnesota Department of Revenue conducted a large-scale field experiment to measure the effectiveness of different strategies to increase voluntary tax compliance (Coleman, 1997). Employed in this project as academic advisors, Slemrod, Blumenthal, and Christian (2001) report on findings of the part of the experimental design testing for the effects of messages warning taxpayers of an increased probability of audit (for the part of the study concerned with effects of normative appeals, see Blumenthal, Christian, and Slemrod, 2001). A random sample of Minnesota taxpayers, stratified by income (low, medium, high), were either sent a letter from the Commissioner of Revenue, informing them they had been randomly selected in a program that would closely examine individual tax returns, or were

not contacted (control). The changes in reported taxable income between previous and current tax return were analyzed and compared between treatment and control groups. A relatively greater increase in reported taxable income for low and medium income taxpayers in the treatment group indicated that the audit threat increased compliance for these income groups. However, the treatment seemed to have backfired for high-income taxpayers who reported less income when being alerted to the review of their tax return compared to the control group. Slemrod et al. (2001) speculate that wealthier taxpayers, who have more complicated tax cases and access to legal advice, might regard an audit as a negotiation that they strategically open with a 'low bid' (i.e., reporting less income). Alternatively, however, their behavior could also be interpreted as defiance or reactance (Brehm & Brehm, 1981), perhaps against what they consider an inappropriate intrusion into their financial affairs. The findings are interesting and document the value of an experimental approach; however, they also show that the processes involved in deterrence-based interventions are less than self-evident and require further investigation.

3. Tax-reporting schedules in the Australian context

Similar interventions as the one tested by Slemrod et al. (2001) are being used by tax authorities in Australia. Letters informing taxpayers of an increased probability of audit, or of penalties that apply if taxpayers are found to make incorrect statements, are here often combined with the use of schedules on which taxpayers are asked to provide further details than normally required in the tax return. For instance, such an approach has been taken in the case of rental property owners.

Australian residents who own rental property must declare in their tax return all the rental income they obtain in a financial year from these properties. Similarly, they are able to claim rental deductions legitimately associated with their rental properties. In the past, the ATO

identified potential risk characteristics of certain groups of taxpayers where compliance with the tax laws in terms of rental income could be questionable (e.g., no rental income reported but rental deductions claimed) and targeted these “high risk” taxpayers in a proactive way prior to the next lodgment period. It sent out to these “risk” taxpayers *Rental Property Schedules* (RPS) and Instructions as well as a booklet with detailed information on tax issues relating to rental properties. The RPS required taxpayers to provide specific details about income and expenses as related to their rental property. That is, the figure entered for net rental income (gross rental income less claimable rental deductions) was to be justified explicitly in the schedules - something not required in ordinary tax return statements. Taxpayers using electronic lodgment were excluded from this approach because this lodgment mode already required provision of the details the RPS asked for. The ATO claimed to have empirical evidence for the effectiveness of the approach in increasing the amount of net rental income declared.

However, the practice of RPS raises a number of further questions surrounding the issues of their general effectiveness and the mode of their application. First, of course, it would be preferable to have definitive evidence as to the effectiveness of RPS, by comparing taxpayers who were sent a RPS with control groups of taxpayers not contacted at all and taxpayers who were contacted in different ways. Without such comparisons, the apparent effectiveness of the RPS could be due to anomalies in the sample or could indeed be surpassed by better alternatives. Second, the ATO used RPS in the past for what it defined as risk groups of taxpayers.¹ It would be of interest to see whether RPS would produce equivalent positive effects for a broader group of taxpayers, and whether the application of RPS could thus be extended and perhaps routinized. Third, especially if a broader application was to be considered, it would be of interest to know whether the schedules could be used in a more cost-effective way. Is it necessary to send out heavy information booklets with the schedules as was ATO practice, or are schedules with brief

instructions or simply a letter just as effective in increasing compliance and tax revenue? Fourth, which accompanying message is more conducive to the effectiveness of RPS? A letter with a harder tone that refers to penalties and audits and attempts to deter taxpayers from making false statements, or a letter with a softer message that claims to be helpful and informative, seeking a cooperative relationship with taxpayers? Lastly, and important to all these questions, if RPS are effective in increasing net rent declared, *why* are they effective and what are the responsible mechanisms? Answers to this latter question in particular would have implications for our understanding of taxpaying behavior more generally.

