universität
wien

# MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Gamification and Tax Experiments“

verfasst von / submitted by

Andreas Hausberger BSc.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2019 / Vienna 2019

| | |
|---|---|
| Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet: | A 066 840 |
| Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet: | Masterstudium Psychologie UG2002 |
| Betreut von / Supervisor: | Univ.-Prof. Dr. Erico Kirchler |
| Mitbetreut von / Co-Supervisor: | Mag. Dr. Jerome Olsen |

## Abstract

This study aims to explore the impact of game elements on participant behavior in standard tax experiments. Based on existing literature, the implementation of game elements (generally known as *Gamification*) was expected to have a positive impact on participant performance, as well as to increase tax evasion. These effects were expected to be moderated by how familiar participants are with various types of games. To control for the potential impact of visual design alone, a plain condition, a richly designed condition, and the same richly designed condition with added game elements were compared against each other. The three conditions did not cause any significant differences in either performance or tax evasion behavior. Familiarity with games was shown to have a positive effect on performance in the *Plain* condition, while having a significantly negative effect on performance in the *Gamified* condition. Possible explanations of the lack of differences between conditions are the limited effects of short-term Gamification as opposed to long-term Gamification and the implementation of unenjoyable game mechanics.


*Keywords*: Gamification, tax experiments, tax evasion

## 1. Introduction

Research about tax behavior has increased significantly over the last several decades. As demonstrated by Kirchler (2007), the amount of research containing key-words such as “tax”, “taxes”, “taxation” and “evasion” has grown exponentially since 1980. While the field of tax behavior research appeared to be dominated by an economic approach rather than a psychological one at the time (Kirchler, 2007), there have been several landmark studies that investigate the impact of psychological aspects on tax behavior.

Psychological tax research has, among other topics, addressed the impact of financial knowledge on tax behavior, or the impact of social and societal norms (Kirchler, 2007). Many of these studies have relied on self-report questionnaires.

For instance, Eriksen and Fallan (1996) reported that an increase in financial knowledge was followed by an increase in perceived fairness of taxation, and a stricter attitude towards tax evasion. In that study, a self-report questionnaire was used to collect data about the participants’ attitudes.

Using self-report questionnaires is a valid method of collecting data about attitudes, but  it has obvious limits: The response patterns are likely influenced by personal and social norms (Wenzel, 2004), and the correlation between self-reported attitude and actual behavior is spurious at best (Hessing, Elffers, & Weigel, 1988). Furthermore, it is generally argued that questionnaire responses are prone to social desirability biases. This could result in much more positive tax attitude measures than people truly hold. (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003).

It stands to reason that an experimental approach which more closely simulates a tax paying situation is more advantageous. Such an approach could potentially yield more information about actual tax behavior. Tax experiments are characterised by a controlled (laboratory) environment and clearly defined ruleset for participants. Their key strengths are high reliability, internal validity and the ability for researchers to systematically vary relevant factors. However, some criticism has also been raised, namely, their assumed lack of realism and low external validity (e.g., Falk & Heckman, 2009).

To address the low external validity,  context-rich items are used. Context-rich questions are worded in a way that relate to the participants in a meaningful way (e.g., to their real life situation or their emotional state). The impact of these context-rich questions is not uniform. In some studies, a positive impact on tax compliance is found when context-rich questions were used (Choo, Fonseca, & Miles, 2014), while others found no such effect (Alm, McClelland, & Schulze, 1992).

One possible method of increasing context-richness (and therefore, possibly, external validity) is the use of game elements. The use of game elements in non-game situations is generally known as Gamification, and is explored further in 1.3. While different types of experimental set-ups have been tested and compared, the possibility of using gamified materials has not yet been tested. While tax experiments have been characterized as "gambling-like" (e.g., Kirchler & Muehlbacher, 2016), it is not yet tested how the behavior of participants is impacted when actual game elements are used.

## 1.1 Effort-Based Tax Experiments

### 1.1.1 Utilization and design.

An experiment that simulates a tax paying situation usually consists of two phases. In an *earnings* phase, participants are given the chance to earn an amount of currency. After that, they should declare what they earned, and pay a fraction of their earnings as a "tax". The latter is called the *declarations* phase.  The *declarations* phase is essentially a choice under uncertainty. Participants must decide how much of their earnings they report without knowing whether they will be audited. Participants may opt to “play it safe”, and correctly report their earnings to avoid a penalty. Alternatively, participants can choose a “risky gamble”, and under-report their earnings to save taxes. The latter strategy would be penalized in case of an audit.

For the *earnings* phase, effort-based tasks seem more preferable than knowledge-based tasks or chance-based ones. For one, a well-designed effort-based task is easy to learn and does not require any expert knowledge. Additionally, there should not be an element of chance within the task. Ideally, participants should feel that the locus of control is entirely within themselves (see also the classification introduced by Roth, 1995 (as cited in Mascagni, 2017)).

Gill and Prowse (2012) designed a computerized effort-based task. In their task, participants would manipulate sliders to set them to a specified value. Participants would repeat this with multiple sliders, often under a time constraint. For every slider that had been set to the correct value, an amount of currency was paid. For an example of a slider that can be displayed in any web browser, see Figure 1. Generally, these sets of slider tasks are repeated over several rounds, depending on the specific design of the experiment. Also, the conditions of the *declarations* phase may be varied over the course of several rounds.

This design satisfies the criteria laid out above: It is easy to learn, and there is an obvious connection between the participants’ performance and their reward. Due to their ease of use and simple implementation, slider tasks have become a staple of tax experiments. Several studies use slider task rounds for their “earnings” phases (e.g., Kogler, Mittone, & Kirchler, 2016).

*Figure 1 - A HTML slider*

### 1.1.2 Monetary incentives.

Due to their connection with economics and real-world tax behavior, the question of decision-contingent monetary incentivization is important. Besides a base reward for participation (as is common in many psychological experiments), some tax researchers opt for linking in-experiment currency with real money. For instance, in a recent study by Chan (2018), participants received a base participation fee of $1.50, and additionally were eligible for a bonus based on their performance in slider tasks.

The role of monetary incentives, however, is not without controversy. There have been questions as to how the use of monetary incentives could influence participants' effort-task performances, their behavior, and their tax compliance.

Hertwig and Ortmann (2001) saw some performance increases in experiments with monetary incentives, but could not find a uniform effect. For decision-making tasks, which would encompass effort-based tax experiments, monetary incentivization is correlated with higher participant performance and reduced variance in the data.

