

Stephan Muehlbacher/Erich Kirchler

Taxperiments

About the external validity of laboratory experiments in tax compliance research



Stephan
Muehlbacher



Erich Kirchler

Experimente in der Steuerforschung

Eine kritische Diskussion der experimentellen Methodik zur Untersuchung des Steuerverhaltens

■ Schlüsselbegriffe

Externe Validität; Laborexperimente; Methoden; Steuerhinterziehung; Steuerverhalten

■ Keywords

External validity; laboratory experiments; methods; tax compliance; tax evasion

Zusammenfassung

Der Einsatz von Laborexperimenten in der Steuerforschung wird häufig kritisiert, da die hohe interne Validität auch eine geringe externe Validität mit sich bringt. Ein Teil der Einwendungen betrifft die experimentelle Methodik als Ganzes, andere beziehen sich auf spezifische Aspekte der Erforschung des Steuerzahlverhaltens im Labor. Im vorliegenden Beitrag wird diese Kritik zusammengefasst und soweit vorhanden werden empirische Prüfungen der Vorwürfe vorgestellt. Aus der bestehenden Literatur können zwar erste vorsichtige Empfehlungen für die experimentelle Praxis abgeleitet werden, im Gesamten wird aber ein dringender Bedarf an systematischer Forschung zur externen Validität von Steuerexperimenten festgestellt.

Abstract

Laboratory experiments are commonly applied in research on tax compliance. Their use is frequently criticized because the price for their high internal validity is low external validity. Whereas some of the concerns about this method apply to experimental methodology in general, others are specifically related to its application in tax compliance research. We provide a summary of the critique and a review of empirical research that has addressed these methodological issues. Although some cautious recommendations for experimental practice can be gleaned from existing studies, there is a strong need for more systematic research on the external validity of studying tax compliance in the lab.

Autoren

Dr. Stephan Muehlbacher (corresponding author), Faculty of Psychology, University of Vienna, Universitaetsstr. 7, A-1010 Vienna, Austria, E-Mail: stephan.muehlbacher@univie.ac.at; Prof. Dr. Erich Kirchler, Faculty of Psychology, University of Vienna.

Acknowledgments: The authors would like to thank two anonymous reviewers for their valuable comments on an earlier version of the manuscript.

1. Introduction

Since Friedland, Maital, and Rutenberg (1978) published their pioneering work, laboratory experiments have become widespread in tax research. Experiments are a relatively cost-efficient way to study taxpayers' behavior in situations that cannot be easily manipulated in the »real« world. The experimental method also allows researchers to control for confounding variables that are likely to affect behavior along with the variables of interest. But the high internal validity of the experimental method comes at the price of rather low external validity, which limits the generalizability of findings. Accordingly, the use of the experimental approach in studying taxpayers' behavior has frequently been criticized by various authors. The goal of the present paper is to summarize this critique and to review empirical work on the methodological issues in experimental tax research.

2. Comparisons across research methods

Due to the clandestine nature of illegal behavior such as tax evasion, it is not easy to gather valid data on the phenomenon. Besides laboratory experiments, the methods that have been applied to the issue of taxation are as diverse as the scientific disciplines that have addressed it. Field experiments (e.g., Slemrod/Blumenthal/Christian, 2001; Torgler, 2013) may be considered the method of choice, but because such studies are costly and intrusive, they are not often conducted. Natural quasi-experiments (e.g., Torgler, 2003) are feasible only if comparable groups that differ on the variable of interest exist in the field. Archival data – for instance, from tax audits or tax authorities' measurement initiatives

(e.g., Ali/Cecil/Knoblett, 2001; Clotfelter, 1983) – are hardly accessible. Such data are often available only in an aggregate form for groups of taxpayers rather than for individuals. Further, archival data typically cover a relatively small set of variables and are therefore useful for only a limited set of research questions. Self-reports on compliance behavior can easily be collected in interviews or surveys (e.g., Gërxhani, 2007; Muehlbacher/Kirchler, 2013) but are prone to socially desirable responses. A similar problem may confound findings from studies that are designed to elicit behavioral intentions in hypothetical evasion scenarios without financial incentives (e.g., Henderson/Kaplan, 2005; Muehlbacher/Kirchler/Hoelzl/Ashby/Berti/Job/Kemp/Peterlik/Roland-Levy/Waldherr, 2008). All methods in tax compliance research have their advantages and disadvantages, and whether one is superior to the others largely depends on the research question to be answered.

Systematic research comparing the various methods has been scarce (e.g., Elffers/Robben/Hessing, 1992; Elffers/Weigel/Hessing, 1987; Hite, 1988), and only a few studies have compared observations from laboratory experiments to findings obtained by other methods. A notable exception is an extensive study by Elffers/Robben/Hessing (1992). The authors received assessments of income tax returns from the Dutch Tax Service done by tax inspectors during their regular work. To ensure the accuracy of the compliance status of each tax return, a second tax officer re-assessed the files, and three expert tax officers discussed potential discrepancies between the two assessments until a consensus was reached. The taxpayers whose tax returns were re-assessed were contacted by a notary who guaranteed anonymity and confidentiality. They were invited to complete a survey and to indicate whether they evaded taxes in the year of

	Tax officer classification	Self-reports	Behavior in experiment
Tax officer classification	–		
Self-reports	.10	–	
Behavior in experiment	–.02	.02	–

Note. n = 120 except for the correlation between tax officer classification and self-reports, which includes a second subsample that did not participate in the experiment for a total of n = 209.

Tab. 1: Descriptive statistics and correlations from Elffers/Robben/Hessing, 1992

the re-assessed tax return. Half of the participants were also asked to participate in a tax evasion experiment. This procedure allowed for comparisons to be made between the official classification, the self-reported behavior on the survey, and the compliance behavior in the tax experiment. Table 1 shows correlations between the three compliance measures. All correlations between assessed, observed, and self-reported tax behavior were non-significant and extremely low. Compliance behavior as assessed by the tax officers was not associated with self-reported behavior on the survey or with participants' behavior in the experiment.

