



Instructor's Guide: Tax Compliance and Audits

Laboratory implementation, theoretical foundations,
experimental evidence, and variants

Behavioral and Experimental Economics Research Lab
Department of Economics



Revised edition | October 2026

<https://beer-lab.org/>

How to use this guide

This guide supports preparation, implementation, and classroom discussion of the Tax Compliance and Audits family in the Behavioral and Experimental Economics Research Lab (BEER Lab). It is written for instructors in public economics, behavioral economics, microeconomics, and experimental methods, including instructors who have not previously used oTree or tax-compliance experiments.

The document offers three reading paths. The quick guide is enough to organize a safe session. Operational chapters describe what students and instructors see. Conceptual chapters develop the Allingham–Sandmo model, behavioral extensions, laboratory and field evidence, and the use of every treatment in classroom feedback.

Common rules appear once. Basic and extended variants then focus on what changes: parameters, predictions, metrics, and questions. The activity deliberately simplifies a tax system. It should teach incentives and behavior, not judge people or represent the law of any country.

Author

This guide was prepared by Stanislaw Maldonado, research professor in the Department of Economics at Tecnológico de Monterrey, Monterrey Campus. For questions or suggestions about its content, please contact the author at stanislao.maldonado@tec.mx.

For the instructor

The platform keeps decisions private and reveals aggregates only. Never project individual codes, describe a low report as proof of dishonesty, or connect the exercise to students' actual tax obligations.

Contents

How to use this guide	i
1 Quick implementation guide	1
1.1 Access through the updated laboratory website	1
1.2 Family summary	4
1.3 Minimum sequence	5
1.4 Learning objectives	5
2 Access, rooms, and family isolation	6
2.1 Path from the laboratory website	6
2.2 Participant and instructor routes	6
2.3 Technical preparation	6
3 Participant experience	7
3.1 Cover and rules	7
3.2 Instructions and parameters	7
3.3 Decision, audit draw, and result	7
3.4 Explicit completion	7
4 Economic mechanism and calculations	13
4.1 Notation	13
4.2 Complete example	13
4.3 Marginal incentive	13
4.4 Fiscal return	14
5 Theoretical foundations	15
5.1 Allingham–Sandmo	15
5.2 Yitzhaki’s observation	15
5.3 Deterrence and probability weighting	15
5.4 Tax morale, legitimacy, and reciprocity	15
5.5 Social interactions	16
5.6 Dynamics and audit history	16
6 Empirical evidence and scope	17
6.1 Laboratory evidence	17
6.2 Field evidence	17
6.3 Basic editions	17
6.4 One decision	18
6.5 Ten periods with feedback	18
6.6 Ten periods with end-only feedback	18
7 Extended treatments	19

7.1	Benchmark institution	19
7.2	High audit probability	19
7.3	High penalty	20
7.4	High tax rate	20
7.5	Fiscal return	20
7.6	Transparent use	21
7.7	Norm information	21
7.8	Responsive enforcement	21
8	Instructor report and interpretation	23
8.1	Controlled revelation	23
8.2	Indicators	23
8.3	Evolution	25
8.4	Feedback script	25
8.5	Lesson plans	25
8.6	Public economics	25
8.7	Behavioral economics	25
8.8	Institutional design	25
8.9	Methods seminar	26
9	Solved exercises and discussion	27
9.1	No-audit calculation	27
9.2	Audit calculation	27
9.3	Expected value	27
9.4	High audit	27
9.5	Fiscal return	27
9.6	Discussion prompts	28
9.7	Design exercise	28
10	Advanced interpretation and future extensions	29
10.1	Observed reports and identification	29
10.2	Realized randomness and compliance	29
10.3	Revenue, compliance, and welfare	30
10.4	Administrative costs and audit selection	30
10.5	Internal and external validity	30
10.6	Equity, heterogeneity, and legitimacy	31
10.7	Extensions not yet implemented	31
10.8	Debrief rubric	31
10.9	Comparisons across classes and semesters	31
11	Data, privacy, and troubleshooting	33
11.1	Recorded data	33
11.2	Common problems	33
11.3	Post-session checklist	33

List of Figures

1.1	Current institutional website. Select CLASSROOM in the top navigation menu to open the teaching experiments. Captured on October 7, 2026.	1
1.2	Current catalog: search by title, topic, or JEL code. Then locate this family's card and select language and edition. Captured on October 7, 2026.	2
1.3	Demo contains only the configurations of this game family, in both languages and editions.	3
1.4	Rooms contains the four rooms for this family: basic and extended editions in Spanish and English.	3
1.5	Session-creation form for the basic English room. Check the selected configuration and participant requirement before creating the session.	4
3.1	English cover with institutional identity, purpose, and laboratory rules.	8
3.2	Current instructions for the basic edition: initial resources, decision rules, and experimental units.	9
3.3	Current comprehension check with an incorrect answer, its specific explanation, and retained responses.	10
3.4	Income-report screen with account, decision control, and preview.	11
3.5	Period outcome without an audit. Audited outcomes add recovered tax and penalty separately.	12
3.6	Final screen with explicit completion and private participant summary.	12
8.1	Hidden instructor report before completion and explicit revelation.	23
8.2	Revealed report with indicators, histogram, round evolution, and post-detection comparison.	24

Quick implementation guide

1.1 Access through the updated laboratory website

Start at <https://beer-lab.org/en/>. Select **CLASSROOM** in the top navigation menu; it opens <https://teaching.beer-lab.org/en/>. In the **Classroom experiments** catalog, use the search box to locate the activity by name. You can also filter by **Topic** or **JEL code**. Clear any filters left from an earlier search before looking for another game.

Each family's card distinguishes **Español** and **English** access links. The **Show extended version** checkbox selects that edition before following a link. Catalog language and session language are separate choices: verify both. Students use the room specified by the instructor; the instructor prepares and monitors the session through this family's own administrative pages.



Figure 1.1: Current institutional website. Select CLASSROOM in the top navigation menu to open the teaching experiments. Captured on October 7, 2026.



Figure 1.2: Current catalog: search by title, topic, or JEL code. Then locate this family's card and select language and edition. Captured on October 7, 2026.

Checking resources, comprehension, and participant requirements

Before opening the room, check the selected configuration and the number of students present. The session-creation form states the group size and minimum required by that configuration. An individual activity accepts any positive participant count; an interactive activity must satisfy its stated requirement: complete groups when size is fixed, or a minimum participant count when size is flexible. Requirements may change across treatments, for example when adding a third party or changing an institution's size. If you edit a grouping parameter, check the requirement for the edited configuration, rather than relying only on the demo's suggested count.