De Araujo et al. (2015) found that when increasing the per-slider reward from $0.005 to $0.08, participants' performances in each round increases by roughly four percent. The researchers of this study caution however, that the increase was not uniform among participants. Also, there was no comparison with performance in an unincentivized slider task.

Among researchers, the notion of using monetary incentivization to increase external validity is common (e.g. Loewenstein, 1999). Using real money in an experiment about tax behavior is often thought to be more resembling of a real-world tax declaring situation. Madsen and Stenheim (2015) however argue that monetary rewards could also lower the external validity. According to them, presenting monetary rewards as a direct result of participants' behavior or performance drives the experience in the experiment farther from its real-world counterpart.

#### 1.1.3 Monetary incentives and replication.

Another potentially problematic aspect of monetarily incentivized studies are their financial costs and the ramifications for replication. In the aforementioned study by Kogler, Mittone, and Kirchler (2016), roughly 1460€ were spent on incentivizing a sample of n=126. In this study, participants could earn up to 15€ depending on their tax compliance. A mean amount of 11.57€ was paid per person. Obviously, any attempt to replicate such a study with a larger sample would cost an amount of money that often is prohibitive for researchers without reliance on grants or external financial aid.

Given the financial constraints of a monetarily incentivised study, other means of incentivization are needed. One possible solution is offered by the design pattern of Gamification.

### 1.2 Gamification

Gamification is a term in the field of user interaction design. It is defined as implementing game elements (or game mechanics) in non-game situations (Prince, 2013; Huotari & Hamari, 2012).

While there have been several marketing strategies in the past that arguably use game mechanics (e.g., collection programs such as store loyalty cards or frequent-flyer-programs (Prince, 2013)), current use of the term Gamification almost always describes specific design of software-based platforms.

Game elements are often implemented to motivate users to behave in a specific way: Several examples show the widespread use of Gamification.

For instance, participants in gamified trading applications were shown to be more active over a longer period of time (Hamari, 2015). Another example is the social network *Foursquare*, which is described a “location-based social network” (Rimon, 2014) uses Gamification as a central feature. It rewards users for “checking in” at certain locations (using a mobile app) with points, achievements and a level system. Yet another is the fitness platform Runtastic which allows its users to use a leaderboard for competition (Runtastic Team, 2015). Within a private group (whose members regularly use Runtastic applications for tracking their activities), users can compare their sport data among each other. This feature is advertised as a way to “boost your motivation and push your limits” (Runtastic Team, 2015).

In a paper by Basten (2017), distinctions are made between three aspects of Gamification design: Mechanics (or game elements), dynamics (reactions to participant behavior) and aesthetics (emotional responses to the application). In sections 1.2.1 and following, these aspects will be explored further.

#### 1.2.1 Mechanics.

Mechanics are software features that are used to make an application more game-like (Basten, 2017). They can be added to an existing system without fundamental redesign.

Table 1 lists some of the most common game elements that are used in Gamification. There is no clear consensus about how many of these elements are sufficient for an application or system to be called “gamified”.

*Table 1 - Common game elements*

| Name | Definition | Real-world example | Goal |
|---|---|---|---|
| Currency | Rewarding a certain behavior with a virtual currency, points or “Experience”. | *Todoist* (To-Do-Lists and Time-Tracking): Completing tasks within their deadlines is rewarded with Experience Points. | Sense of progression, rewarding certain behaviors. |
| Levels | Assigning ranks and levels to users; giving users the chance to level up over time (often in combination with an Experience-system). | *Todoist:* Users are assgined levels, depending on their experience points. | Sense of progression. |
| Badges | Giving users badges (or “achievements”) for certain accomplishments. These are displayed prominently and / or can be revisited at a later time. | *Apple Watch Activitiy*: Users who reach their activity goals often enough in a time period are awarded virtual medals. | Sense of accomplishment, collection of prizes. |
| Leader-board | Allowing users to enter their score to a public board, displaying their accomplishments to others | *Runtastic:* A group of users who are connected via a mobile app can compare their activity data, and are ranked in a leader-board based their runs. | Sense of competition. |
| Narrative, Lore and Story | Introducing a fictional setting in which user interaction has meaningful consequences. Allowing users to experience and/or influence stories in a fictional world | *Human Resource Machine*: In an educational game which teaches its players Assembly programming languages, players experience a story about a corporate conspiracy told through videos and dialogue. | Motivation to improve; motivation to keep engaging with the application. |

### 1.2.3 Dynamics.

Dynamics are the “runtime behavior of mechanics concerning players’ inputs and outputs over time” (Basten, 2017). This includes all forms of feedback to the participants, including notifications, pop-ups, progress indicators and so forth.

One important property of application dynamics is their instant nature. For any meaningful action, participants should receive feedback immediately (e.g., be shown a visual cue that their action was registered and is meaningful).

#### 1.2.4 Aesthetics.

Aesthetics are the “desirable emotional responses evoked in users when they interact with the gamified system” (Basten, 2017). A game-like optical design and game-like features (such as collecting items, using playing cards or controlling a virtual character) are part of the aesthetics aspect of gamification.

#### 1.2.5 Gamification as a means to increase motivation.

A common thread among all attempts to *gamify* a serious application is the increased focus on intrinsic motivation. While playing games is obviously facilitated mostly by intrinsic motivation (Basten, 2017), the use of non-game applications is motivated by extrinsic factors (e.g., the user is working towards a mandated or externally defined goal).

Platforms which expect their users to have difficulties with sustained motivation (e.g. platforms dedicated to activity tracking or learning) can incorporate gamification elements to motivate their users (Basten, 2017).

Hamari, Koivisto and Sarsa (2014) write that Gamification provides motivational affordances, which influence psychological outcomes, and therefore in turn behavioral outcomes.

In a recent study, a gamified student quiz application was evaluated by its participants, who reported a positive impact on their enjoyment and motivation (Cheong, Cheong, & Filippou (2013)). However, the study did not compare gamified and non-gamified version of the app. Therefore, no conclusions on a possible effect on student performance was possible.

### 1.2.6 Possible downsides of Gamification.

However, there are also several potential pitfalls to the use of Gamification (Deterding, 2010). Many attempts at *gameful design* are suboptimal, and may lead to negative outcomes.

Some findings of positive effects of the implementation of Gamification may be explained by a novelty effect. For example, while the online platform *Foursquare* was one of the initial proponents of Gamification and *gameful design*, few long-term effects regarding regular user visits were measured (Forrester, 2010, after Detering, 2010).