However, not all comparisons of methodological approaches are so disillusioning. Hite (1988) found a low but positive correlation between self-reported compliance, compliance decisions in hypothetical scenarios, and government-reported statistics. Regarding the experimental approach, Alm/Bloomquist/McKee (2015) compared the behavior of subjects in the laboratory with data from »real« taxpayers obtained by the National Research Program of North America's Internal Revenue Service. By restricting the field data to taxpayers whose taxable income was not reported by a third party, the two data sets were guaranteed to share an important feature of the compliance decision: In both situations – the lab and the real world – the tax liabilities were self-reported. On average, participants in the experiments were slightly more honest in their income declarations than the real taxpayers. Interestingly, in both samples, the distribution of compliance rates was bimodal with peaks at the upper and lower ends. This means that whereas the absolute compliance levels observed in experiments should not be generalized to the real world, evasion strategies in the laboratory seem to be similar to those in a real-world setting: The majority of taxpayers tend to be either completely honest or to evade the tax due as a whole, with a random distribution of evasion falling between the two extremes.

Cummings and colleagues (Cummings/Martinez-Vazquez/McKee/Torgler, 2009) collected survey data from the so-called Afrobarometer and conducted lab experiments to compare tax compliance between Botswana and South Africa. Due to the diverse political history of these countries, the citizens' respect for state and governmental institutions differs strongly, a difference that should also be reflected in the taxpaying culture. Analyses

of self-reported evasion collected in the Afrobarometer and results from identical laboratory experiments conducted in both countries showed that compliance in Botswana was substantially higher than in South Africa. Hence, asking for self-reports on the survey yielded results that were similar to those from the experiments. Both methods seemed to be able to capture the social norms that prevailed in the different cultures.

A review of empirical studies on the four parameters (probability of an audit, fine, income, and marginal tax rate) from the economic model on income tax evasion allows for a further comparison of methods in tax compliance research (Kirchler/Muehlbacher/Kastlunger/Wahl, 2010).¹ Table 2 shows a summary of findings achieved by various research methods. The clearest effects on compliance were observed for audit probabilities and the tax rate. Experimental studies on these parameters yielded results that were similar to those from analyses of aggregate data on »real« taxpayers. Fine levels had either the intended effect or no effect at all, regardless of the method that was applied. Findings on income effects are as ambiguous as the outcomes that are predicted by the theoretical model (Allingham/Sandmo, 1972). Independent of the research method, a high income level was found to have a positive, negative, or no effect on compliance. Hence, the different methods applied in the empirical studies that were reviewed converged with respect to the direction of the effects that were observed.

To summarize, the tax literature lacks systematic research on its methodology that would allow an evidence-based assessment of the external validity of tax experiments. The few exceptions that have directly compared different methodological approaches have produced puzzling results, ranging from zero correlations between participants' behavior in tax experiments and their actual behavior in the real world to concordant assessments of variables that have been hypothesized to affect compliance.

Many ideas have been discussed about what the limits for generalizing experimental results are and how to increase external validity. We review this critique in the following sections.

¹ See Blackwell (2010) for a meta-analysis of the effects of the standard economic model's parameters observed in lab experiments.

Effect of income on compliance				
Method	Nr. of studies	Negative	Zero	Positive
Aggregate data	9	4 (44 %)	1 (11 %)	4 (44 %)
Experiment	4	2 (50 %)	1 (25 %)	1 (25 %)
Self-reports	6	2 (33 %)	3 (50 %)	1 (17 %)
Effect of tax rate on compliance				
		Negative	Zero	Positive
Aggregate data	7	6 (86 %)	0 (0 %)	1 (14 %)
Experiment	7	5 (71 %)	1 (14 %)	1 (14 %)
Self-reports	2	1 (50 %)	1 (50 %)	0 (0 %)
Effect of audit probability on compliance				
		Negative	Zero	Positive
Aggregate data	5	0 (0 %)	0 (0 %)	5 (100 %)
Experiment	11	0 (0 %)	1 (9 %)	10 (91 %)
Self-reports	3	0 (0 %)	2 (67 %)	1 (33 %)
Effect of fines on compliance				
		Negative	Zero	Positive
Aggregate data	3	0 (0 %)	2 (67 %)	1 (33 %)
Experiment	8	0 (0 %)	4 (50 %)	4 (50 %)
Self-reports	1	0 (0 %)	1 (100 %)	0 (0 %)
Note. Frequencies indicate how often a positive, negative, or no effect of the particular factor was observed with the various methods.				

Table 2: Summary of empirical findings obtained with different methods regarding the effects of income, tax rate, audits, and fines on compliance (frequencies from Kirchler/Muehlbacher/Kastlunger/Wahl, 2010)

3. Critique on tax experiments

Various authors have raised several issues that call into question the external validity of laboratory experiments in tax compliance research. In the following, the most frequent critique will be summarized and classified into arguments that have not yet been addressed empirically and objections that have been evaluated by at least rudimentary empirical research.

3.1. Potential weaknesses without systematic research

The most general critique addresses the artificiality of the laboratory setting (e.g., Elffers/Weigel/Hessing, 1987; Torgler, 2002). Experiments are designed to mimic the most important aspects of the rich con-

text in which tax compliance decisions are embedded in real life. Naturally, the complexity of reality must be reduced to maximize an experiment's internal validity. The argument about a level of realism that is too low may be countered by pointing out that the artificiality of the experimental environment should not matter as long as the experiment includes the important variables that induce the same psychological mechanisms that are produced in the real world. Webley and his colleagues (Webley/Robben/Elffers/Hessing, 1991) provided the analogy of a painter who emphasizes a distinct aspect of reality by neglecting or blurring other details that the artist considers to be less important. By reducing the less important aspects in the picture, the more important things become apparent. However, it is still necessary to identify the essential aspects of reality that should be included in the experimental environment in order to induce realistic behavior.

The issue of artificiality also concerns the generalizability of the experimental results. If conclusions for the whole population of taxpayers need to be made, the researcher must determine *whose* reality needs to be modeled in the lab. For instance, the task of self-declaring income in a typical tax experiment is very similar to the situation faced by self-employed taxpayers but seems completely unrealistic for employees whose income is reported to authorities by a third party. Further, in tax experiments, a flat tax system seems to be standard, whereas many countries apply a progressive tax scheme. Progressive tax rates increase the complexity of the system, may create fairness issues, and leave self-employed taxpayers uncertain about how much tax is actually due until the end of the fiscal year. The total annual income and therefore the final tax rate can be determined only after the last earnings are made and the final costs are deducted. Another challenge for generalizability is that in the laboratory, it is common to provide objective audit probabilities, but in reality, such information is vague, authorities adjust the probability by behavior, and taxpayers have to rely on subjective probabilities in their decisions (Kirchler/Muehlbacher/Kastlunger/Wahl, 2010; Torgler, 2002).