4. Possible mechanisms underlying tax-reporting schedules

There are various possible reasons for the effectiveness of RPS as applied by the ATO. A first possibility is that recipients of RPS simply feel targeted and watched; they perceive a greater likelihood of being audited and are made aware of penalties for unlawful behavior. In this case, the provision of schedules should be a rather superfluous element of the approach. Rather, it will suffice to send taxpayers a convincing and deterring ‘information letter’ demonstrating that the ATO has an eye on them (Slemrod et al., 2001). This is dubbed here a ‘mere surveillance effect’.

A second possibility is that the RPS and the information presented with them have an ‘education effect’. While taxpayers could be generally willing to comply with the tax laws, they might lack the knowledge to do so. The RPS and the instructions to fill them in would basically provide taxpayers with assistance and support in making correct tax statements. (However, it could also mean helping taxpayers make correct statements in their own favor, increasing compliance but reducing revenue.) In this case, RPS should have an effect even when they are obviously not used by the ATO to scrutinize tax return details. That is, RPS will be effective

even when the schedules are not to be sent back to the ATO, but rather used by taxpayers only for their own record-keeping and preparation of the tax return.

A third possibility is that both elements interact in their effects. RPS might provide taxpayers with detailed information about what they are required to do, and educate taxpayers about their tax rights and obligations. At the same time, taxpayers might need to believe that the tax office reviews the data and reacts on any suspected irregularities. Specific knowledge and instructions on how to comply combined with deterrence inducing a motivation to comply could be the key to the approach's effectiveness. That is, if this mechanism holds, RPS will increase compliance when they are sent out to taxpayers with the instruction to return them to the ATO: a 'proper schedule effect'. Here, we will also explore whether this effect is achieved only with the inclusion of a detailed information booklet (as the ATO practiced the program in the past), or whether a more economical set of instructions is sufficient. The latter would be more strictly comparable with the other experimental conditions (which do not include a booklet) and would make sure that no confounding processes are involved (e.g., taxpayers reciprocating a valuable booklet with being more compliant; see Cialdini, 1997).

The relative effectiveness of a deterring or a cooperative accompanying message could also tell us something about the mechanisms underlying the effects of RPS. For instance, a deterring letter having generally more positive effects on compliance could indicate that deterrence is the main driving process. Conversely, a generally more positive result for a cooperative message could indicate that compliance is based on taxpayers' positive evaluation and perceived legitimacy of the tax authority (Tyler, 1990, 1997).

Further, the tone of the accompanying message could interact with the three mechanisms discussed above and enhance their effects. A cooperative message that emphasizes the tax office's help, support and assistance could encourage taxpayers to make the best use of the

schedules and thus be conducive to an ‘education effect’. In contrast, a deterring message emphasizing the increased possibility of an audit and the certainty of penalties for noncompliance could be conducive to a ‘mere surveillance effect’.

However, while deterrence is the traditional and widely accepted perspective to tax compliance (e.g., Allingham & Sandmo, 1972), the studies by Schwartz and Orleans (1967) and Slemrod et al. (2001) suggest that deterrence measures can sometimes be counterproductive. Indeed, research into psychological reactance (Brehm & Brehm, 1981) has shown that the use of threat and coercion, particularly when perceived as illegitimate, can produce behavior in the opposite direction from that advocated (see also Kirchler, 1999). That means a deterring message to taxpayers that does not allow them to demonstrate their freely-chosen, voluntary compliance with the tax laws and that uses threat on the a-priori assumption that they will be non-compliant, could motivate them to re-establish their freedom by resisting the enforcement attempt. Specifically, this could be the case for taxpayers who have been sent RPS for the second time in a row, who may thus feel unfairly targeted by the tax office and regard the coercive attempt as particularly illegitimate. To test this, the present study included a group of taxpayers who had been subjected to the RPS program in the previous year as well as a group of taxpayers who were subjected to the program for the first time. It has to be noted, however, that the group of taxpayers who had already been sent RPS in the previous year had been selected by the ATO based on a risk assessment and thus differed from first-time participants of the program in other, confounded characteristics.