Also, Ecker, Slawik, and Broy (2010) warn of unintended side-effects that arise due to the implementation of Gamification. In their study, participants drove cars with a gamified dashboard user interface that would encourage fuel efficiency. In practice however, some participants tended to drive more recklessly (for example, ignoring red traffic lights, or not halting at stop signs) to save fuel. This is an example of unforeseen (and undesirable) behavior as a result of Gamification.

## 1.3 The Gamification of a Tax Experiment

This study aims to explore how a tax experiment can be gamified, using the methods and ideas that were introduced in 1.2.

Experiments in the field of Economic Psychology appear as potential targets for gamification: They often involve decision making tasks and simulated risk (Mascagni, 2017). Also, recruiting enough participants who are willing to partake in an experiment while not receiving any monetary rewards is a common problem in academic psychological research. One key opportunity, however, would be to use Gamification as a means to increase the external validity of a tax experiment. In the following sections, several game elements are introduced that can be used to *gamify* a tax experiment.

#### 1.3.1 Virtual currency.

The first, and perhaps most obvious feature is the introduction of a virtual currency as a replacement for monetary incentives. The currency can be earned via the completion of effort tasks, and is dependent on the participants' performances in these tasks. Using points as a measure of task completion and general performance has been used in tax experiments before (Gill & Prowse, 2012).

To make the situation more game-like, earning points should elicit a feeling of gaining something valuable. This can be achieved by using language and presentation that would evoke an association with real money (e.g., the use of a currency symbol, words like “virtual currency” and “earn”).

#### 1.3.2 Achievement badges.

Achievement Badges (also “achievements” or “trophies”) are defined as optional sub-goals in a secondary reward system (Montola et al., 2009). In games, Achievement Badges are given to players who accomplish goals (or show certain behaviors) that are optional or not necessary to complete the game. For example, achievement badges may be given to players who collect a number of hidden items. On the other hand, a subset of  achievement badges are given for accomplishing the “main goals” of a game. This is used to show players that there are achievement badges available. Also, giving out a subset of badges for completing “main goals” may motivate players to collect the optional badges as well (Hamari & Eranti, 2011).

In a tax experiment with multiple effort-task rounds, both kinds of achievement badges can be implemented easily. “Progress” badges can indicate how many rounds the participants have completed, thus giving overview of their total progress. “Optional” badges can be given to participants who show certain types of behaviors (e.g., a badge that is given to players who never under-report their earnings).

A great deal of thought is given to the optical placement of achievement badges. In games, achievement badges are announced with a notification pop-up, which often contains a unique icon, a distinct name (sometimes containing a pun or an inside joke) and a description of why the achievement badge was earned (Hamari & Eranti, 2011).

While players often engage with games for several hours over multiple days, a tax experiment is usually completed within a single hour. Therefore, a salient way to display achievement badges is needed.

To prevent badges from influencing participant behavior, the badges should not be explained beforehand.

Displaying the achievement badge collection in a fixed frame (that follows the participants' viewports as they scroll) with empty "sockets" for badges that are still available should be effective: The participants are aware of available badges, and are possibly motivated to fill the empty spots in their collection. When a new badge is earned, a salient notification is displayed in a corner of the screen.

#### 1.3.3 Leaderboard.

Leaderboards are lists which display a participant's chosen nickname and his/her final score after playing a game. They show a ranking of the "best" players who chose to enter their score (Symonds, 2010). Historically, leaderboards were featured heavily in arcade game systems. Recently, these leaderboards are used for online comparisons between players (Symonds, 2010).

Again, the implementation of a leaderboard in a tax experiment is fairly obvious: After completing the earning and taxation parts of the study, participants are asked whether they want to enter their score into a leaderboard with a chosen nickname, thus protecting their anonymity.

### 1.3.4 Implementation of other game elements.

Depending on the experiment, other game elements are also possible. Some experiments generate a random number to simulate the probability of a tax audit (for an example, see Chan, 2018). This mechanic can also be presented in a game-like manner, such as throwing a die or drawing a card.

For experiments that are comprised of multiple sessions over a longer timespan, the implementation of long-term gamification elements are also possible.

### 1.3.5 Central questions.

By now it is apparent that applying the principles of gamification to a tax experiment is as promising as it is difficult. On the one hand, the use of gamification in several applications and services has shown an increase in participant interactions (Basten, 2017). On the other hand, the methods laid out above may appear daunting to implement. To investigate the effects of gamification in a tax experiment, and to distinguish the effects of gamification from those of the visual design itself, this study aims to examine four distinct hypotheses:

1. **Performance**: A gamified design will result in better slider task performance and a steeper performance increase over the course of several rounds.
2. **Tax Evasion:** A gamified design will result in greater tax evasion due to higher risk-taking behavior.
3. **Design:** Design itself will not be a significant factor. There will be no difference between two designs with identical content which only differ in their visual design.

4. **Familiarity:** The effects in hypotheses 1 and 2 are moderated by the participants' familiarity with games, since people who enjoy playing games are more likely to be motivated by game elements.

Furthermore, the study explores whether there are differences in the experienced degree of realism between the three conditions.

# 2. Method

## 2.1 Participants

### 2.1.1 Demographics.

A sample with the total size of $N$ = 175 was recruited. 72% of the participants were women. The sample had an age range from 17 to 35 years, with a median of 20 years, and a mean of 21.15 years. The majority of participants was not actively employed (~ 43%) or worked ten or fewer hours per week (*Geringfügigkeit*, ~31%). The sample is reflective of the general student population at the University's bachelor program.

## 2.2 Materials

### 2.2.1 Experimental conditions.

To investigate the proposed hypotheses, a trimodal interactive application was developed. Participants were presented the application in an opened web browser in the laboratory. Every participant was randomly assigned to one of three conditions:

1. Mode *Plain*

Participants in the *Plain* condition completed an experimental questionnaire without much thought to visual design. It used a plain grey background, a standard system font, and standard HTML elements. In the *Plain* condition, no game elements were implemented.

2. Mode *Rich*

Participants in the *Rich* condition were presented an experimental questionnaire with modern design. It used a custom font (*Rubik* - a modern, open source sans-serif font [1]), a slight gradient background, and redesigned HTML elements. Also, no game elements were implemented.