Alm (2012) pointed out that participating in experiments is voluntary, which might introduce a bias in sample selection. A related aspect concerns remuneration for participating in the study. Generally, disciplines seem to differ in their views on financial incentives in experimentation (for a discussion, see Hertwig/Ortmann, 2001). Economists frequently emphasize the importance of incentivizing decisions in experiments. Incentives should be linked to behavior and must be large enough to compensate for the time and effort of participating in the study (e. g., Alm, 1991). They provide a clear goal for participants' performance, and their value does not satiate over the course of an experiment. In a review of the experimental literature on judgment and decision making, less variability in the data and decisions that were closer to the theoretical predictions were found when the choices in the experiments were incentivized (Hertwig/Ortmann, 2001). Psychologists, on the other hand, often voice the criticism that monetary incentives would crowd out participants' intrinsic motivation (for a review, see Deci/Koestner/Ryan, 1999). This could also be true for the motivations behind tax-paying behavior in the lab. If remuneration for

participating in tax experiments depends on participants' behavior in the experiment, taxes might be evaded solely due to the extrinsic motive to earn money. Whereas cheating in the lab is somewhat artificial (and therefore the psychological costs of evasion are low), the money paid for such behavior is real. On the other hand, tax honesty observed in the absence of monetary incentives may be considered hypocritical because evading taxes would yield only a fictitious gain. These considerations about the pros and cons of monetary incentives should be kept in mind when interpreting and generalizing experimental data. Motives in the real world for (not) paying taxes differ widely (Braithwaite, 2003; Kirchler/Hoelzl/Wahl, 2008), and whether an effect that is observed under a particular motivation holds for the general population seems questionable.

An important question seems to be what particular behavior should be observed when speaking of tax compliance. A point that is sometimes brought up by practitioners is that in real life, the evasion of value added tax is the biggest problem. Most studies are concerned with income tax evasion instead. It is unclear whether income tax evasion is driven by the same factors that drive evasion of the VAT (Webley/Ashby, 2010). Further, Elffers and colleagues (Elffers/Weigel/Hessing, 1987) remarked that taxpayers in the real world evade in many ways. To reduce the tax burden, they underreport their income, exaggerate deductions, or avoid taxes by exploiting loopholes in the law. Typically, (not only experimental) studies focus on only one of these behaviors, but most of them speak of »tax compliance.«

Webley and colleagues (Webley/Robben/Elffers/Hessing, 1991) brought up the issues of experimenter effects and social desirability in experiments. Participants might tend to show the desired behavior when they know they are being observed – a concern that is definitely true in all areas of experimental research. For instance, generosity in the dictator game decreases with the social distance between the participants and the experimenter (Hoffman/McCabe/Smith, 1996). Considering the interdisciplinary nature of tax research, it would be particularly problematic if the behavior that participants consider desirable depends on how they perceive the experimenter. Students in economics, for instance, act more selfishly in public goods experiments than students from other disciplines

(Marwell/Ames, 1981). This might be due to training or to self-selection effects. However, the observation could also be due to an experimenter effect. If the experimenter is an economist, participants might tend to show their knowledge about finding the right strategy for maximizing profits; if the experimenter is from psychology, participants might prefer to present themselves as advocates for the various theories about justice. Although this exemplification was hypothetical only, in tax research, such differential experimenter effects would mean that participants' behavior in the laboratory will depend on their training, the variables being manipulated, and the scientific discipline of the research team conducting the study. Clearly, systematic research on experimenter effects in tax experiments is needed.

Surely there are further examples of the aspects of reality that are usually neglected when studying tax behavior in the lab. For instance, Torgler (2002) discussed the absence of severe sanctions such as jail or social blaming in the lab, the role of democratic participation through voice or voting, and the problem of capturing the dynamic process of paying taxes in an experiment that lasts for only a few rounds. More research and a consensus on appropriate designs in the scientific community are definitely necessary. For now, researchers should at least keep in mind the many pitfalls of experimental tax research when interpreting their observations. Luckily, some of the critiques on tax experiments have been addressed in empirical studies; these will be reviewed in the next section.

3.2. Potential weaknesses previously addressed by empirical research

Relatively few studies have systematically addressed the specific methodological issues of experimentation in tax research. They have addressed the criticism of the gambling-like character of tax experiments and that paying taxes in the lab serves no purpose. Some studies have compared different subject pools because – as in other fields – experimental tax research is typically based on student samples that lack experience in taxpaying. Further concerns that are based on empirical observations are related to the impact of receiving a windfall income in experiments and the temporal aspects of the compliance decision.

3.2.1. Gambling-like character of tax experiments

Several authors have criticized the gambling-like character of tax experiments and have pointed out that important determinants of honesty (e. g., moral considerations) could play a subordinate role in the laboratory (e. g., Elffers/Weigel/Hessing, 1987; Kirchner/Muehlbacher/Kastlunger/Wahl, 2010). What if participants in the laboratory act as if they were in a casino? Another concern was reported by Baldry (personal communication, as cited in Webley/Robben/Elffers/Hessing, 1991). He pointed out that if participants perceive the experiment as a game, they might engage in competition against each other instead of »playing« against authorities. To avoid giving participants the impression that they are participating in a game, Torgler (2002) advises that participants be instructed to complete a tax report rather than to maximize their income.

Several empirical studies have compared decisions embedded in a tax-related context with choices in a tax-neutral context or in abstract lotteries. Whereas Alm and colleagues (Alm/McClelland/Schulze, 1992) observed similar compliance rates regardless of whether »tax language« (taxes, audit, reported income, penalty) or »neutral language« (payment, check, disclosed money, shortfall) was used in the instructions, Baldry (1986) reported higher compliance when choices were made in a tax context than in a neutral setting. Also Choo/Fonseca/Myles (2014) found higher compliance when experiments were framed as involving taxes than when they were presented in a neutral context. Cadsby/Maynes/Trivedi (2006) manipulated whether the experimental instructions emphasized that correct payments were expected from participants or whether participants were invited to decide if they wanted to gamble. Compliance was considerably higher in the former than in the latter condition. Further, emphasizing the expectation of honesty ruled out the impact of audits and fines on compliance. Durham and colleagues (Durham/Manly/Ritsema, 2014) found a complex three-way interaction between the source of income (endowed vs. earned), level of income, and whether payments were framed as taxes or as a market fee that had to be paid. Whereas higher income was related to higher compliance in all other treatments, the opposite was true in the tax-framed condition when income had to be earned.