To summarize, the purpose of the present study is to investigate:

- (1) whether the ATO’s RPS program is effective and works best when schedules are sent out with the instruction that they are to be returned (proper schedule effect), or

- whether it does not require the instruction to return the schedules (education effect), nor even the inclusion of the schedules (mere surveillance effect);
- (2) whether a hard and deterring accompanying message or a soft and cooperative message produces greater levels of compliance;
 - (3) whether the tone of the accompanying letter message interacts with the use (return vs. no return, and inclusion vs. no inclusion) of the schedules;
 - (4) whether the RPS program is equally effective for first-time and second-time recipients of the schedules; and
 - (5) whether the tone of the accompanying letter has different effects for first-time vs. second-time participants of the RPS program.

5. Method

5.1 Participants

Nine thousand taxpayers who, according to ATO records, owned rental property were randomly selected by the ATO for the study. All taxpayers included in the study were paper lodgers, who lodged their tax returns on paper rather than electronically. As a consequence, the majority of cases were self-preparers (76%), because most professional tax preparers use an electronic lodgment facility; the remainder of cases used a tax preparer who lodged a paper tax return.

Samples of 4500 cases each were randomly drawn from two populations of rental property owners: taxpayers who had not been sent RPS before (first-time) and taxpayers who had been sent RPS to complete and return to the ATO in the year before (second-time). First-time participants were drawn from the general population of rental property owners, while second-time participants had been subjected to the RPS program in the previous year because they met

ATO risk criteria. Table 1 details the background characteristics for the two general samples of taxpayers (included as covariates in the analyses below).

5.2 Design

From each participant *group* (first-time vs. second-time), five hundred taxpayers were randomly allocated to one of eight experimental conditions that combined the factors *letter tone* (soft vs. hard) and *letter content* (information only, no-return schedule, return schedule only, return schedule plus booklet). That is, the experimental conditions constituted a 2(group) x 2(letter tone) x 4(letter content) design. In addition, 500 cases of each participant group served as untreated control groups that were not contacted by the ATO at all. For ease of illustration, the design with 16 cells plus two control groups is shown in Table 2.

5.3 Independent variables

Tone of letter. The “soft” letters began with the sentence “At the Australian Taxation Office (ATO) we are committed to helping taxpayers to correctly prepare their income tax returns”. The emphasis in these letters was on the role of the ATO as being informative and helpful. No mention was made in these letters of penalties or audit action. The “hard” letters began with “Over the past few years the ATO has conducted an extensive review program which has enabled us to collect and analyze rental property income and deductions data. The program has resulted in a substantial number of adjustments to rental property claims”. It was emphasized that taxpayers could be selected for audit action, and that penalties for non-compliance could be imposed.

Letter content. The “information only” condition consisted of bringing to the attention of taxpayers the need to complete their tax returns correctly in relation to rental property, and how to obtain further information about tax requirements concerning rental property. No schedule was included within this condition. The “no-return schedule” letter contained RPS with brief

instructions on how to complete the schedules. However, it was emphasized that the schedules were for the benefit of the taxpayer, that they should be kept with their taxation records, and that the schedules were *not* to be returned to the ATO. The “return schedule only” letter also contained RPS and brief instructions on how to complete the enclosed schedules as well as when and where to return them to the ATO. The “return schedule plus booklet” letter contained RPS plus a booklet of instructions on how to complete the schedules as well as a booklet detailing tax issues related to rental property. The letter advised that the RPS had to be returned to the ATO and when and where to do so.

5.4 Procedure

The letters and schedules were mailed out to taxpayers in June 2000, just prior to the end of the 1999/2000 financial year (which, in Australia, is from 1 July to 30 June). For conditions where schedules were required to be returned to the ATO, the letters specified that the schedules were to be returned at the same time as the year 2000 tax return (but to different addresses).

5.5 Data extraction

In January 2001 (approximately two months after the official deadline for lodgment of individual tax returns), the ATO provided the authors with de-identified anonymous data for each taxpayer included in the sample on the following variables: age, sex, time of lodgment (of the tax return in weeks since beginning of the filing period), whether or not a rental schedule had been lodged, the total taxable income reported in the 2000 tax return (excluding rental income), gross rental income for 2000 and 1999 tax return, and rental deductions claimed in 2000 and 1999.