3. Mode *Gamified*

The *Gamified* condition used the same design elements as the *Rich* condition. In addition, several game elements were implemented. See Figures 2a-d for screenshot comparisons between the different versions.

#### 2.2.2 Content of experimental questionnaire.

Tutorial: In the tutorial, participants read explanations of the tasks, and were instructed to try out the slider tasks (both with and without a time limit). Also, participants completed their first tax declaration item. The *Gamified* condition also featured explanations about the game elements, namely the virtual currency, the achievement badges and the leaderboard.

Demographics: In this short section, five demographic details were asked. Participants were asked to enter their ages, genders, nationalities, as well as their education and employment status.

Slider Tasks (*Earnings* and *Declarations*): The main portion of the experimental questionnaire was taken up by ten rounds of slider tasks (the *earnings* phase), each followed by a tax declaration (*declarations* phase). Each slider task consisted of ten items: The slider itself and a short explanation text. Within the explanation text, the current slider goal was stated (e.g., “Put the slider precisely at the 65% mark”).

[1] https://fonts.google.com/specimen/Rubik

*Figure 2a - Slider Tasks in the* Plain *condition*

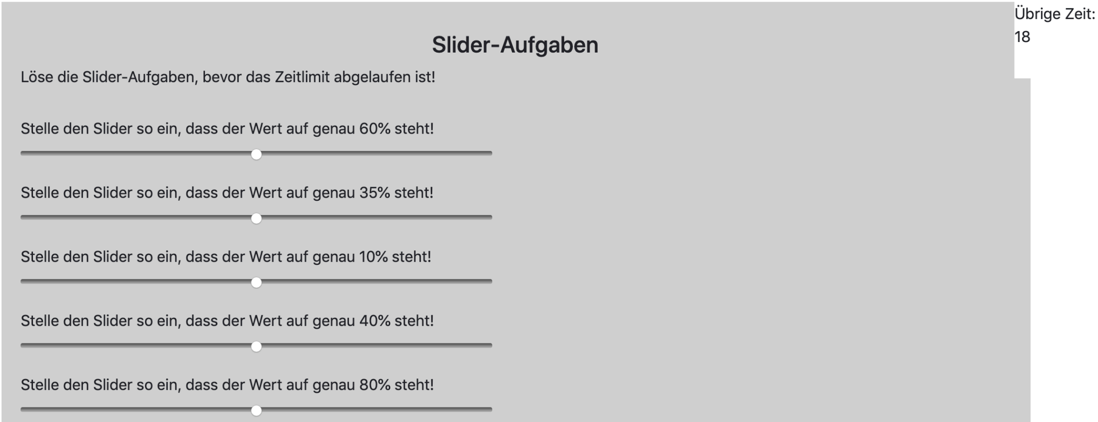



*Figure 2b - Slider Tasks in the* Rich *and* Gamified *conditions*

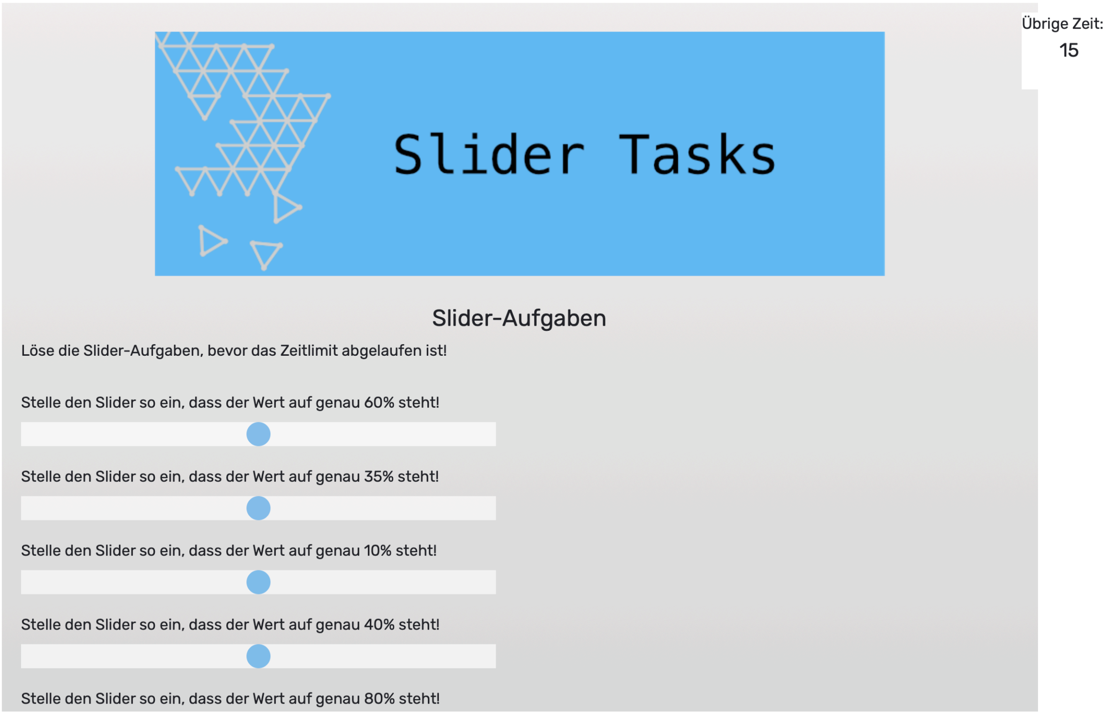



*Figure 2c - Tax declaration in the* Plain *condition*

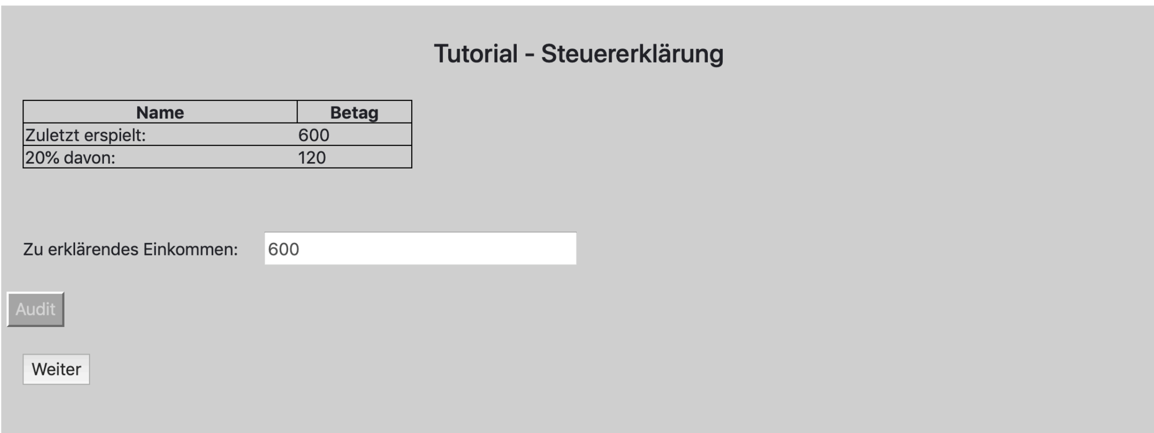
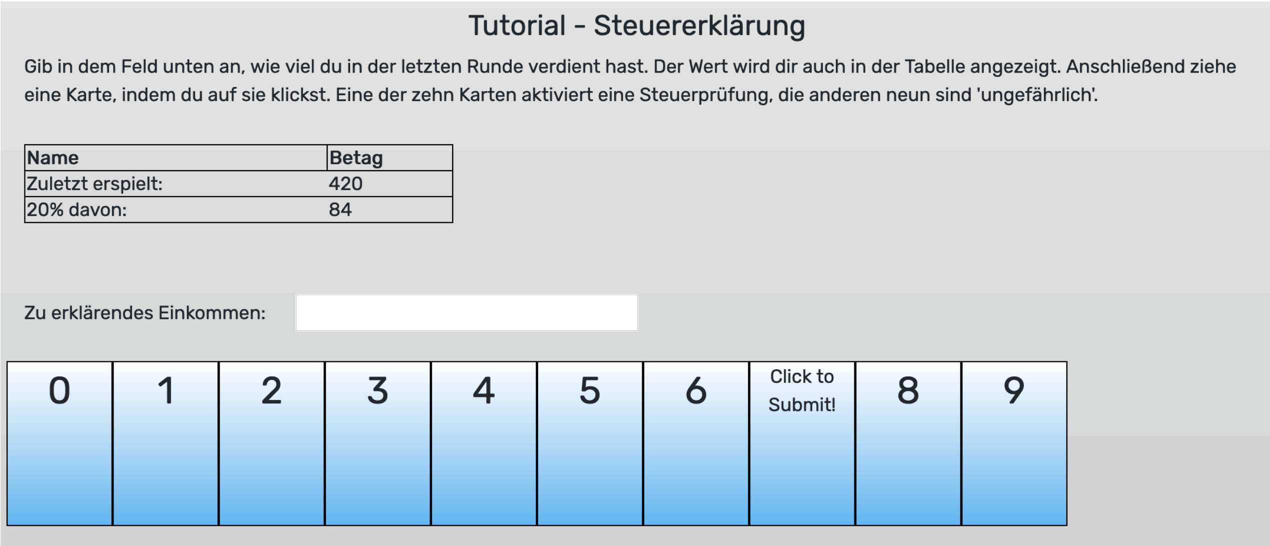


*Figure 2d - Tax declaration in the* Gamified *condition (cards turn when hovered over with the cursor)*




On releasing the mouse cursor from the slider, the current position was displayed next to the item.

For each slider, participants could earn up to 200 points / “EcoBucks”. If the slider is put within 10% over or under the correct mark, a percentage of the reward is paid. In order to create a challenging task, each round had a time limit of 25 seconds. A countdown element was clearly displayed during each slider task. When the time limit was up, all sliders were deactivated and participants were no longer able to manipulate them.

Each tax declaration task displayed clearly how much participants had earned in the previous round. Participants then were asked to enter the sum they wish to declare, thus giving them the opportunity to be honest with their earnings, or to avoid some tax. Each round income was taxed with 20%, and in each round, there was a 10% chance of an audit. In case of an audit, participants who avoided taxes (by under-reporting their earnings) would have to pay the remainder of the taxes they owed, as well as the whole tax amount as an additional fine. After finishing all ten slider and tax declaration rounds, some participants (those who are in the *Gamified* condition) could enter their score in the leaderboard.