Mittone (2006) observed similar choices for reporting decisions in a tax setting and in abstract lotteries with the same expected values. However, in the first 16 periods out of a total of 60 periods, risk taking was more pronounced in lottery choices than in tax-compliance decisions. He attributed this finding to the fact that the whole sample had experienced at least one audit (or had lost one of the lotteries) at this point in time. Receiving the feedback that one had lost after gambling seemed to affect only choices in the lottery context but not in the tax-compliance setting.

A further study on the effect of context on risk taking was conducted by Wartick/Madeo/Vines (1999). Participants were told that the purpose of the research was either to learn about economic decision making or to learn about tax-reporting decisions. In one experimental condition, the task for the subjects was to allocate income between a risk-free account on which a 30% commission had to be paid and a risky account for which a die decided whether no or a 60% commission was due. In the other experimental condition, the same task was framed as a tax-reporting decision with the risk of an audit. Interestingly, the effect of context depended on the age of the participants. Whereas participants under the age of 25 made their decisions regardless of context, older participants were more risk-averse (i.e., compliant) in the tax context. The authors argued that the contextual cues about taxpaying could have triggered role playing and that due to their experience, the older participants were more likely to access mental scripts for tax behavior.

It seems that using tax-related terminology in experiments affects participants' behavior, and the task of reporting taxes in the lab is not simply perceived as a gamble. However, little is known about potential interactions with personal characteristics such as experience in paying taxes and with other parameters of the experiment such as audit rates and fines. Also unclear is the number of cues that are necessary to evoke realistic behavior while maximizing control over potentially triggered mental scripts.

3.2.2. Tax money in the lab is wasted

An important difference between the real world and the laboratory lies in issues of exchange fair-

ness. In reality, paying taxes may be perceived as an investment in public goods, whereas in a typical tax experiment, tax payments are meaningless because they are »lost« to the experimenter (e.g., Alm/Jackson/McKee, 1992; Alm/McClelland/Schulze, 1992; Kirchler/Muehlbacher/Kastlunger/Wahl, 2010; Mittone, 2006). Several studies have implemented a mechanism by which tax payments are redistributed to the participants. By doing so, the task of paying taxes changes from an individual decision under risk into a decision about cooperation in a social dilemma situation (Dawes, 1980).

In an experiment by Mittone (2006), the tax money that was collected was redistributed equally to each participant regardless of how much he or she had contributed. Across all 60 periods, 72% of the tax reports were identified as completely honest in the condition with redistribution, whereas the rate was 47% in a control condition. In similar experiments, the sum of the tax payments was multiplied by some factor before it was redistributed to the group of participants in the session. Doubling the collected tax payments increased the average compliance rates from 44% to 54%, and multiplying the sum by a factor of 6 further increased compliance to 59% (Alm/Jackson/McKee, 1992; Alm/McClelland/Schulze, 1992). In another experiment by Alm and colleagues (Alm/Sanchez/de Juan, 1995), the provision of tax-financed public goods had no effect.

A crucial aspect of paying taxes seems to lie in the use of the tax money. Payments in a typical tax experiment are meaningless, and participants might have the perception that their money is »wasted.« Introducing a redistribution mechanism to the experimental environment strongly affects compliance and might bring into play further determinants of tax behavior. If participants benefit from what others contribute to the public good, their behavior is likely to be affected by social norms (Wenzel, 2004), their trust in other taxpayers' willingness to pay, and the power of authorities to protect the community from non-cooperators (Alm/Kirchler/Muehlbacher, 2012). Further, although the role of group size for the provision of public goods has long been acknowledged in social dilemma research (Kopelman/Weber/Messick, 2002), the previously studied groups were rather small (5–8 participants), and tax compliance in larger groups has not been studied yet.

3.2.3. Student samples

As in all other disciplines, students seem to be overrepresented as participants of tax experiments. The practice of using students has often been called into question, particularly in the study of tax behavior (Alm, 2012; Torgler, 2002; Webley/Robben/Elffers/Hessing, 1991). Students are younger than the average taxpayer, typically have a higher IQ, and come from wealthier families (Torgler, 2007). In general, students are a relatively homogenous group, and it is unclear whether their behavior is generalizable to the whole population (Hite, 1988). Further, the role of academic training in students' behavior in experiments is unclear. The participants' area of study could make a difference. Also frequently mentioned is students' lack of experience in paying taxes. With experience, taxpayers may internalize the social norms of their occupational group (Ashby/Webley/Haslam, 2009; Wenzel, 2004), and their perceptions of tax obligations and of authorities seem to change (Kirchler, 1999). Whether lack of experience with tax issues is a problem for experimentation largely depends on the research question. For some issues, naïve subjects might also be advantageous.

An empirical study by Wartick and colleagues (Wartick/Madeo/Vines, 1999) that allowed for a comparison to be made between students' behavior in tax experiments and that of employees was already mentioned in the section about context effects. Participants in this experiment were recruited from among students and university staff members and thus were quite diverse in their experience with taxes. Older and presumably more experienced participants made riskier choices when the experimental task was framed as an economic investment decision than when it was described as a tax-reporting decision. For younger participants, the decision context made no difference. Alm/Bloomquist/McKee (2013) also compared student subjects with university staff members. Although the two subject pools differed in their absolute levels of compliance, the effects of the experimental treatments were the same in the two groups. Wahl/Kastlunger/Kirchler (2010) studied the roles of trust and power in tax compliance and reported similar findings obtained between a student sample in the first experiment and self-employed participants in the second. Cause for concern gives a recent study by Choo/Fonseca/Myles (2014) comparing students

with employed and self-employed participants: As in the previous experiments, the students were less compliant than the other two groups. They also reacted more strongly to the manipulations in the experiment (e.g., the fine rate or whether detailed information about the probability of an audit was provided). Further, students exhibited a stronger bomb crater effect (cf. Mittone, 2006) than other participants (i.e., a drop in compliance after an audit occurred). The interaction effects observed in this study raise serious doubts about the generalizability of research that relies solely on student samples.