In the instructions, ensure that students can identify what each role receives, what they can decide, and how the outcome is calculated. An initial account, a distributable budget, and final earnings need not be the same amount. Refer to the resources panel and the example for the selected treatment. Do not carry the basic edition's endowment into a treatment that changes accounts or introduces additional positions.

ECU stands for *Experimental Currency Units*: the activity's unit of account. An outcome of 100 ECU does not automatically equal 100 pesos or 100 dollars. Conversion depends on the session configuration. Before announcing a payment, confirm with the laboratory whether money is paid, the conversion rate, and which decisions count for payment. In hypothetical modules, describe amounts as outcomes in the stated situation.

When a comprehension check is included, an incorrect response receives an explanation beside that question, identifying the submitted answer and the correct answer. Other answers remain in the form. Ask the student to return to the example, locate the relevant rule, and redo the calculation before correcting the response. Avoid revealing answers to the entire class for an individual error, and do not interpret a comprehension mistake as an economic choice in the game.

to tributario — una decisión	Cumplimiento tributario y auditorías / Tax Compliance and Audits — BEER Lab Seleccione una variante de este juego. Select a version of this game.
to tributario — 10 periodos con retroalimentación	
to tributario — 10 periodos con retroalimentación final	
nce — one decision	
nce — 10 periods with feedback	
nce — 10 periods with end feedback	
to tributario extendido — referencia	
to tributario extendido — auditoría alta	
to tributario extendido — sanción alta	
to tributario extendido — tasa impositiva alta	
to tributario extendido — retorno fiscal	
to tributario extendido — uso transparente	
to tributario extendido — información normativa	
to tributario extendido — fiscalización responsiva	
x Compliance — benchmark	
x Compliance — high audit probability	
x Compliance — high penalty	
x Compliance — high tax rate	
x Compliance — fiscal return	
x Compliance — transparent use	
x Compliance — norm information	
x Compliance — responsive enforcement	

Figure 1.3: Demo contains only the configurations of this game family, in both languages and editions.

Rooms

Current rooms:

[ES] Cumplimiento tributario y auditorías
[EN] Tax Compliance and Audits
[ES] Cumplimiento tributario y auditorías: extendido
[EN] Tax Compliance and Audits: extended

Figure 1.4: Rooms contains the four rooms for this family: basic and extended editions in Spanish and English.

Room: [EN] Tax Compliance and Audits

Create a new session

Session Config

[EN] Tax Compliance — one decision

Number of participants

Individual decisions: any positive participant count is allowed.

Create

[Configure session](#) 

App sequence

tax_compliance_game (10 rounds)

Tax compliance, audits, penalties, fiscal exchange, and tax morale.

Figure 1.5: Session-creation form for the basic English room. Check the selected configuration and participant requirement before creating the session.

Downloading a session with several rounds

For the instructor

Open **Data** within the relevant session and use that session's CSV or CSV-for-Excel download link. The general export has one row per participant and columns for the different rounds. Ten rounds do not necessarily mean ten rows per person. When the family provides its own long-format export, observations may instead be organized by participant and round or another unit defined by the design.

Save the original file together with the configuration name, session code, and active-round count. Check that columns for the last active round are present and that the data belong to the selected session. Empty cells can represent inactive rounds, a pending decision, or a field that does not apply to a role; do not automatically recode them as zeros. To compare learning, preserve round and group identifiers alongside participant identifiers. Share only anonymous aggregates in class, and record session incidents before constructing derived tables.

1.2 Family summary

Purpose	Study how tax rates, audits, penalties, institutional information, and nonmonetary motivations affect income reporting.
Experimental unit	One person makes independent decisions; there are no pairs, social roles, or waits.
Participants	Any positive number; ten or more is recommended for cautious distributional interpretation.

Basic edition	One decision; ten periods with feedback; ten periods with end-only feedback.
Extended edition	Benchmark, high audit, high penalty, high rate, fiscal return, transparency, norm information, and responsive enforcement.
Decision	Report 0–100 ECU from true taxable income of 100 ECU.
Payoff	Base account plus income, minus taxes and penalties, plus fiscal return when applicable.
Topics	tax compliance; tax evasion; audits; penalties; tax morale; public economics.
JEL	A22; C91; H26; H30.
Suggested duration	20–30 minutes for one decision; 35–55 minutes for ten periods and feedback.
Basic room	https://taxcompliance.beer-lab.org/room/tax_compliance_en .
Instructor access	https://taxcompliance.beer-lab.org/demo .

1.3 Minimum sequence

Before class, decide whether the objective is to introduce compliance, compare enforcement tools, or study nonmonetary motivation. A short session uses one decision. A learning session uses ten periods with feedback. For treatment comparisons, create separate sessions or allocate course sections deliberately; do not mix configurations and later attribute any difference to treatment.

1. Open <https://beer-lab.org/> and continue to <https://teaching.beer-lab.org/>.
2. Select the **Tax Compliance and Audits** card.
3. Enter the instructor domain, open **Rooms**, and select the correct language and edition.
4. Create the session with the number of students present and keep **Monitor** and **Report** open.
5. Share only the room or individual participant links.
6. Wait for completion. The report keeps aggregates hidden while decisions remain pending.
7. Elicit predictions, reveal aggregates, and discuss facts before mechanisms and outside evidence.

Checkpoint

Before starting, confirm language, edition, treatment, and number of periods. Demo should list 22 configurations from this family and Rooms should show exactly four rooms. If another game appears, stop and contact the laboratory coordinator.

1.4 Learning objectives

Students should be able to calculate an outcome with and without an audit; derive the marginal expected incentive; explain why audit probability and penalty severity are distinct instruments; recognize that the effect of the tax rate is not mechanically signed; distinguish deterrence, tax morale, reciprocity, and norms; read a reporting distribution; and discuss limits to extrapolation from a classroom laboratory to real tax administration.

Access, rooms, and family isolation

2.1 Path from the laboratory website

The recommended route begins at the institutional website and continues through the teaching catalog. The catalog brings together title, language, topics, guide, and stable domain. The proposed family address is <https://taxcompliance.beer-lab.org/>. Deployment should map this domain to an oTree instance with `GAME_FAMILY=taxcompliance` and a separate database. The code fails closed in production instead of exposing the full repository.

The English basic room is https://taxcompliance.beer-lab.org/room/tax_compliance_en; the Spanish basic room ends in `cumplimiento_tributario_es`. Extended rooms end in `tax_compliance_extended` and `cumplimiento_tributario_extendido`. A basic-room form contains three configurations and never selects an extended one. Each extended room contains eight treatments.

2.2 Participant and instructor routes

Students do not need administrative credentials. They enter through a stable room or an individual link created with the session. The room can be reused; an individual link belongs to one session and may fail after a database reset. The instructor enters `/demo`, opens the session, uses Monitor for progress, Report for aggregates, and Data for post-class export.