Familiarity With Games: To investigate how familiar participants are with games, five questions were asked. Participants were asked how often they played computer or video games (including games on their phones), board games, puzzles or riddles (e.g. Sudoku or crossword puzzles), as well as how often they participated in competitive sport actively or as a spectator.

Participants chose from a dropdown menu with the following options: “(Almost) Daily”, “2-5 times per week”, “1-4 times per month”, “Fewer than that”.

Attitude Towards Taxation: To allow better comparisons with other tax experiments, three self-reported questions were included. These questions included two items about the perceived realism of the experiment and one item about the perceived fairness of evading taxes.

Design Evaluation: In this section, participants should state how interesting and visually appealing they found the experiment. The questions asked whether the tutorial was informative, whether the slider tasks were interesting or boring, and whether they would rate the visual design of the web application as modern and beautiful or outdated and ugly. Participants answered these items on a seven point Likert scale.

In addition, some questions about potential technical issues were included. These questions were included for software quality assessment, and did not necessarily yield useful information for the study itself.

Debriefing: After concluding the experiment, participants reached the final page of the application, which featured further information about the study. Participants were debriefed and could read explanations about the study’s goals, its hypotheses, information about the three modes, as well as the potential badges of the *Gamified* condition.

#### 2.2.3 Game elements.

In the *Gamified* condition, several game elements are implemented. These were explained to the participants in the tutorial section.

Virtual Currency: Instead of “generic” points, participants earned a virtual currency named “EcoBucks”. In all relevant instructions, they were encouraged to think about their earned units as money (for example, the word “Account balance” was used, instead of “Score”).

Badges: Participants of the *Gamified* condition earned badges as they completed the experiment. Both “progress” badges (which participants earned simply by completing the experiment) and optional ones were included (e.g., one badge would be rewarded to participants who earned more than 50% of the maximal value). See Table 2 for a detailed overview of the badges.

Leaderboard: After completing all slider tasks, participants in the *Gamified* condition were able to enter their score in a leaderboard. They were instructed to use a nickname. On the same screen participants would enter their nickname, they saw the scores of other players in a descending list. Leaderboard entry was optional.

Gameplay Elements: Tax declarations in the *Gamified* condition featured ten virtual “cards” numbered from 0 to 9. Participants were told to “draw a card” by clicking on it. To represent a 10% chance of an audit, participants were told that one of the ten cards would trigger an audit, while the other nine were “harmless”.