In sum, it seems that students act differently in tax experiments than the average taxpayer. It is particularly problematic if they react differently to the parameters that are manipulated in the study, as reported by two of the empirical studies comparing different groups of subjects. However, the use of students as subjects in experimentation most likely will continue due to their availability and willingness to participate in research. As mentioned before, for some research questions, their greenness might also be advantageous.

3.2.4. Windfall income

A potential limitation of the experimental method in general is that income in experiments is typically a windfall gain (Arkes/Joyner/Pezzo/Nash/Siegel-Jacobs/Stone, 1994), whereas in reality, money has to be earned with some degree of effort. Source of income has been shown to affect decisions in several experimental paradigms such as the ultimatum game (Ruffle, 1998), the dictator game (Cherry/Frykblom/Shogren, 2002), and the public goods game (Muehlbacher/Kirchler, 2009). Participants usually tend to be less cooperative when their endowments are earned rather than provided by the experimenter. For tax compliance, however, the opposite seems to be true. In a tax experiment, non-cooperation means taking the risk of paying a fine in the case of an audit, and according to the »reverse« sunk cost effect (Zeelenberg/van Dijk, 1997), when a great deal of effort is invested into earning one's income, one tends to make risk-averse choices. In a laboratory experiment by Kirchler and colleagues (Kirchler/Muehlbacher/Hoelzl/Webley, 2009), participants either earned their income (e.g., by completing items from

an intelligence test) or they obtained their income without much effort. In line with the »reverse« sunk cost effect, tax compliance was higher in the high-effort condition, and evasion occurred more often when the income was obtained without much effort. Similar results were found with a different method when scenarios on a questionnaire described the income to be declared as achieved by either hard work or low effort (Muehlbacher/Kirchler, 2008; Muehlbacher/Kirchler/Hoelzl/Ashby/Berti/Job/Kemp/Peterlik/Roland-Levy/Waldherr, 2008). Bühren/Kundt (2013) reported the opposite finding. In their study, tax cheating was more likely when effort was high. Participants in Choo/Fonseca/Myles' (2014) study earned income in accordance with their performance in a working task. Hence, income was achieved through a mixture of effort and skill. While higher income was related to higher compliance, wealth that accumulated over experimental periods was negatively related to compliance. No main effect of income source was found by Durham and colleagues (Durham/Manly/Ritsema, 2014), but they observed an interaction between effort and income level. Boylan/Sprinkle (2001) reported an interaction between tax rate and income source. Participants in their experiment reacted more strongly to different tax rates when their income was earned rather than provided by the experimenters.

For tax behavior, it seems to matter whether income was earned by hard and effortful work. As Zeelenberg/van Dijk (1997) put it, sometimes »too much [has been] invested to gamble« (p. 689). Hence, in experimental tax research, it seems advisable to let participants work for their experimental income. Particularly in multi-trial experiments, introducing working tasks between the tax reporting phases offers the additional advantage of breaking up the monotony of completing tax reports again and again.

3.2.5. Temporal aspects of compliance

The decision to evade taxes in the real world may be conceptualized as an inter-temporal choice between the options of accepting a sure loss now or taking the risk of evasion and probably paying a fine at some point in the future. In most countries, the timespan in which a tax audit may occur is large; in Austria, for example, it can take up to

7 years for taxpayers to know whether their tax returns are being checked. On the one hand, the temporal spacing between the compliance decision and its potential consequences could lead to a discounting of the uncertain fine because it is so far in the future. This would reduce the deterring effects of sanctions. On the other hand, anticipating the uneasy feeling of waiting for a potential audit may also increase compliance. Empirical evidence has shown a preference for facing unpleasant events such as a painful tooth surgery immediately to avoid the dread of waiting for it (Loewenstein, 1987). In a tax-compliance decision, the simplest strategy for avoiding the fear of being caught while waiting for a potential audit is to file an honest tax return. The effect of the temporal distance between the compliance decision and the resolution of uncertainty of whether an audit will take place was tested in an experiment by Muehlbacher and colleagues (Muehlbacher/Mittone/Kastlunger/Kirchler, 2012). By solving crossword puzzles, participants earned a gross income of 20 Euro, which was taxed at a rate of 30%. Participants were instructed to put the amount of taxes they wanted to pay in an envelope and hand it to the experimenter. In a control condition, immediately after receiving the envelope, the experimenter rolled a die to decide whether it would be opened. Hence, as in typical tax experiments, participants received immediate feedback about whether their tax payment was checked. In another condition, however, participants had to wait for this uncertainty to be resolved. They were required to come back to the laboratory after three weeks; at this point, the die was thrown and if they had evaded their taxes, they had to pay their penalty. Evasion was much higher in the condition with immediate feedback than in the condition where three weeks passed until participants were informed about whether an audit would take place. Kogler, Mittone, and Kirchler (submitted manuscript) conducted a similar experiment with multiple periods. In one condition, participants received feedback about whether an audit would occur after each period, and in another condition, they completed their tax reports without immediate feedback. Instead, in this group, a summary was provided at the end of the experiment showing which periods were audited and whether a fine had to be paid. Again, tax compliance was lower in the condition with immediate feedback.

A slightly different temporal facet of the reporting decision was studied in an experiment by Cadsby and colleagues (Cadsby/Maynes/Trivedi, 2006). Participants in one treatment were allowed to consider their tax reporting decision for one week before the tax payments were collected. Compliance was substantially lower compared to a control condition in which taxes were paid immediately.

Before the significance of temporal aspects in tax reporting can be assessed, further studies are needed to replicate the scarce empirical evidence that has been reported so far, and more importantly, studies are needed to explore potential interactions with the variables of interest in the experiment.

4. Conclusions

Tax behavior is a complex issue, and all methods available for research have their own disadvantages. Identifying the appropriate methodological approach largely depends on the particular research question. Laboratory experiments are often the only option that can be used to manipulate specific variables or to observe a certain behavior under controlled conditions. They allow individual choices to be studied rather than their aggregate, have high internal validity, and thus also enable researchers to determine causality.