Attention

Never share the administrative route or project the Data page. Report is designed for projection because it removes identifiers and keeps results hidden until an explicit action.

2.3 Technical preparation

Create the session before students arrive, check computer connectivity, and test one link in a private browser window. Keep the instructor password available but never place it on slides or in a class message. For ten-period sessions, allow time for slower participants. The report's completion rule also prevents the last student's decision from being inferred from changing aggregates.

An individual activity accepts any positive number of participants. No even number or group multiple is required. This flexibility does not make a small class statistically large: with five people, one decision changes a proportion by twenty percentage points. Emphasize distributions and calculations when samples are small.

Participant experience

3.1 Cover and rules

The first screen presents institutional logos, purpose, and laboratory rules. A checkbox confirms silence, privacy, and remaining on the page. The cover avoids hypotheses and moralizing language. Neutral wording is essential because tax compliance has legal and ethical meaning outside the experiment.

3.2 Instructions and parameters

Instructions distinguish the 50 ECU base account from 100 ECU of true taxable income. The base account prevents negative outcomes under the strongest penalty. Parameter cards show rate, audit probability, and total collection factor. The 60 ECU example makes clear that true income remains 100 when a smaller amount is reported.

The comprehension check requires three rules: true income is fixed; without an audit, tax applies to the report; with an audit, hidden tax is recovered and a penalty applies. Incorrect answers produce an explanatory message. Attempt counts support diagnosis but are absent from the projected report.

3.3 Decision, audit draw, and result

The decision control runs from 0 to 100 ECU. Two cards update as it moves: hidden income and tax without audit. Institutional parameters remain visible. The audit draw occurs only after confirmation, preserving a choice under risk.

With round feedback, participants see the draw and accounting after every decision. With end-only feedback, these screens are omitted during the battery and summarized later. This comparison changes learning and response to audit experience; it is more than a difference in round count.

3.4 Explicit completion

The final screen confirms completion and tells students to wait. It summarizes mean reporting, cumulative outcomes, revenue, audits, and a period table. There is no final button leading to a missing page. This design prevents the confusing “No more pages left” endpoint.



Welcome to the BEER Lab

The Behavioral and Experimental Economics Research Lab of the Department of Economics at Tecnológico de Monterrey is a space designed to learn about economic decision-making through controlled activities.

In this activity, you will decide how much income to report to a tax authority. Decisions are private and will be analyzed only in aggregate.

General laboratory rules

- Remain silent throughout the activity.
- Do not talk to or observe other participants.
- Make decisions individually and privately.
- Do not refresh, close, or leave this page.
- If you have a problem, raise your hand and wait quietly for help.

☐ I understand and will follow these rules.

Start

Powered by [oTree](#)

Figure 3.1: English cover with institutional identity, purpose, and laboratory rules.



Instructions

1 · ECU

Resources and unit of account

Income subject to the exercise's decisions: 100 ECU.

ECU stands for Experimental Currency Units. It is the unit of account for this activity.

This session does not convert ECU into real money.

Benchmark institution

Tax institution in this session

20%

Tax rate

20%

Audit probability

×4

Total factor on audited hidden income

The audit probability and penalty remain constant.

In every period, you receive a 50 ECU base account and 100 ECU of taxable income. Report an integer from 0 to 100 ECU. Your true taxable income does not change when you report a smaller amount.

If there is no audit, the tax rate is charged only on reported income. Unreported income remains in your account.

If there is an audit, tax is charged on reported income and the tax on hidden income is recovered. In addition, the total charge on each hidden ECU is multiplied by the displayed penalty factor.

An independent random draw with the stated probability determines the audit. You cannot know the outcome before reporting.

After each decision, you will see whether an audit occurred and how the period outcome was calculated.

Numerical example

With a 20% rate, reporting 60 ECU creates 12 ECU of reported tax. The other 40 ECU are hidden. Without an audit, 12 ECU are collected. With an audit and factor 4, the hidden-income charge is $4 \times 20\% \times 40 = 32$ ECU; total collection is 44 ECU.

The activity simplifies a tax system to isolate incentives, risk, and institutions. It is not tax advice and does not describe any country's law.

Continue

Con tecnología de gTree

Figure 3.2: Current instructions for the basic edition: initial resources, decision rules, and experimental units.



Review the marked questions. Each shows your answer and an explanation; your other answers are retained.

Comprehension check

20% · 20%

Resources and unit of account

ECU stands for Experimental Currency Units. It is the unit of account for this activity.

This session does not convert ECU into real money.

Benchmark institution

If you report less than 100 ECU, what happens to your true taxable income?

It falls to the reported amount

Your answer: It falls to the reported amount. The correct answer is It remains 100 ECU. Review the rules: true income remains fixed; without an audit, reported income is taxed; an audit recovers hidden tax and applies the penalty.

If there is no audit, which amount is taxed?

Reported income only

If there is an audit, which components may be collected?

Reported tax, recovered tax, and penalty

Continue

Con tecnología de oTree

Figure 3.3: Current comprehension check with an incorrect answer, its specific explanation, and retained responses.



Income report

Period 1 of 1

Benchmark institution

Your account this period

50 ECU

Base account

100 ECU

True taxable income

How many of the 100 ECU of taxable income do you want to report?

Reported income (0–100 ECU)

0 ECU 100 ECU

0 ECU

Unreported income

20 ECU

Tax if there is no audit

20%

Tax rate

20%

Audit probability

×4

Total factor on audited hidden income

The audit probability and penalty remain constant.

Confirm report

Powered by [oTree](#)

Figure 3.4: Income-report screen with account, decision control, and preview.



Period outcome

Period 1 of 1

The report was not audited

Reported income	30 ECU
Unreported income	70 ECU
Tax on reported income	- 6 ECU
Total collection	6 ECU
Account outcome	144 ECU

Continue

Powered by oTree

Figure 3.5: Period outcome without an audit. Audited outcomes add recovered tax and penalty separately.



Activity completed

The activity has ended. Keep this screen open and wait for the instructor's directions.

30 ECU
Mean report

144 ECU Cumulative outcome		6 ECU Cumulative collection		0 Audits	
Period	Reported income	Unreported income	Audit	Total collection	Outcome
1	30	70	No	6	144

Powered by oTree

Figure 3.6: Final screen with explicit completion and private participant summary.

Economic mechanism and calculations

4.1 Notation

Let $Y = 100$ denote taxable income, $B = 50$ the base account, $D \in [0, Y]$ the report, t the tax rate, p the audit probability, and m the total factor applied to tax associated with hidden income. Hidden income is $H = Y - D$. Without an audit, revenue is tD and the outcome is $B + Y - tD$. With an audit, revenue is $tD + mtH$ and the outcome is $B + Y - tD - mtH$. A fiscal-return treatment adds r times revenue.