5.6 Data screening

At the time of accessing and analyzing the data, 1435 taxpayers of the 9000 cases selected had not lodged a tax return.² Loglinear analyses showed that the rate of non-lodgment

was not significantly associated with the experimental variables letter tone and letter content. However, there was a significant association with taxpayer group. The non-lodgment rate was significantly greater for first-time participants (18.1%) than for second-time participants (13.8%), $\chi^2(1) = 31.53, p < .001$. Another 719 cases were excluded from analyses because these taxpayers reported having zero net rental income and zero total rental deductions and thus no longer owned rental property in a tax-relevant sense. This left 6,846 cases.

Parametric statistical tests were used to analyze the data; these required that variables were normally distributed. Typical for monetary variables, however, the distributions of income and deductions variables were positively skewed. All monetary variables were therefore subjected to square root transformations, which substantially improved the distributions towards normality. Note that the transformations mean that the values can no longer be interpreted on a dollar scale.

Following the transformations, further cases were excluded on the basis of outlier analyses. First, univariate outliers are cases that have extreme values on single variables and thus could have distorting effects on statistical analyses. Univariate outliers were here defined as cases that differed more than 4 standard deviations from the mean. Monetary variables, which have no natural range limits, are particularly likely to have extreme cases. Because first-time and second-time participants came from different populations of taxpayers, with potentially different value ranges on the tax return variables, this outlier analysis was applied to both groups separately. Altogether 88 cases were excluded from analyses because they had ‘extreme’ values on either total taxable income, gross rental income, previous or current rental deductions (i.e., the monetary variables considered in the analyses reported in this paper).

Second, multivariate outliers are cases that do not fit the statistical model used to estimate the relation between multiple variables. Multivariate outliers were here defined as cases for

which the multivariate analyses produced estimates of the dependent variable that deviated more than 4 standard deviations from the observed value (Cohen & Cohen, 1983). Depending on the specific analysis, 53 cases (for the model involving the three factors of the experimental design) and 60 cases (for the model that included the control groups but dropped letter tone) were excluded from the analyses.

Our definition of outliers follows conventions and recommendations by Cohen and Cohen (1983). It uses relatively relaxed criteria, given our large sample size. In total only about 2% of valid cases were excluded, which is considered as “few” (Cohen & Cohen, 1983, p. 128). Indeed, re-running the data analyses without exclusion of univariate or multivariate outliers produced very similar results. For more accurate and generalizable estimates of the experimental effects, however, we report here analyses that excluded outliers. In the tax context, outliers are likely due to differences or changes in economic circumstances that can have dramatic effects on tax positions. Outliers do not indicate a problem in the design of the study; their exclusion does not represent ignorance to any such problems but is rather a reasoned approach.

6. Results

The main dependent variable investigated in this article was the amount of rental deductions claimed in the 2000 tax return, because the RPS asked taxpayers to detail their rental expenditures that they claimed as deductions in the tax return; the RPS also asked for the total amount of rental income but not for any further details on this issue. Thus, the focus was on deductions and we expected only treatment effects to occur for rental deductions, but not for rental income. In fact, equivalent analyses for rental income, as the ones reported here for rental deductions, did not yield any differences between the various treatments, nor between treatment and control groups.

First, a 2(group) x 2(letter tone) x 4(letter content) analysis of covariance was conducted for (square-root transformed) deduction claims as the dependent variable, controlling for effects of the following covariates: age, sex, time of lodgment, schedule lodgment, total taxable income (excluding rent), gross rental income, and previous rental deductions. The inclusion of the latter covariate meant that differences in previous rental deductions were statistically partialled out, and thus experimental effects could be interpreted as *change* in rental deduction claims. According to Cohen and Cohen (1983), in the behavioral sciences this approach to model change is more appropriate than creating a new dependent variable from difference scores.

Schedule lodgment (i.e., whether a rental schedule had been lodged or not) was included as a covariate because, although two of the experimental conditions and the control condition did not require a schedule to be returned, about 17% of taxpayers in each of these three conditions lodged schedules. One main reason for this was that taxpayers, although selected on the basis that they had lodged a paper return in the previous year, chose to file their tax return electronically. By controlling for this and other background variables, we accounted for a great part of the large heterogeneity in the taxpayer population compared to which cell sizes could be considered rather modest. This way, we increased the statistical power and the sensitivity of the statistical procedures.