However, little is known about the external validity of experiments in tax compliance research. Although its application in this field has often been criticized, only a few empirical studies have directly addressed problems with the generalizability of experimental results. The growth of research addressing methodological issues has been slow, and the concerns brought up in a similar review more than a decade ago (Torgler, 2002) have remained largely the same. More research on the external validity of tax experiments is needed. The essential cues for the framing of the reporting decision in an experiment need to be determined to ensure that the same psychological mechanisms that drive tax behavior in the real world are induced in experiments.

All of the issues discussed here would be of minor importance if changing the criticized aspects of the experimental environment would simply shift tax compliance in one or the other direction. It would be more problematic if certain features of the

design were to interact with the variables of interest in the study. Such interactional effects would create severe limitations in interpreting experimental results. Hence, methodological research should not only explore the main effects of altering the experimental environment but should also test for potential interactions with parameters that are commonly known to affect compliance such as the probability of an audit or the tax rate. If a specific experimental setting allows observations made with non-experimental research methods to be replicated, this would – at least to some degree – indicate the external validity of the respective design.

From the few empirical studies conducted so far, conclusions for experimental practice can be drawn only cautiously. (i) It seems advisable for researchers to use tax-framing rather than »neutral« language and to avoid instructing participants to maximize their income. Compliance in a tax setting seems to be higher than risk-taking rates in a pure gambling context, and the experimental framing was observed to interact with other variables such as participants' age and audit probabilities. (ii) To increase the realism of the simulated environment, the collected tax money should be redistributed in some way to the participants. This may slightly complicate the experimental procedure, but introducing a public good financed by the tax payments changes the compliance decision from an individual decision under risk to a more realistic coordination problem with authorities and the community. Whereas taking the risk of evading taxes might be affected by audit probabilities and fines, cooperation with other taxpayers could also depend on different factors such as social norms or trust. (iii) Another recommendation for experimentation is to allow participants to earn their income by engaging in some task (e.g., by playing trivia games or solving crosswords). Windfall money seems to be treated differently than earned income. Moreover, source of income was found to interact with the essential parameters of the experimental design such as levels of income and the tax rate. Further, completing tasks between the taxpaying phases of an experiment offers participants a little variety in the experimental session. (iv) Regarding the subject pool, in theory, student participants should be avoided whenever possible, in particular when the goal is to generalize the results to experienced taxpayers. However, assuming that the use of students as subjects will continue for practical

reasons, further research should more systematically explore the limitations of relying on student samples.

If more studies would directly address methodological issues, this would help to improve the design and increase the external validity of tax compliance experiments. But is it desirable to reach a consensus in the tax research community with regard to a standard design for tax compliance experiments? At the moment, the most basic elements are operationalized differently in each study. Even the measurement of compliance – the dependent variable – is quite diverse. In an experiment by Mittone (2006), participants' task was to indicate the amount of taxes they wanted to pay. In another experiment conducted by Torgler (2007), participants had to report the amount of income they wanted to declare. The first situation emphasizes losses from taxpaying more than the latter, a difference that could potentially affect participants' compliance. Given the variety of designs, an evidence-based discussion about the essential features of the experimental environment could improve the overall validity of results and would enhance the comparability of observations across studies conducted by different research teams. A »tax game« could be defined with suggestions regarding, for instance, the level of income, the information provided about the probability of being audited, or penalties for evasion. On the other hand, a strict specification as is common in other areas of experimentation – where, for instance, the rules for an ultimatum game or a public goods game are normally not changed without good reason – may not be possible for a topic as broad as tax behavior. Even the manner in which such features as basic as the dependent variable are operationalized (e.g., whether non-compliance can be shown solely by underreporting income or also by exaggerating deductions) largely depends on the research question. Further, it can be argued that heterogeneity in experimental designs is advantageous as it allows for replications in different settings and that the »best« design will evolve and survive automatically over time. Hence, the ideal of a standardized »tax game« should probably be lowered to the goal of defining »dos« and »don'ts« for good experimental practice in tax compliance research.

In general, experimental research is undertaken for different purposes. Roth (1995) established a classification regarding the underlying research

rationale. »Speaking to theorists« describes a set of experiments that provide empirical feedback for theorists by setting an experimental environment based on the parameters of a given theory. »Searching for facts« refers to the exploration of variables about which little theoretical background is given. By replicating and accumulating empirical observations, such experiments may lead to the formulation of new theories. As Torgler (2002) stated, most tax compliance experiments fall into these two categories. The motivation behind a third set of experiments is described as »whispering into the ears of princes.« Their design tries to mimic important aspects of real-life situations to provide evidence-based arguments for policy makers. Considering the concerns about external validity that are raised in this review, it seems reasonable to verify experimental findings by replication through other methods before »whispering into the ears of princes.« However, for some research questions, experiments are the only option for empirical clarification. For instance, in studies on tax amnesties (Alm/McKee/Beck, 1990; Torgler/Schaltegger, 2005), the experimental method was chosen because »[f]ield data simply do not exist to examine the long run effects of an amnesty« (Alm/McKee/Beck, 1990, p. 23). Hence, experiments were the only way to keep up the dialogue between researchers and policy makers.

The popularity and acceptance of lab experiments as a method for empirically studying tax compliance has increased. Their widespread use and the lack of alternatives for some research projects call for more systematic research on the external validity of experimental observations. Until more evidence has been gathered, results obtained in the lab should be interpreted with care and – whenever possible – replicated with other empirical methods.

References

- Ali, M. M./Cecil, H. W./Knoblett, J. A. (2001): The effects of tax rates and enforcement policies on taxpayer compliance: A study of self-employed taxpayers. In: *Atlantic Economic Journal*, Vol. 29 (2001), 2, pp. 186–202.
- Allingham, M. G./Sandmo, A. (1972): Income tax evasion: A theoretical analysis. In: *Journal of Public Economics*, Vol. 1 (1972), 3–4, pp. 323–338.
- Alm, J. (1991): A perspective on the experimental analysis of taxpayer reporting. In: *The Accounting Review*, Vol. 66 (1991), 3, pp. 577–593.