Theoretical foundation

The penalty is expressed as a total factor. When $m = 4$, audited hidden income is charged four times the corresponding tax. One part recovers omitted tax and three parts are the penalty. Preserve this definition in every calculation.

4.2 Complete example

Suppose $D = 60$, $t = 0.20$, $p = 0.20$, and $m = 4$. Reported tax is 12 and hidden income is 40. Without an audit, the outcome is $50 + 100 - 12 = 138$. With an audit, the hidden-income charge is $4(0.20)(40) = 32$, total revenue is 44, and the outcome is 106. The report creates a lottery paying 138 with probability 0.8 and 106 with probability 0.2, with expected monetary value 131.6.

Full reporting creates certain revenue of 20 and an outcome of 130. A risk-neutral person prefers reporting 60 by 1.6 expected ECU, but that option exposes the account to a 32 ECU drop. A sufficiently risk-averse person may prefer the certain outcome. This distinction links expected value to expected utility.

4.3 Marginal incentive

Hiding one additional ECU saves t without an audit. With an audit, it pays mt rather than the truthful t , a relative loss of $(m-1)t$. Under linear utility the expected change is $(1-p)t - p(m-1)t = t(1 - pm)$. Hiding raises expected value when $pm < 1$, reporting raises it when $pm > 1$, and the participant is marginally indifferent when $pm = 1$.

The rate scales both gain and loss and drops out of the sign. This does not imply tax rates never matter. Concave utility, income effects, moral costs, or another penalty base can change the response. The high-rate treatment separates a precise model prediction from an intuitive claim.

4.4 Fiscal return

The return treatment adds rR , with R denoting revenue. The net private cost of one collected ECU is $1 - r$. At $r = 0.5$, paying one ECU reduces the account by fifty cents. This represents valued services rather than a real transfer or a strategic shared fund, which preserves the individual unit.

Theoretical foundations

5.1 Allingham–Sandmo

Allingham and Sandmo (1972) model reporting as a choice under uncertainty. A taxpayer knows income, rate, detection probability, and penalty and maximizes expected utility. The framework connects evasion to Becker’s economics of crime and shows why compliance depends on risk preferences and enforcement design. Its teaching value comes from turning a complex institution into an explicit lottery.

The model does not imply one report without utility and parameters. Under risk neutrality, the benchmark predicts hiding all income. Concave utility can produce an interior solution. Observed reporting alone does not identify risk aversion because nonmonetary motives can also enter utility.

5.2 Yitzhaki’s observation

Yitzhaki (1974) showed that the rate effect depends on the penalty base. When the fine is proportional to evaded tax, a higher rate raises the gain from hiding and the audited consequence. Under conditions such as decreasing absolute risk aversion, the income effect can raise compliance, contrary to the claim that higher rates always increase evasion. The platform uses a tax-based penalty and exposes this structure.

5.3 Deterrence and probability weighting

Audits and penalties are deterrence tools but may not be behaviorally equivalent. People may overweight a small probability, attend to a recent audit, or react differently to a penalty viewed as disproportionate. Alm, McClelland, and Schulze (1992) found compliance above simple risk-neutral benchmarks and discussed overweighting of audits.

Comparing high audit and high penalty opens questions about administrative cost, credibility, and perception. An extreme fine with virtually no audit may collect less than credible enforcement. The classroom session does not estimate structural elasticities, but it displays the ingredients of a policy tradeoff.

5.4 Tax morale, legitimacy, and reciprocity

Tax morale includes nonpecuniary motives such as duty, identity, reciprocity, trust, legitimacy, and norms. Luttmer and Singhal (2014) stress that these are distinct mechanisms. They may affect

compliance without changing rates, audits, or penalties. Transparency and fiscal return separate information about revenue use from material service value.

Higher reporting under transparency does not prove altruism. It may reflect comprehension, experimenter demand, or a changed reading of the institution. Lower reporting does not prove rejection. Credible causal inference requires treatment assignment and adequate sample size; a teaching session primarily supports conceptual comparison.

5.5 Social interactions

Fortin, Lacroix, and Villeval (2007) distinguish endogenous conformity, fairness effects, and correlated factors. Similar reports within a group do not establish imitation. The norm treatment transparently labels its benchmark as hypothetical rather than claiming that classmates reported a particular amount. Its sign depends on prior beliefs.

5.6 Dynamics and audit history

An audit can generate learning or a temporary reaction. Responsive enforcement makes history objectively change future audit probability. This distinguishes a rule change from a purely psychological response, but a simple post-detection comparison still mixes mechanisms. The report shows denominators and describes rather than overclaims.

Empirical evidence and scope

6.1 Laboratory evidence

Tax experiments manipulate enforcement and observe decisions that are hard to measure in administrative data. Alm and coauthors found reporting above risk-neutral predictions and responses to audits, penalties, and public services. ECU, simple rules, and short horizons strengthen control but limit direct generalization to real liabilities.

Evidence on norms and morale is heterogeneous. Messages can raise reporting, do nothing, or produce different responses across prior beliefs. The guide therefore does not promise a direction. The instructional goal is to compare hypotheses and read a distribution rather than reproduce a published mean.

6.2 Field evidence

Kleven and coauthors (2011) studied a Danish audit experiment and found large differences between third-party-reported and self-reported income. Opportunity matters alongside willingness. The platform represents self-reportable income and should not be generalized to wages automatically reported by employers.

Castro and Scartascini (2015) tested deterrence, reciprocity, and peer messages in an Argentine municipality. The enforcement message increased average compliance, while other messages produced heterogeneous responses. The result illustrates why transparency and norm effects depend on beliefs.

Slemrod (2019) reviews audits, third-party reporting, and remittance regimes. This platform covers only individual reporting under risk. It does not represent firm networks, invoices, withholding, legal avoidance, or administrative capacity. Stating the boundary improves feedback.

Attention

Never use results to classify students as honest or evasive. The decision occurs in an artificial game and does not measure real legal conduct. Project aggregates and preserve the educational purpose.

6.3 Basic editions

6.4 One decision

The one-decision edition removes learning and response to audit history. It is appropriate for introducing the model, eliciting predictions, and spending time on accounting. The report is cross-sectional. In a small class, emphasize ranges and examples rather than strong inference.

6.5 Ten periods with feedback

After each report, participants see the audit and outcome. The sequence reveals learning, exploration, and responses to recent events. Benchmark draws are independent, so a prior audit does not change objective probability. This supports discussion of the gambler's fallacy and the distinction between experience and rules.