The results of the analysis are summarized in Table 3. All of the covariates, except taxpayer gender, were significantly related to rental deduction claims. Previous deduction claims was the strongest predictor ($\beta = .69$); the more deductions had been claimed in the previous year, the more deductions were claimed in the current year. Likewise, gross rental income was positively related to rental deductions ($\beta = .32$), with taxpayers reporting more rental income also claiming greater rental deductions. Similarly, total (non-rental) taxable income was positively related to the amount of rental deductions claimed ($\beta = .07$). Age of taxpayer proved

significantly related to rental deduction claims ($\beta = -.08$), with younger taxpayers claiming greater amounts of deductions. Taxpayers who lodged their tax return earlier and those who lodged a RPS claimed greater deductions ($\beta = -.03$ and $\beta = .03$ respectively) than those who did not; however these latter two effects were very small and for this reason will not be further discussed.

Controlling for the covariates, the analysis yielded significant main effects for group and letter content as well as a significant interaction effect of group and letter tone. First-time participants claimed a greater amount of rental deductions than second-time participants did ($M_s = 89.14$ vs. 80.87), despite the fact that the second-time participants had been defined as risk cases. However, the riskiness of the cases might be controlled by the inclusion of the covariates; that is, in particular, previous rental deduction claims and current rental income. Indeed, without taking covariates into account, first-time participants showed lower deduction claims than second-time participants did ($M_s = 69.42$ vs. 98.76).³

The significant interaction between group and letter tone indicates that for first-time and second-time participants the tone of the letter had different effects. For first-time participants, a hard and deterring letter yielded overall significantly lower deduction claims than a softer letter ($M_s = 88.44$ vs. 89.84 , $p < .05$) while letter tone did not impact significantly on second-time participants ($M_s = 81.34$ vs. 80.40 , $p > .1$). While these results suggest that a threatening letter may deter taxpayers upon first exposure to it, whereas threat has less effect upon more seasoned taxpayers, this interpretation is not clear cut and will be discussed below.

More important for the present study, rental deduction claims differed significantly depending on the experimental treatment taxpayers received. To see how the treatment conditions compared with the control groups on rental deductions, a separate 2 (group) x 5 (intervention) analysis of covariance was conducted (dropping letter tone as this factor was not

applicable to the control groups). The intervention factor included the four letter content conditions and the control conditions. The same set of covariates was included in the analysis as before; their effects were essentially the same as in the earlier analysis and need not be discussed again (see Table 4). The analysis again yielded significant main effects for group and intervention, but no interaction, indicating that the effects of intervention were similar for both groups. Pairwise comparisons showed that rental deduction claims in the no-return schedule condition and the information only condition did not differ from the control group (see Table 5). In contrast, taxpayers in each of the two return schedule conditions claimed significantly less deductions than taxpayers in the control condition. Further, although not reaching conventional levels of significance, rental deductions were marginally lower in the return schedule only condition compared with the return schedule and booklet condition, $p = .053$. These findings suggest that the demand to return schedules to the ATO was crucial for the intervention to have an effect, while the addition of an information booklet produced no further advantages.

7. Discussion

The present study applied a controlled experimental design to investigate the effects of tax-reporting schedules as used by (Australian) tax authorities to proactively encourage taxpayers to comply with tax rules and be truthful in deduction claims for expenditures. The study clearly showed that the schedules can be an effective intervention, but only when they are used for taxpayers to account towards the tax office for their deduction claims. The schedule did not show any effect as an educational tool when sent to taxpayers for their personal use only. Importantly, the schedule increased compliance not simply because it warned taxpayers generally of a greater probability of audit. A mere warning letter did not achieve any reduction in deduction claims compared to control groups. It seems then that tax-reporting schedules need to

achieve both clear instructions and education about what the tax office expects taxpayers to do *as well as* creating the impression that the tax authorities scrutinize taxpayers' statements and penalize non-compliance. Good news for tax authorities is that the approach does not need to involve detailed and expensive booklets; the schedules worked just as well when accompanied by brief instructions. Further good news is that, according to estimates based on analyses with untransformed monetary variables, schedules in our study yielded a reduction of predicted rental deduction claims by 5% to 7.5% (for conditions with and without booklet, respectively). This is not an inconsequential saving to the tax office.