- Alm, J. (2012): Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. In: *International Tax and Public Finance*, Vol. 19 (2012), 1, pp. 54–77.
- Alm, J./Bloomquist, K./McKee, M. (2015): On the external validity of laboratory tax compliance experiments. In: *Economic Inquiry*, Vol. 53 (2015), 2, pp. 1170–1186.
- Alm, J./Jackson, B. R./McKee, M. (1992): Estimating the determinants of taxpayer compliance with experimental data. In: *National Tax Journal*, Vol. 45 (1992), 1, pp. 107–114.
- Alm, J./Kirchler, E./Muehlbacher, S. (2012): Combining Psychology and Economics in the analysis of compliance: From enforcement to cooperation. In: *Economic Analysis & Policy*, Vol. 42 (2012), 2, pp. 133–151.
- Alm, J./McClelland, G. H./Schulze, W. D. (1992): Why do people pay taxes? In: *Journal of Public Economics*, Vol. 48 (1992), 1, pp. 21–38.
- Alm, J./McKee, M./Beck, W. (1990): Amazing grace: Tax amnesties and compliance. In: *National Tax Journal*, Vol. 43 (1990), 1, pp. 23–37.
- Alm, J./Sanchez, I./de Juan, A. (1995): Economic and noneconomic factors in tax compliance. In: *Kyklos*, Vol. 48 (1995), 1, pp. 3–18.
- Arkes, H. R./Joyner, C. A./Pezzo, M. V./Nash, J. G./Siegel-Jacobs, K./Stone, E. (1994): The Psychology of windfall gains. In: *Organizational Behavior and Human Decision Processes*, Vol. 59 (1994), 3, pp. 331–347.
- Ashby, J. S./Webley, P./Haslam, A. S. (2009): The role of occupational taxpaying cultures in taxpaying behaviour and attitudes. In: *Journal of Economic Psychology*, Vol. 30 (2009), 2, pp. 216–227.
- Baldry, J. C. (1986): Tax evasion is not a gamble: A report on two experiments. In: *Economics Letters*, Vol. 22 (1986), 4, pp. 333–335.
- Blackwell, C. (2010): A meta-analysis of incentive effects in tax compliance experiments. In: Alm, J./Martinez-Vazquez, J./Torgler, B. (Eds.): *Developing alternative frameworks for explaining tax compliance*. Oxon 2010, pp. 97–112.
- Boylan, S. J./Sprinkle, G. B. (2001): Experimental evidence on the relation between tax rates and compliance: The effect of earned vs. endowed income. In: *The Journal of the American Taxation Association*, Vol. 23 (2001), 1, pp. 75–90.
- Braithwaite, V. (2003): Dancing with tax authorities: Motivational postures and noncompliant actions. In: Braithwaite, V. (Ed.): *Taxing democracy*. Aldershot, UK 2003, pp. 15–39.
- Bühren, C./Kundt, T. (2013): Worker of shirker: Who evades more taxes? A real effort experiment. In: *Joint Discussion Paper Series in Economics*, No. 26–2013, pp. 1–30.
- Cadsby, C. B./Maynes, E./Trivedi, V. (2006): Tax compliance and obedience to authority at home and in the lab: A new experimental approach. In: *Experimental Economics*, Vol. 9 (2006), 4, pp. 343–359.
- Cherry, T. L./Frykblom, P./Shogren, J. F. (2002): Hardnose the dictator. In: *The American Economic Review*, Vol. 92 (2002), 4, pp. 1218–1221.
- Choo, L./Fonseca, M./Myles, G. (2014): Do students behave like real taxpayers? Experimental evidence on taxpayer compliance from the lab and from the field (*Discussion Paper 004-14*). Manuscript retrieved from <http://tarc.exeter.ac.uk/publications/discussionpapers/> (Last access: 3.11.2015)
- Clotfelter, C. T. (1983): Tax evasion and tax rates: An analysis of individual returns. In: *The Review of Economics and Statistics*, Vol. 65 (1983), 3, pp. 363–373.
- Cummings, R. G./Martinez-Vazquez, J./McKee, M./Torgler, B. (2009): Tax morale affects tax compliance: Evidence from surveys and an artefactual field experiment. In: *Journal of Economic Behavior & Organization*, Vol. 70 (2009), 3, pp. 447–457.
- Dawes, R. M. (1980): Social dilemmas. In: *Annual Review of Psychology*, 31 (1980), pp. 189–193.
- Deci, E. L./Koestner, R./Ryan, R. M. (1999): A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. In: *Psychological Bulletin*, Vol. 125 (1999), pp. 627–668.
- Durham, Y./Manly, T. S./Ritsem, C. (2014): The effects of income source, context, and income level on tax compliance decisions in a dynamic experiment. In: *Journal of Economic Psychology*, Vol. 40 (2014), pp. 220–233.
- Elffers, H./Robben, H. S. J./Hessing, D. J. (1992): On measuring tax evasion. In: *Journal of Economic Psychology*, Vol. 13 (1992), 4, pp. 545–567.
- Elffers, H./Weigel, R. H./Hessing, D. J. (1987): The consequences of different strategies for measuring tax evasion behavior. In: *Journal of Economic Psychology*, Vol. 8 (1987), 3, pp. 311–337.
- Friedland, N./Maital, S./Rutemberg, A. (1978): A simulation study of income tax evasion. In: *Journal of Public Economics*, Vol. 10 (1978), 1, pp. 107–116.
- Gërxhani, K. (2007): »Did you pay your taxes?« How (not) to conduct tax evasion surveys in transition countries. In: *Social Indicators Research*, Vol. 80 (2007), 3, pp. 555–581.
- Henderson, B. C./Kaplan, S. E. (2005): An examination of the role of ethics in tax compliance decisions. In: *The Journal of the American Taxation Association*, Vol. 27 (2005), 1, pp. 39–72.
- Hertwig, R./Ortmann, A. (2001): Experimental practices in economics: A methodological challenge for psychologists? In: *Behavioral and Brain Sciences*, Vol. 24 (2001), 3, pp. 383–451.
- Hite, P. A. (1988): An examination of the impact of subject selection on hypothetical and self-reported taxpayer noncompliance. In: *Journal of Economic Psychology*, Vol. 9 (1988), pp. 445–466.
- Hoffman, E./McCabe, K./Smith, V. L. (1996): Social distance and other-regarding behavior in dictator games. In: *The American Economic Review*, Vol. 86 (1996), 3, pp. 653–660.
- Kirchler, E. (1999): Reactance to taxation: Employers' attitudes towards taxes. In: *Journal of Socio-Economics*, Vol. 28 (1999), 2, pp. 131–138.
- Kirchler, E./Hoelzl, E./Wahl, I. (2008): Enforced versus voluntary tax compliance: The »slippery slope« framework. In: *Journal of Economic Psychology*, Vol. 29 (2008), 2, pp. 210–225.
- Kirchler, E./Muehlbacher, S./Hoelzl, E./Webley, P. (2009): Effort and aspirations in tax evasion: Experimental evidence. In: *Applied Psychology: An International Review*, Vol. 58 (2009), 3, pp. 488–507.
- Kirchler, E./Muehlbacher, S./Kastlunger, B./Wahl, I. (2010): Why pay taxes? A review of tax compliance decisions. In: Alm, J./Martinez-Vazquez, J./Torgler, B. (Eds.): *Developing alternative frameworks for explaining tax compliance*. Oxon 2010, pp. 15–31.
- Kogler, C./Mittone, L./Kirchler, E. (submitted manuscript): Delayed feedback on tax audits affects compliance and fairness perceptions.
- Kopelman, S./Weber, J. M./Messick, D. M. (2002): Factors Influencing Cooperation in Commons Dilemmas: A Review of Experimental Psychological Research. In: Ostrom, E./Dietz, T./Dolšak, N./Stern, P. C./Stonich, S./Weber, E. U. (Eds.): *The Drama of the Commons*. Washington DC 2002, pp. 113–156.
- Loewenstein, G. (1987): Anticipation and the valuation of delayed consumption. In: *The Economic Journal*, Vol. 97 (1987), 387, pp. 666–684.