6.6 Ten periods with end-only feedback

Decisions accumulate without intermediate audit outcomes. This limits learning from results and produces a profile under constant stated parameters. Comparing it with immediate feedback asks whether experience changes behavior. Use between-group allocation when previous exposure would contaminate the comparison.

For the instructor

For a 75-minute class, allocate twenty minutes to instructions and decisions, ten to predictions, twenty to the report, twenty to theory and evidence, and five to closure and export.

Extended treatments

7.1 Benchmark institution

Mechanism. a 20% tax rate, 20% audit probability, and total factor 4 on tax due on hidden income.

Prediction and interpretation. With linear utility, each hidden ECU saves 0.20 ECU without an audit and creates a relative loss of 0.60 ECU when audited. The marginal expected gain is positive because $0.8(0.20) - 0.2(0.60) = 0.04$. The extreme prediction is to hide all income. Positive reporting motivates discussion of risk aversion, overweighting of small probabilities, tax morale, and institutional trust.

What to use in the report. Use this treatment as the comparison group. Examine the entire distribution rather than only the mean, and distinguish full compliance, partial reporting, and zero reporting.

Feedback script

Ask students to state a prediction before seeing the aggregates, identify the denominator for each percentage, and distinguish what the treatment changes from mechanisms that remain observationally equivalent.

7.2 High audit probability

Mechanism. a 40% audit probability with the benchmark tax rate and penalty.

Prediction and interpretation. The product of audit probability and the total collection factor exceeds one. Hiding one ECU lowers expected payoff for a risk-neutral participant, so full reporting is the benchmark prediction. Partial responses may reflect nonlinear preferences, calculation mistakes, or gradual adjustment.

What to use in the report. Compare mean reporting and full compliance with the benchmark. Discuss sampling variation and the realized number of audits before assigning substantive meaning to a small difference.

Feedback script

Ask students to state a prediction before seeing the aggregates, identify the denominator for each percentage, and distinguish what the treatment changes from mechanisms that remain observationally equivalent.

7.3 High penalty

Mechanism. total factor 6 on tax associated with hidden income.

Prediction and interpretation. The treatment raises the loss conditional on an audit without changing audit frequency. Here $pm = 1.2$, so hiding ceases to be profitable under risk neutrality. The comparison with high audit probability separates enforcement frequency from consequence severity.

What to use in the report. Ask whether frequent moderate audits feel equivalent to rare severe consequences. Connect responses to probability weighting, legitimacy, and procedural preferences.

Feedback script

Ask students to state a prediction before seeing the aggregates, identify the denominator for each percentage, and distinguish what the treatment changes from mechanisms that remain observationally equivalent.

7.4 High tax rate

Mechanism. a 35% tax rate with the benchmark audit and penalty parameters.

Prediction and interpretation. In this proportional design, the tax rate scales both the gain from hiding and the audited loss. Under linear utility it does not change the sign of the marginal incentive. With risk aversion and income effects, reporting may change. The treatment makes Yitzhaki's point about penalty definitions concrete.

What to use in the report. Do not state that higher rates mechanically cause more evasion. Compare the exact model prediction with the observed pattern and identify alternative channels.

Feedback script

Ask students to state a prediction before seeing the aggregates, identify the denominator for each percentage, and distinguish what the treatment changes from mechanisms that remain observationally equivalent.

7.5 Fiscal return

Mechanism. each ECU collected produces 0.50 ECU of service value for the participant's account.

Prediction and interpretation. The net burden of reporting falls because part of revenue returns as valued services. This is a simplified fiscal-exchange treatment. It is not a strategic public-good interaction among participants, which keeps the experimental unit individual.

What to use in the report. Display reporting, revenue, and net outcomes together. Ask whether behavior reflects the monetary return, reciprocity, or a changed interpretation of the institution.

Feedback script

Ask students to state a prediction before seeing the aggregates, identify the denominator for each percentage, and distinguish what the treatment changes from mechanisms that remain observationally equivalent.

7.6 Transparent use

Mechanism. monetary parameters remain unchanged while revenue is explicitly linked to collective services.

Prediction and interpretation. The treatment changes institutional information and framing. The basic monetary model predicts the same report as in the benchmark. Reciprocity, legitimacy, and tax-morale theories permit a response when the use of revenue becomes salient.

What to use in the report. Treat a difference as descriptive evidence from the class. A single, nonrandomized comparison cannot establish that transparency caused the change.

Feedback script

Ask students to state a prediction before seeing the aggregates, identify the denominator for each percentage, and distinguish what the treatment changes from mechanisms that remain observationally equivalent.

7.7 Norm information

Mechanism. a hypothetical median-report benchmark of 80 ECU is displayed.

Prediction and interpretation. The number is explicitly identified as part of the treatment rather than behavior by current students. It permits a transparent discussion of social influence. Descriptive information can raise or lower reports depending on prior beliefs, so theory does not imply a universal sign.

What to use in the report. Elicit prior beliefs verbally before disclosure. Then compare the distribution with 80 ECU and discuss why social-interaction effects are difficult to identify.

Feedback script

Ask students to state a prediction before seeing the aggregates, identify the denominator for each percentage, and distinguish what the treatment changes from mechanisms that remain observationally equivalent.

7.8 Responsive enforcement

Mechanism. next-period audit probability rises from 20% to 60% after detected hidden income.

Prediction and interpretation. The rule creates history dependence. An audit changes the current outcome and future risk. Optimal behavior therefore depends on previous detection and repeated decisions. The report's post-detection comparison is descriptive rather than automatically causal.

What to use in the report. Use the round profile and post-detection panel. Separate learning, history selection, and the response to higher risk. If nobody is detected, explain why the relevant denominator is zero.

Feedback script

Ask students to state a prediction before seeing the aggregates, identify the denominator for each percentage, and distinguish what the treatment changes from mechanisms that remain observationally equivalent.

Instructor report and interpretation

8.1 Controlled revelation

The report opens with participants, completed decisions, pending decisions, and observations. While anything is pending, reveal remains disabled and the page refreshes. Once complete, the instructor reveals aggregates and may hide them again. This protects participants from feedback before they finish.