The effects of the schedules were not further moderated by the tone of the letter, contrary to what was predicted for a mere education or a mere surveillance process. Indeed, it could be argued that schedules operating through a combination of education and implied deterrence require both respectful encouragement to cooperate and threat of punishment when cooperation is lacking (Ayres & Braithwaite, 1992; Braithwaite & Braithwaite, 2000). One limitation of this research is that we had not pretested the letters to ensure that their messages were perceived as intended. Thus, it could be argued that the letter tone failed to moderate the effectiveness of the different letters simply because the desired tones were not achieved. However, the fact that letter tone did have a significant impact for first-time participants suggests that different letter tones were indeed successfully established. Nonetheless, future research should determine more carefully prior to their experimental application how taxpayers perceive and process certain letters from the tax authority.

While our findings suggest that rental schedules are effective tools for tax administrators to increase compliance, it could be argued that these effects were only due to schedules signaling to taxpayers that they were likely to be audited. That is, it could be that the schedules are simply a tool for signaling potential audit and have no additional value beyond that. However, this

interpretation is not convincing for three reasons. First, it presumes a very sophisticated knowledge and understanding by taxpayers about the use of schedules, the targeting of risk groups and the wider audit strategy of the tax office. In fact, the schedules had thus far only been used on a relatively small scale. Second, such taxpayer knowledge is unlikely to be equally distributed; we would expect taxpayers who have first-hand experience with rental schedules to know (or speculate) more about their use. Nonetheless, the schedules proved equally effective for first and second-time recipients. Third, if this argument were valid, one would expect a mere warning letter to also signal potential audit. However, the mere warning letter had no impact. We contend therefore that the effects of the schedules are distinct from those of mere audit threats.

Despite its limitations, we think the present study very much demonstrates the virtues of the experimental method in evaluating strategies and concepts of tax administration and regulation more generally (Wenzel & Taylor, in press). Randomization, experimental manipulation, and the use of control groups allow us to unambiguously attribute observed differences to the respective treatments. The study used actual tax return data that constitutes the ultimate outcome variables tax administrators are interested in; their validity is immediate and unquestionable. In this respect, the field-experimental approach is superior to focus groups, interviews, or customer satisfaction questionnaires that are widely used by tax administrators as well as in other fields of regulation. Likewise, the tested interventions are neither simulated (as in laboratory experiments) nor approximated, but are the ones the tax authority uses or plans to use. The result is external validity at the highest level.

Although the present study could not answer all the questions involved in the use of tax-reporting schedules as a proactive means to increase tax compliance, it was consistently found that RPS as used by the ATO were effective in reducing rental deduction claims and thus, presumably, increasing tax compliance. It needs to be noted, however, that the present research

focused on self-preparers only; it is unclear whether schedules would be equally effective with taxpayers who use professional tax preparers.⁴ On the one hand, tax professionals might be less susceptible to such communications from the tax office because they have greater routine and are less easily impressed or intimidated. On the other hand, they might pass on the function of the schedules by using them to make their clients more accountable for their claims. This is very much a question of the general role of tax professionals (e.g., Klepper & Nagin, 1989) and requires further research.

We hope the present study encourages tax administrations and other regulating bodies to adopt field-experimental evaluations as an important tool for evidence-based decision-making. Not only can the experimental approach directly evaluate the usefulness and effectiveness of certain interventions and strategies but, if thoughtfully designed and used systematically and cumulatively, it can also contribute to our theoretical understanding of compliance and regulation. That understanding can then be extrapolated – not in an empiricist manner, but at the level of theoretical analysis – to a wealth of other regulatory questions and interventions. The derived predictions can then be experimentally evaluated again, feeding into further and sound theoretical advancement.

Acknowledgments

This study was conducted as part of a research project funded by the Australian Taxation Office.

Thanks go to Tony Goddard, Geoff Whyte, Daniel Heavey, Leanne Bleakley, Fiona Zani, and Daryll D'Mello from the Australian Taxation Office for their collaboration.