*Table 2 - Badges*

| Badge | Category | Challenge | Image |
|---|---|---|---|
| Start | Progress | Finish the tutorial. |  |
| Money | Optional | Earn more than 10.000 EcoBucks. |  |
| Honesty[a] | Optional | Always fully report your taxes. |  |
| Evasion[a] | Optional | Never get caught evading taxes. |  |
| Completion | Progress | Multiple Badges: Awarded when completed 20%, 40%, 60%, 80% and 100% of all slider tasks. |   |

[a] To avoid influencing participants' behavior, no information about the existence of these badges was given until the slider tasks were completed.

#### 2.2.4 Release of application.

To facilitate replication, comparability of results and provide a robust framework for creating gamified experimental questionnaires, the software used for this experiment is available under an open source license. See the appendix for further information.

### 2.3 Procedure

#### 2.3.1 Technical implementation.

The experiment application used Node.js for backend organisation, and React.js for the frontend display. The application was hosted on an external server (Heroku platform), which also served as a temporary data storage.

#### 2.3.2 Recruitment.

Participants were recruited among undergraduate psychology students at the University of Vienna. The participants signed up on their own accord, and no invitations were sent out. Participants were notified by the university's LABS system, in which they received course credits for participating. To comply with the university's rules regarding said LABS system (and the goals of the study), no financial incentives of any kind were offered to the participants.

#### 2.3.3 Conducting the experiment.

As mentioned in 2.2, participants were presented the application in an opened web browser on a desktop computer. They were given a short verbal instruction, which would give them an idea about the experiment (without mentioning the actual research questions). After the introduction, participants were instructed to start the experiment. A researcher was present at all times to help participants with any potential questions.

The experiment could be conducted without major problems. Few computers would randomly shut down and restart. In these cases, the participants were instructed to start over the experiment.

## 3. Results

### 3.1 Testing the Hypotheses

#### 3.1.1 Gamification and overall performance.

The first hypothesis states that participants in the *Gamification* condition would have a better performance in the slider tasks generally, and a steeper increase over the course of ten rounds specifically.

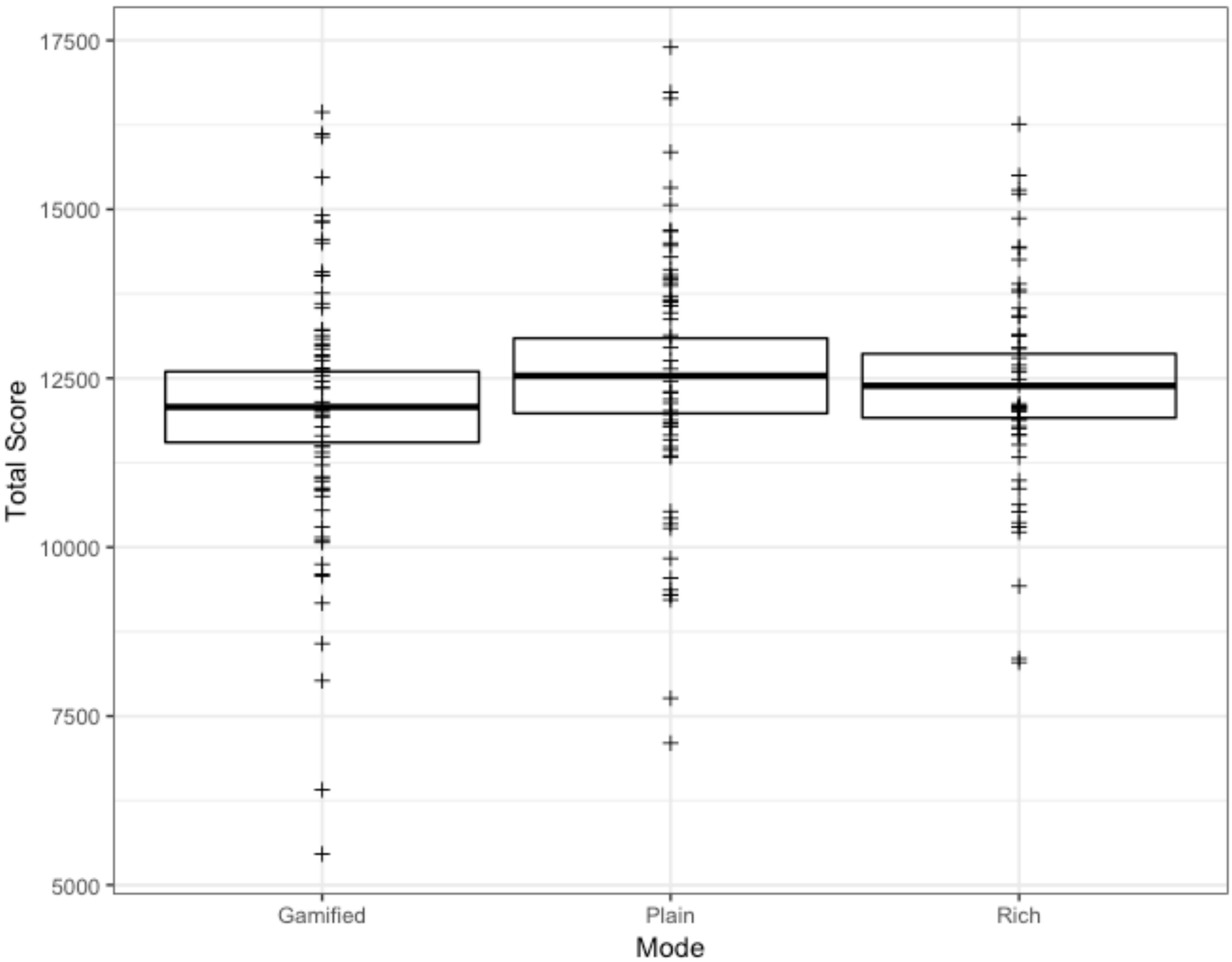
As for the first part of the hypothesis, there is no significant difference between participants in any of the three conditions. The total scores were homogenous in variance according to a Levene's test, $F(2, 172) = 1.61$, $p = .204$. The total scores of all three conditions were compared with a one-way ANOVA, $F(2, 172) = 0.855$, $p = .427$. See also Table 3, Figure 3.