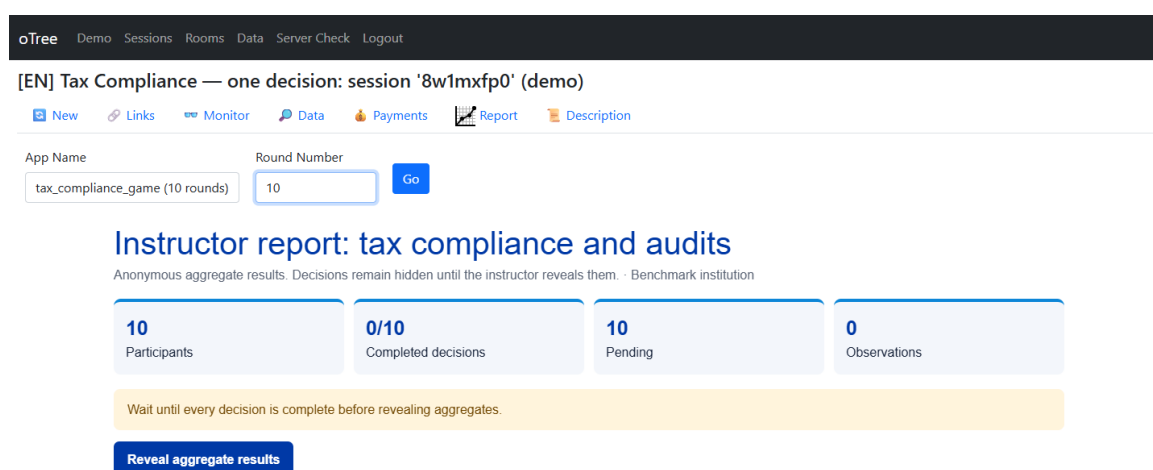


Figure 8.1: Hidden instructor report before completion and explicit revelation.

8.2 Indicators

Mean reporting summarizes level but is sensitive to extremes. The median identifies the center. Full compliance is the share reporting 100 ECU. Audit rate uses all observations; audits with underreporting use audits as the denominator. Revenue includes reported tax, recovered tax, and penalties. Mean account outcome is not social welfare.

The histogram preserves mass at zero, intermediate reports, and complete reporting. Two classes can share a mean while one clusters at 60 and another polarizes between 20 and 100. Distributional shape prevents an average from erasing qualitatively different behavior.

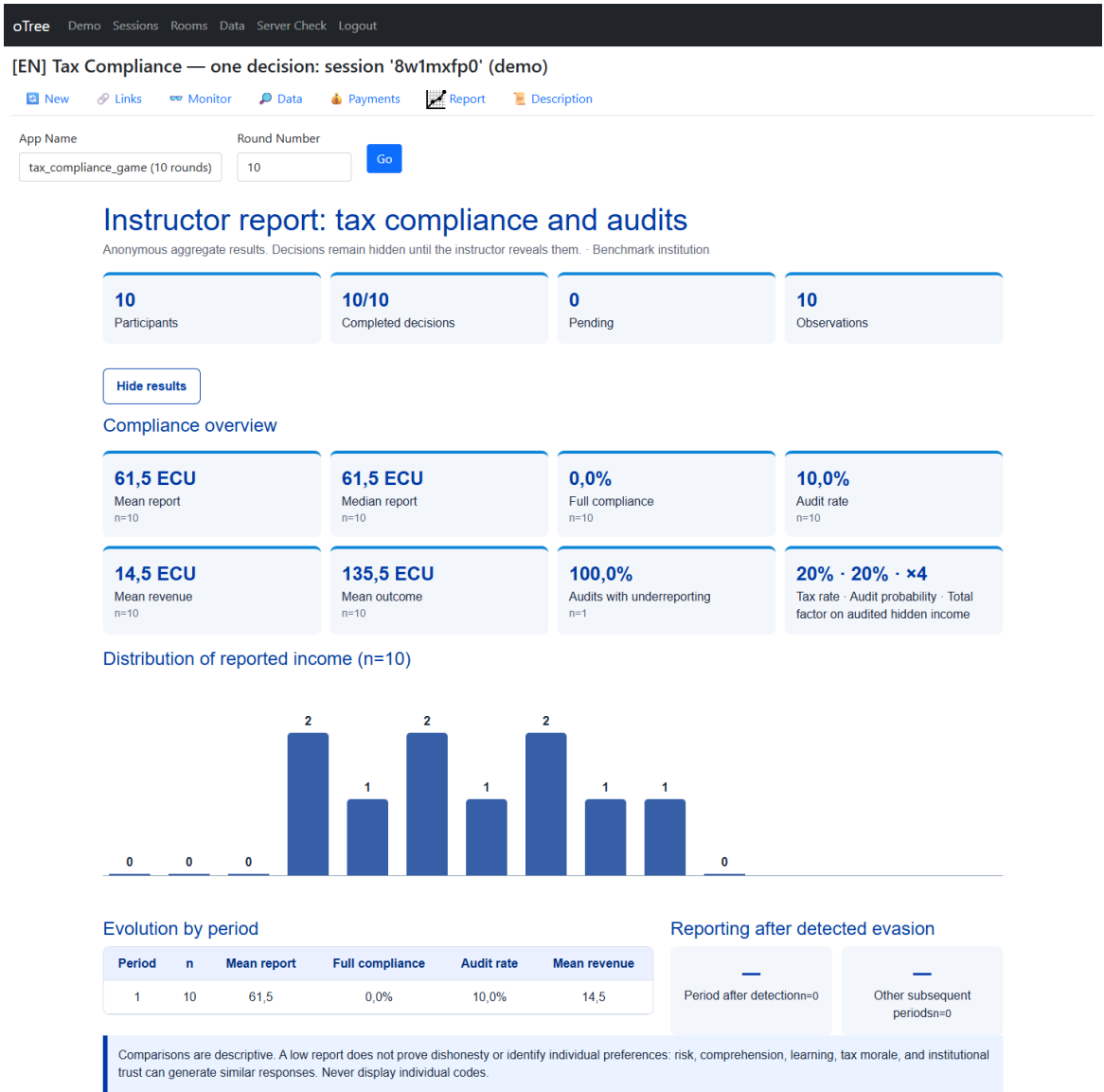


Figure 8.2: Revealed report with indicators, histogram, round evolution, and post-detection comparison.

8.3 Evolution

The round table shows reporting, full compliance, audits, and revenue. A one-decision session has one row; a repeated session reveals trends, jumps, and stabilization. Declining reports after several nonaudits may reflect experience, mistaken beliefs, or chance. Increases after audit may reflect learning, emotion, or a real probability change under responsive enforcement.

The post-detection panel reports means and sample sizes for periods following audited hidden income and other subsequent periods. If $n = 0$, it displays a dash instead of an invalid mean. Even with observations, detection combines prior behavior and random audits, so the comparison is descriptive.

8.4 Feedback script

Start with the risk-neutral prediction. Reveal the histogram and ask for description without explanation. Then ask for denominators, whether median and mean tell the same story, and what followed an audit. Only after that introduce risk aversion, morale, and reciprocity. Close with external evidence and limits.

Feedback script

Fact: mean reporting was X and full compliance Y with n observations. Interpretation: the pattern is compatible with several mechanisms. Prior evidence helps evaluate plausibility. Limit: this session alone does not identify preferences or causal effects.