- Marwell, G./Ames, R.E. (1981): Economists free ride, does anyone else?: Experiments on the provision of public goods, IV. In: *Journal of Public Economics*, Vol. 15 (1981), 3, pp. 295–310.
- Mittone, L. (2006): Dynamic behaviour in tax evasion: An experimental approach. In: *Journal of Socio-Economics*, Vol. 35 (2006), 5, pp. 813–835.
- Muehlbacher, S./Kirchler, E. (2008): Arbeitsaufwand, Anspruchsniveau und Steuerehrlichkeit [Workload, aspiration level, and tax compliance]. In: *Zeitschrift für Arbeits- und Organisationspsychologie*, Vol. 52 (2008), 2, pp. 91–96.
- Muehlbacher, S./Kirchler, E. (2009): Origin of endowments in public good games: The impact of effort on contributions. In: *Journal of Neuroscience, Psychology, and Economics*, Vol. 2 (2009), 1, pp. 59–67.
- Muehlbacher, S./Kirchler, E. (2013): Mental accounting of self-employed taxpayers: On the mental segregation of the net income and the tax due. In: *FinanzArchiv: Public Finance Analysis*, Vol. 69 (2013), 4, pp. 412–438.
- Muehlbacher, S./Kirchler, E./Hoelzl, E./Ashby, J./Berti, C./Job, J./Kemp, S./Peterlik, R.U./Roland-Levy, C./Waldherr, K. (2008): Hard-earned income and tax compliance: A survey in eight nations. In: *European Psychologist*, Vol. 13 (2008), 4, pp. 298–304.
- Muehlbacher, S./Mittone, L./Kastlunger, B./Kirchler, E. (2012): Uncertainty resolution in tax experiments: Why waiting for an audit increases compliance. In: *Journal of Socio-Economics*, Vol. 41 (2012), 3, pp. 289–291.
- Roth, A. (1995): Introduction to Experimental Economics. In: Kagel, J./Roth, A. (Eds.): *The Handbook of Experimental Economics*. Princeton 1995, pp. 3–110.
- Ruffle, B.J. (1998): More is better, but fair is fair: Tipping in dictator and ultimatum games. In: *Games and Economic Behavior*, Vol. 23 (1998), 2, pp. 247–265.
- Slemrod, J./Blumenthal, M./Christian, C. (2001): Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. In: *Journal of Public Economics*, Vol. 79 (2001), 3, pp. 455–483.
- Torgler, B. (2002): Speaking to theorists and searching for facts: Tax morale and tax compliance in experiments. In: *Journal of Economic Surveys*, Vol. 16 (2002), 5, pp. 657–683.
- Torgler, B. (2003): Does culture matter? Tax morale in an East-West-German comparison. In: *FinanzArchiv: Public Finance Analysis*, Vol. 59 (2003), 4, pp. 504–528.
- Torgler, B. (2007): Tax compliance and tax morale: a theoretical and empirical analysis. Cheltenham, UK 2007.
- Torgler, B. (2013): A field experiment in moral suasion and tax compliance focusing on underdeclaration and overdeduction. In: *FinanzArchiv: Public Finance Analysis*, Vol. 69 (2013), 4, pp. 393–411.
- Torgler, B./Schaltegger, C.A. (2005): Tax amnesties and political participation. In: *Public Finance Review*, Vol. 33 (2005), 3, pp. 403–431.
- Wahl, I./Kastlunger, B./Kirchler, E. (2010): Trust in authorities and power to enforce tax compliance: An empirical analysis of the »slippery slope« framework. In: *Law & Policy*, Vol. 32 (2010), 4, pp. 383–406.
- Wartick, M.L./Madeo, S.A./Vines, C.C. (1999): Reward dominance in tax-reporting experiments: The role of context. In: *The Journal of the American Taxation Association*, Vol. 21 (1999), 1, pp. 20–31.
- Webley, P./Ashby, J. (2010): The economic psychology of value added tax compliance. In: Alm, J./Martinez-Vazquez, J./Torgler, B. (Eds.): *Developing alternative frameworks for explaining tax compliance*. Oxon 2010, pp. 238–259.
- Webley, P./Robben, H.S.J./Elffers, H./Hessing, D.J. (1991): *Tax evasion: An experimental approach*. Cambridge, UK 1991.
- Wenzel, M. (2004): An analysis of norm processes in tax compliance. In: *Journal of Economic Psychology*, Vol. 25 (2004), 2, pp. 213–228.
- Zeelenberg, M./van Dijk, E. (1997): A reverse sunk cost effect in risky decision making: Sometimes we have too much invested to gamble. In: *Journal of Economic Psychology*, Vol. 18 (1997), 6, pp. 677–691.