*Table 3 - Slider Task Overall Performance*

| Condition | Plain | Rich | Gamified |
|---|---|---|---|
| Group Mean | 12536.05 | 12387.53 | 12074.97 |
| Standard Deviation | 2131.3 | 1685.52 | 2119.11 |
| Highest | 17400 | 16260 | 16438 |
| Lowest | 7100 | 8294 | 5459 |

*Note:* The “Group mean” is the mean value of the participants’ total scores, seperated by experiment condition. (n = 175).

*Figure 3 - Slider Task Overall Performance Point Diagram (with mean score & 95% confidence intervals for mean score)*



#### 3.1.2 Gamification and slider task progression.

To determine whether participants of different groups showed differences in their respective progression patterns over multiple rounds, the round mean scores were determined for each group.

The first hypothesis states that participants in the Gamification condition would have a steeper performance increase than the other groups due to a more motivating design.

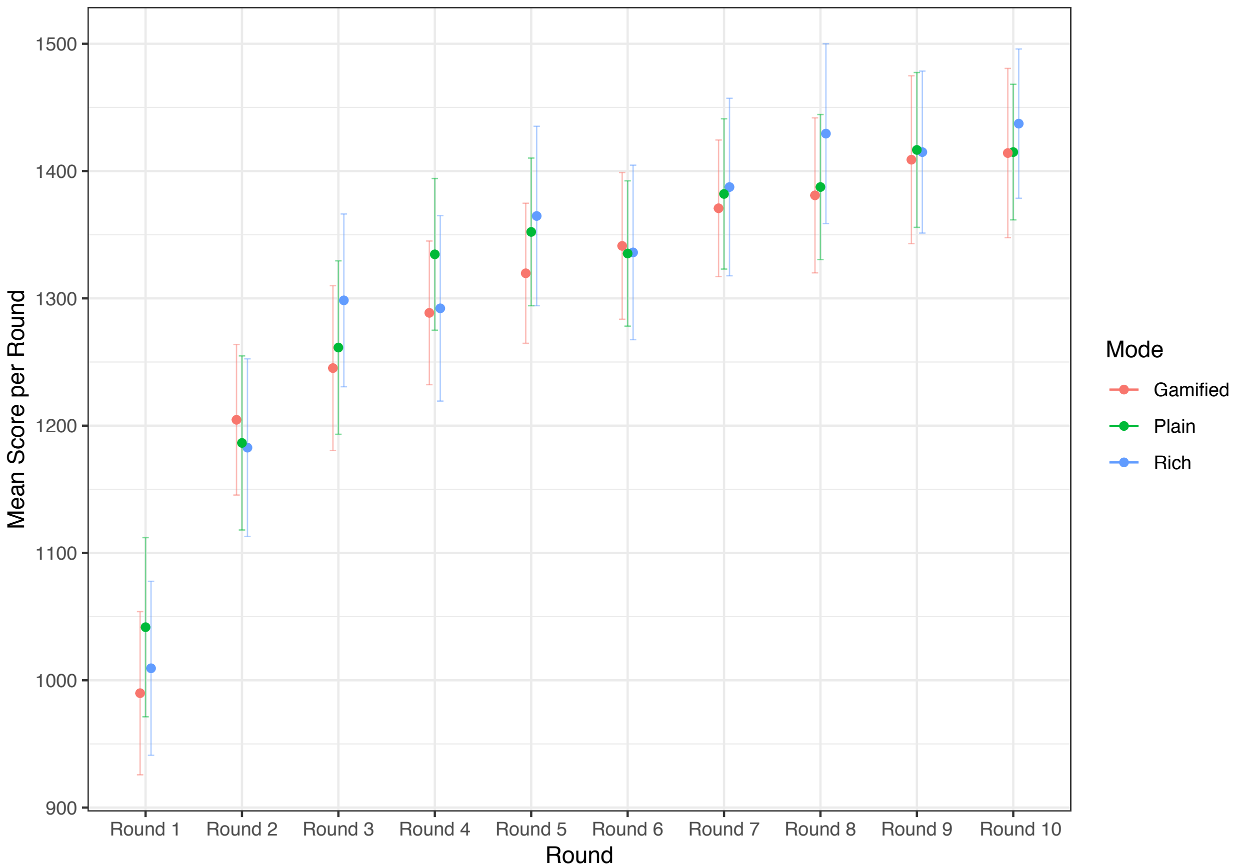
Figure 4 shows the mean scores for each group over the course of ten rounds. It is apparent that whatever differences exist, they are unlikely to be significant. See also Table 4 for an overview of the mean scores per round.

*Table 4 - Mean scores per round and ANOVA results*

| Round | Mean(SD) | *F*-value | *p*-value |
|---|---|---|---|
| 1 | 1013(263.01) | 0.605 | .55 |
| 2 | 1192(253.71) | 0.127 | .88 |
| 3 | 1266(260.54) | 0.608 | .55 |
| 4 | 1305(242.27) | 0.657 | .52 |
| 5 | 1344(235.31) | 0.577 | .56 |
| 6 | 1338(235.14) | 0.012 | .99 |
| 7 | 1379(233.11) | 0.078 | .93 |
| 8 | 1397(243.17) | 0.638 | .53 |
| 9 | 1413(247.89) | 0.016 | 0.98 |
| 10 | 1421(235.03) | 0.168 | 0.85 |

*Note*: n = 175.

*Figure 4 - Mean Score per Round, with Confidence Intervals.*




### 3.1.3 Gamification and tax evasion.

The second hypothesis states that participants in the *Gamification* condition were more likely to evade taxes. This was due to the fact that playing games is associated with making choices in risk-reward situations.

To explore this assumption, the amounts of evaded income over the course of ten rounds were added up. 13 participants who over-declared their income were excluded, since they may not have understood the instructions properly.

To exclude the factor of performance, relative compliance scores with values between 0 and 1 were calculated (relative compliance scores are the quotient between the mean actual income and mean declared income for each condition in each round).

The relative compliance scores in all three conditions are homogenous in variance (Levene's Test, $F(2, 165) = 2.38$, $p = .096$).