8.5 Lesson plans

8.6 Public economics

Derive the taxpayer budget before the session and leave the rate effect unresolved. Run benchmark and high rate in different sections. During feedback, compute $t(1 - pm)$ and show why the rate does not change the sign under linear utility. Introduce Yitzhaki and third-party reporting afterward.

8.7 Behavioral economics

Use benchmark, high audit, and high penalty. Elicit predictions about monetary equivalence and perception. Compare distributions. Connect choices to probability weighting, risk aversion, salience, and recent-outcome effects. Explain that the exercise does not estimate individual prospect-theory parameters.

8.8 Institutional design

Compare benchmark, transparency, fiscal return, and responsive enforcement. Ask students to design an institution with an administrative budget. They must justify spending on audits, information, services, or sanctions and state what data would evaluate the policy.

8.9 Methods seminar

Randomize treatments in advance and record the rule. Define a primary outcome such as mean reporting and a secondary outcome such as full compliance. Discuss experimental unit, standard errors, multiple outcomes, and confirmatory versus exploratory analysis. Data exports include participant-period decisions; the projected report remains anonymous.

Solved exercises and discussion

9.1 No-audit calculation

A participant reports 75 ECU at a 20% rate and is not audited. Tax is 15 and the outcome is $50 + 100 - 15 = 135$. Hidden income is already included in the 100 ECU and must not be added again. Double counting is the most common mistake.

9.2 Audit calculation

With the same report and $m = 4$, the hidden-income charge is $4(0.20)(25) = 20$. Total revenue is 35 and the outcome is 115. Of the 20, five recover omitted tax and fifteen are the penalty. The result screen separates these components.

9.3 Expected value

Reporting 75 pays 135 with probability 0.8 and 115 with probability 0.2, for an expected value of 131. Full reporting pays 130 with certainty. A risk-neutral person prefers hiding 25 by one expected ECU; a risk-averse person may prefer certainty. Choice does not identify a risk coefficient by itself.

9.4 High audit

At $p = 0.40$, the same report has value $0.6(135) + 0.4(115) = 127$. Full reporting remains 130. Risk neutrality predicts full compliance. If the class does not change, examine comprehension and perception before saying audits do not matter.

9.5 Fiscal return

With a 50% return and revenue of 35, services add 17.5, raising the audited outcome from 115 to 132.5. Without audit, revenue 15 creates 7.5 and the outcome is 142.5. The rule lowers the net burden but preserves risk.

9.6 Discussion prompts

Is it more legitimate to raise audits or penalties? Which administrative cost is missing? How does employer reporting change the model? Could an 80 ECU norm increase evasion among people expecting 100? How do illegal evasion and legal avoidance differ? Why is revenue not sufficient for welfare? What would an amnesty do? Every answer should identify a mechanism and required evidence.

9.7 Design exercise

Ask teams to propose a ninth treatment with a question, parameter, prediction, screen, formula, metric, and source. Possible topics are third-party reporting, voting over spending, or joint declarations. Reject designs that alter several dimensions at once or falsely describe classmates' behavior.

Advanced interpretation and future extensions

10.1 Observed reports and identification

An observed report is a reduced-form outcome of several forces. Monetary incentives depend on the rate, audit probability, and penalty; expected utility depends on risk preferences; utility may include moral cost, legitimacy, attention, and mistakes. Two participants who report the same amount may reach it for different reasons. One participant may change reports because the calculation became clearer rather than because tax morale changed. The report therefore describes behavior and never labels latent taxpayer types.

Identification improves when a treatment changes one dimension and assignment is random. High audit and high penalty preserve other parameters, while transparency preserves payoffs. Even then, a comparison requires enough independent units. Ten periods from one participant do not equal ten people for inference because preferences and histories are shared. In formal analysis, the person is the unit for standard errors and rounds are observations nested within a person.

This distinction also matters for classroom language. Saying that “there were one hundred observations” can be numerically correct when ten people make ten reports, yet it overstates independent information. The instructor should report both people and participant-period observations whenever treatment conclusions are discussed.

10.2 Realized randomness and compliance

The realized audit rate may differ from the programmed probability in a small session. With ten observations and $p = 0.20$, zero, one, two, or three audits are plausible. Never infer the treatment from the realized count without recalling the rule. The report separates the parameter from the observed audit share, providing a concrete example of ex ante probability and ex post frequency.

In repeated sessions, one person may be audited several times by chance. The program’s reproducible deterministic draw behaves like a uniform draw based on participant identifier, round, and treatment; it does not respond to the report except under responsive enforcement. The instructor should not promise exactly twenty percent audits in one class or treat a cluster as a software error without checking the underlying probability.

10.3 Revenue, compliance, and welfare

Higher reporting is not identical to higher welfare. Revenue can fund services, but audits use resources, penalties may be transfers, and risk can lower utility. Normative evaluation needs the marginal value of public funds, administrative cost, equity, risk-bearing, and real-income distortions. The activity fixes income, thereby omitting labor supply, business production, and movement into informality.

The participant's account outcome is not social surplus either. A penalty lowers the private account and raises recorded revenue, but the social value depends on where ECU go. Fiscal return sends part back as service value, but does not model congestion, heterogeneous demand, or distribution. These omissions support an important conclusion: maximizing collections and designing a desirable tax system are different objectives.

Ask students to compare a policy that raises revenue through frequent audits with one that raises the same revenue through voluntary reporting. The accounting total is equal, but administrative cost and experienced risk differ. They must state which welfare components are absent before recommending one policy.

10.4 Administrative costs and audit selection

The platform uses an exogenous audit probability. Real authorities select returns using third-party information, cross-checks, and risk models. Targeting can raise detection per audit but creates false positives and equity concerns. Responsive enforcement approximates one transparent history-based rule without claiming to reproduce an administrative algorithm.

Assign a fixed enforcement budget and ask teams to choose among doubling random audits, raising penalties, and improving third-party reporting. A complete proposal identifies resource cost, the margin of behavior affected, and the information required. The exercise connects the game to state capacity and explains why severe statutory penalties do not substitute for credible administration.

Audit selection also changes the interpretation of observed detection. Under random audits, the audited sample represents the population in expectation. Under targeting, detection rates are selected and cannot be compared directly with random rates. This point links a simple classroom game to selection bias in administrative data.

10.5 Internal and external validity

Laboratory control makes true income, parameters, and decisions observable. In the field, true income is usually latent and detected evasion is only a subset. The laboratory removes advisors, legal complexity, filing costs, reputational consequences, and enforcement discretion. These differences limit level comparisons but do not eliminate the value of identifying mechanisms.

External validity improves through triangulation. Laboratory evidence examines causal responses under clear rules; field experiments test messages and enforcement in administrative settings; audit data reveal heterogeneity; theory organizes channels. Rigorous feedback treats the sources as complementary instead of ranking one method as universally decisive.