References

- Allingham, M., Sandmo, A. 1972. Income tax evasion: A theoretical analysis. *Journal of Public Economics* 1, 323-338.
- Ayres, I., Braithwaite, J. 1992. *Responsive Regulation: Transcending the Deregulation Debate*. Oxford University Press, Oxford.
- Blumenthal, M., Christian, C., Slemrod, J. 2001. Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota. *National Tax Journal* 54, 125-38.
- Boruch, R. F. 1989. Experimental and quasi-experimental designs in taxpayer compliance research. In: Roth, J. A., Scholz, J. T., Witte, A. D. (Eds.), *Taxpayer Compliance, Vol. 1: An Agenda for Research*. University of Pennsylvania Press, Philadelphia, pp. 339-379.
- Braithwaite, V., Braithwaite, J. 2000. An evolving compliance model for tax enforcement. In: Shover N., Wright, J. P. (Eds.), *Crimes of Privilege*. Oxford University Press, Oxford, pp. 405-418.
- Brehm, S. S., Brehm, J. W. 1981. *Psychological Reactance: A Theory of Freedom and Control*. Academic Press, New York.
- Cialdini, R. B. 1997. Interpersonal influence. In: Shavitt S., Brock, T. C. (Eds.), *Persuasion: Psychological Insights and Perspectives*. Allyn & Bacon, Boston, pp. 195-217.
- Cohen, J., & Cohen, P. 1983. *Applied Multiple Regression/Correlation Analysis for the Behavioural Sciences* (2 ed.). Erlbaum, Hillsdale, NJ.
- Coleman, S. 1997. Income tax compliance: A unique experiment in Minnesota. *Government Finance Review*, 13, 11-15.
- Judd, C. M., Smith, E. R., Kidder, L. H. 1991. *Research Methods in Social Relations*, sixth ed. Harcourt Brace Jovanovich, Fort Worth.

- Kirchler, E. 1999. Reactance to taxation: Employers' attitudes towards taxes. *Journal of Socio-Economics*, 28, 131-138.
- Klepper, S., & Nagin, D. S. 1989. The role of tax practitioners in tax compliance. *Policy Sciences*, 22, 167-192.
- McGraw, K. M., Scholz, J. T. 1991. Appeals to civic virtue versus attention to self-interest: Effects on tax compliance. *Law and Society Review*, 25, 471-498.
- Roth, J. A., Scholz, J. T., Witte, A. D. 1989. *Taxpayer Compliance, Vol. 1: An Agenda for Research*. University of Pennsylvania Press, Philadelphia.
- Schwartz, R. D., Orleans, S. 1967. On legal sanctions. *University of Chicago Law Review*, 34, 274-300.
- Slemrod, J., Blumenthal, M., Christian, C. 2001. Taxpayer response to an increased probability of audit: Evidence from a controlled experiment in Minnesota. *Journal of Public Economics*, 79, 455-83.
- Tyler, T. R. 1990. *Why People Obey the Law: Procedural Justice, Legitimacy, and Compliance*. Yale University Press, New Haven.
- Tyler, T. R. 1997. The psychology of legitimacy: A relational perspective on voluntary deference to authorities. *Personality and Social Psychology Review*, 1, 323-345.
- Wenzel, M., & Taylor, N. In press. Toward evidence-based tax administration. *Australian Journal of Social Issues*.

Table 1. Means and Standard Deviations of Background Characteristics for the Two Samples, First-Time and Second-Time Participants.

	First-time participants		Second-time participants	
	(n = 4500)		(n = 4500)	
	M	(SD)	M	(SD)
Sex (0 = male, 1 = female)	0.48	(0.50)	0.39	(0.49)
Age	45.52	(11.81)	47.50	(11.07)
Lodgment of tax return (0 = no, 1 = yes)	0.82	(0.39)	0.86	(0.34)
Lodgment week	14.52	(6.82)	15.13	(6.66)
Lodgment of RPS (0 =no, 1 = yes)	0.34	(0.47)	0.36	(0.48)
Taxable (non-rental) income	38229.83	(40456.66)	54751.92	(61361.27)
Gross rental income	5315.23	(5582.64)	12656.64	(14426.54)
Previous rental deductions	5596.61	(7804.26)	12400.48	(16386.96)

Note. Means for dichotomous variables translate directly into percentages; for example, a mean for sex of 0.48 indicates 52% males and 48% females. First-time and second-time participants differed significantly at $p < .001$ on all attributes except for lodgment of RPS.