Students should identify a target population before extrapolating. Evidence about self-employed income need not predict wage reporting with employer withholding. A municipal property-tax message need not predict national income-tax reporting. The mechanism may travel even when the magnitude does not.

10.6 Equity, heterogeneity, and legitimacy

The same proportional rate can carry different meaning when base accounts or effort differ. The current activity holds endowments equal to isolate reporting. Future treatments could introduce earned versus windfall income, progressive rates, or unequal concealment opportunities. Each change alters perceived fairness and therefore cannot be interpreted as a purely mechanical parameter adjustment.

Legitimacy also depends on procedure: audit selection, appeal, transparency, and proportional penalties. These dimensions are not monetized in the current design. Instructors may discuss them but should not say that the session measured them. Measurement would require prespecified choices or questions and a report designed for those outcomes.

Heterogeneity in the histogram is therefore an invitation to formulate hypotheses rather than diagnose individuals. A polarized distribution could reflect different risk preferences, moral costs, or comprehension. Repeated choices help describe stability but still do not identify the source without additional variation.

10.7 Extensions not yet implemented

Third-party reporting could separate verifiable and self-reportable income. An amnesty would require past liability, a reduced surcharge, and a model of future credibility. Voting over expenditure would create a collective institution. A real-effort task could compare earned and endowed income. Joint filing would add coordination and shared liability. Tax withholding could change salience and the timing of payment.

These are not safe text-only switches. Each extension requires new formulas, comprehension checks, exported variables, and report panels. Keeping them outside the first family preserves clarity. The shared bilingual architecture makes future additions possible as documented treatments without changing stable rooms or splitting the family into unrelated apps.

10.8 Debrief rubric

A strong student statement contains four elements: a parameter-based prediction, a description with denominator, an interpretation acknowledging alternatives, and an inference limit. For example: “Because $pm < 1$, risk neutrality predicts hiding. The observed median was 70 among forty participant-period decisions from four people. Partial compliance is consistent with risk aversion or moral cost. Without randomized assignment, I cannot attribute a difference from another class to the treatment.”

This rubric prevents feedback from becoming a search for one preferred story. It also builds transferable skills: separating model and data, distinguishing a statistic from a mechanism, and specifying the design needed to discriminate among explanations.

10.9 Comparisons across classes and semesters

Instructors may wish to compare a current session with a previous semester. Such comparisons are useful for generating questions but require careful metadata. Record the configuration, participant count, language, feedback mode, date, and any interruption. A change in mean reporting can

reflect a treatment, class composition, instruction delivery, or the realized audit sequence. Preserve the full distribution and denominators rather than copying one headline number into slides.

If the goal is cumulative teaching evidence, define a common protocol before collecting sessions. Use the same wording and timing, export data after every class, and document exclusions. Avoid treating repeated classes as randomized treatment groups unless assignment actually was randomized. A simple graph by session can reveal stability and heterogeneity, while uncertainty intervals remind students that estimates vary. This practice turns the laboratory into a transparent example of replication without claiming that convenience samples represent a population.

Cross-semester comparisons also create privacy responsibilities. Aggregate across enough participants, suppress very small cells, and never combine timestamps or seating information in a way that could reidentify someone. The pedagogical value comes from patterns and institutional mechanisms, not from tracing an individual student across exercises.

Data, privacy, and troubleshooting

11.1 Recorded data

Each period records the report, hidden income, parameters, random draw, audit, reported tax, recovery, penalty, revenue, fiscal return, outcome, and decision time. Comprehension attempts diagnose instruction problems. The report never displays codes. Administrative exports contain technical identifiers and require normal laboratory data handling.

11.2 Common problems

If a link says the user does not exist, the database was reset or the session is unavailable; create a new session. If a local page stays on “Please wait”, verify the server process and port. If Demo shows other games, family filtering is wrong. If Report cannot reveal, check pending decisions. If a panel is blank, inspect its denominator.

Localhost works only on the same computer. Other computers require a reachable IP address, network rules, and a server listening on the appropriate interface. In production, use the laboratory subdomain. Do not improvise network changes during class.

11.3 Post-session checklist

Export data if the session must be retained; record configuration and size; close individual links; document problems; never reset the production database; and update code through Git. Before publishing changes, run bots, validate rooms, check both languages, and inspect hidden and revealed reports.

Checkpoint

A session is closed when every participant sees “Activity completed”, Report shows zero pending decisions, aggregates have been discussed, required data are exported, and no individual link remains projected.

References

- Allingham, M. G., and A. Sandmo (1972). “Income Tax Evasion: A Theoretical Analysis”. *Journal of Public Economics*, 1(3–4), 323–338. [https://doi.org/10.1016/0047-2727\(72\)90010-2](https://doi.org/10.1016/0047-2727(72)90010-2).
- Yitzhaki, S. (1974). “A Note on Income Tax Evasion: A Theoretical Analysis”. *Journal of Public Economics*, 3(2), 201–202. [https://doi.org/10.1016/0047-2727\(74\)90037-1](https://doi.org/10.1016/0047-2727(74)90037-1).
- Alm, J., G. H. McClelland, and W. D. Schulze (1992). “Why Do People Pay Taxes?”. *Journal of Public Economics*, 48(1), 21–38. [https://doi.org/10.1016/0047-2727\(92\)90040-M](https://doi.org/10.1016/0047-2727(92)90040-M).
- Fortin, B., G. Lacroix, and M.-C. Villeval (2007). “Tax Evasion and Social Interactions”. *Journal of Public Economics*, 91(11–12), 2089–2112. <https://doi.org/10.1016/j.jpubeco.2007.03.005>.
- Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). “Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark”. *Econometrica*, 79(3), 651–692. <https://doi.org/10.3982/ECTA9113>.
- Luttmer, E. F. P., and M. Singhal (2014). “Tax Morale”. *Journal of Economic Perspectives*, 28(4), 149–168. <https://doi.org/10.1257/jep.28.4.149>.
- Castro, L., and C. Scartascini (2015). “Tax Compliance and Enforcement in the Pampas: Evidence from a Field Experiment”. *Journal of Economic Behavior & Organization*, 116, 65–82. <https://doi.org/10.1016/j.jebo.2015.04.002>.
- Slemrod, J. (2007). “Cheating Ourselves: The Economics of Tax Evasion”. *Journal of Economic Perspectives*, 21(1), 25–48. <https://doi.org/10.1257/jep.21.1.25>.
- Slemrod, J. (2019). “Tax Compliance and Enforcement”. *Journal of Economic Literature*, 57(4), 904–954. <https://doi.org/10.1257/jel.20181437>.