Table 2. Experimental Design.

		Letter content				Control
		Information only	No-return schedule	Return schedule	Schedule + booklet	
Group	Tone					
First-time	soft	(<i>n</i> = 500)	(<i>n</i> = 500)	(<i>n</i> = 500)	(<i>n</i> = 500)	
	hard	(<i>n</i> = 500)	(<i>n</i> = 500)	(<i>n</i> = 500)	(<i>n</i> = 500)	(<i>n</i> = 500)
Second-time	soft	(<i>n</i> = 500)	(<i>n</i> = 500)	(<i>n</i> = 500)	(<i>n</i> = 500)	
	hard	(<i>n</i> = 500)	(<i>n</i> = 500)	(<i>n</i> = 500)	(<i>n</i> = 500)	(<i>n</i> = 500)

Table 3. Analysis of Covariance of Rental Deduction Claims.

Source	β	df	F	p
Covariates				
Age	-.08	1	263.53	.000
Gender	-.01	1	1.78	.182
Lodgment week	-.03	1	44.01	.000
Schedule lodgment	.03	1	19.25	.000
Taxable (non-rental) income	.07	1	156.17	.000
Gross rental income	.32	1	2076.61	.000
Previous rental deductions	.69	1	9036.81	.000
Group		1	279.12	.000
Letter tone		1	.29	.587
Letter content		3	9.15	.000
Group x Letter tone		1	7.51	.006
Group x Letter content		3	.60	.612
Letter tone x Letter content		3	2.31	.074
Group x Tone x Content		3	1.28	.278
Error		5924		

Table 4. Analysis of Covariance of Rental Deduction Claims, including Control Groups.

Source	β	df	F	p
Covariates				
Age	-.08	1	270.57	.000
Gender	-.01	1	1.69	.193
Lodgment week	-.04	1	52.25	.000
Schedule lodgment	.03	1	27.95	.000
Taxable (non-rental) income	.07	1	178.69	.000
Gross rental income	.33	1	2307.89	.000
Previous rental deductions	.69	1	9910.20	.000
Group		1	278.88	.000
Intervention		1	8.41	.000
Group x Intervention		1	.54	.704
Error		6681		

Table 5. Estimated Marginal Means of Rental Deductions for the ANCOVA Model including all Treatment and Control Conditions: Effect of Treatment.

Letter Content	<i>M</i>	<i>SE</i>	<i>n</i>
Information only	86.44 ^a	.46	1496
No-return schedule	86.27 ^a	.46	1479
Return schedule only	82.85 ^b	.47	1477
Return schedule plus booklet	84.03 ^b	.47	1497
Control	86.21 ^a	.63	749

Note. Means with different superscripts differ significantly at $p < .05$.

Footnotes

¹ The definition of risk taxpayers subjected to the RPS program was based on ATO processes and insights and may not be intuitively understood. It included cases with any of the following characteristics: (1) zero gross rental income, or (2) gross rental income of more than A\$35,000, or (3) a combination of net rental income of less than A\$8,000 and taxable income of more than A\$5,400, or (4) a combination of gross rental income between A\$5,000 and A\$35,000, and taxable income of more than \$20,700.

² It is possible that taxpayers who had not lodged an individual tax return by this time had switched to having their return professionally prepared. Returns prepared by tax professionals can be lodged as late as May of the following year.

³ The main effect seems to suggest that the RPS programs of the previous years, to which the second-time participants had been subjected, continued to have an impact on deduction claims, becoming apparent only when their higher risk-level is statistically accounted for. However, the large size of the effect (relative to the concurrent treatment effects) suggests otherwise. In fact, the result may also be the outcome of regression to the mean. Note that the second-time group had been sampled as a risk group based on their earlier tax details. The probability that, over time, these cases become more mainstream is greater than for them becoming even more extreme (regression to the mean).

⁴ In the Australian context, this is a rather academic question because the vast majority of tax professionals file tax returns electronically, where they are required to enter the same details as requested in the schedules anyhow.