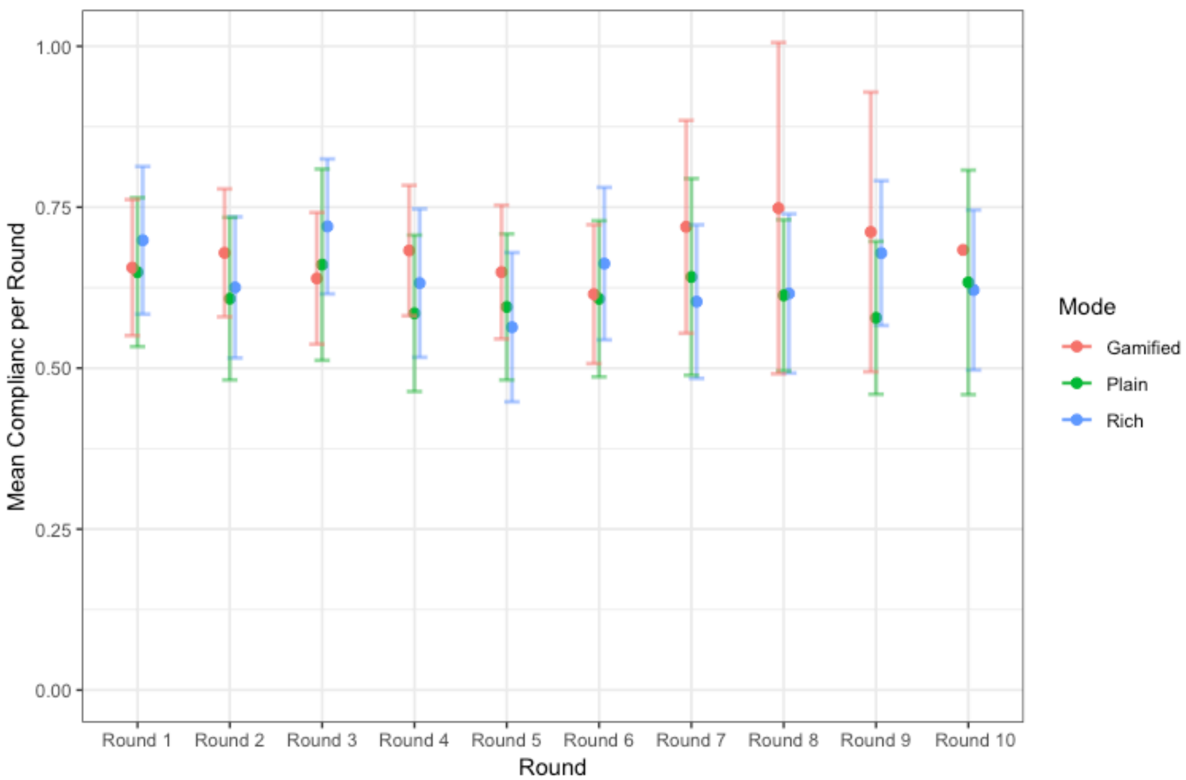
A one-way ANOVA found no significant differences in the relative compliance scores between the three conditions, $F(2, 165) = 1.71$, $p = .183$.

When looking at the progression of tax evasion over the course of the ten rounds, no interpretable effect is visible. While participants in the *Gamified* condition had on average a slight tendency towards tax compliance, there is no overall trend detectable. See also Figure 5.

#### 3.1.4 The evaluation of design.

To assess the perceived impact of the experiement's visual design, participants reacted to five items on a seven-point Likert scale. The items were as follows (translated from German):

*Figure 5 - Relative Compliance over 10 Rounds with Confidence Intervals*



- *Item 1: The tutorial explained the rules and tasks well*
- *Item 2: The slider-and-tax declaration tasks were interesting and/or entertaining*
- *Item 3: The slider-and-tax declaration tasks were boring*
- *Item 4: I think the visual design of the experiment is beautiful and/or modern*
- *Item 5: I think the visual design of the experiment is ugly and/or outdated*

Overall, the *Rich* and *Gamified* conditions were rated higher than the *Plain* condition, both in visual appeal and their interestingness. For the first item, a Welch's test was conducted, since Levene's test revealed significant differences in variance between the groups. For items 2-5, a standard ANOVA was conducted. See Table 5 for an overview of the respective test results.

*Table 5 - Results of ANOVA of Design Evaluation Items*

| Item | M(SD) *Plain* | M(SD) *Rich* | M(SD) *Gamified* | p-Value |
|---|---|---|---|---|
| 1 | 4.46(1.52) | 5.29(0.99) | 5.12(0.96) | 0.002** |
| 2 | 2.64(1.65) | 3.96(1.55) | 4.26(1.36) | 0.08 |
| 3 | 2.29(1.64) | 1.96(1.55) | 1.74(1.64) | 0.144 |
| 4 | 2.90(1.41) | 3.41(1.56) | 3.60(1.39) | 0.02* |
| 5 | 2.61(1.47) | 2.02(1.59) | 2.08(1.51) | 0.07 |

*Note:* The Likert scale items were coded as values from 0 to 6. n = 175.

The differences in design evaluation suggest that the different conditions were salient for participants. Hypothesis 3 stated that "the design itself (would) not be a significant factor" (see 1.3.5). Due to the lack of significant differences in either performance (see 3.1.1, 3.1.2) or tax compliance (see 3.1.3), it can be assumed that the design differenes did not have a significant effect on participant behavior.

### 3.1.5 Familiarity with games.

Participants chose their answers for the “Familiarity with Games” section from a dropdown menu (see 2.2.3 for item descriptions). Since all items were mandatory to complete the study, no values were excluded. To assess how the results in this section are correlated with the overall performance, the answers were dummy-coded categorically as 0, 1, 2, and 3, relative to their frequency from lowest to highest. The global *Familiarity* score is the sum of said variables. See Table 6 for an overview of the results.

*Table 6 - Familiarity with Games (Results)*

| Category | (Almost) Daily | 2-5 Times / Week | 1-4 Times / Month | Fewer than that |
|---|---|---|---|---|
| Video Games | 15 | 22 | 33 | 104 |
| Tabletop Games | 1 | 12 | 106 | 56 |
| Puzzle Games | 6 | 10 | 51 | 107 |
| Active Sports | 4 | 14 | 27 | 129 |
| Passive Sports | 0 | 0 | 17 | 158 |

*Note:* n = 175.

Compared with all conditions simultaneously, only two small correlations are found: *Familiarity with Video Games* (Pearson’s $r$ = .17) and *Familiarity with Active Sports* (Pearson’s $r$ = .11) are positively correlated with the overall score.

The correlation between *Familiarity With Video Games* and the overall score is bigger in the *Plain* ($r$ = .36) and *Rich* ($r$ = .27) conditions, while there is no correlation (r = -.07) in the *Gamification* condition.

For a complete overview of the correlations between the *Familiarity* scores and the overall scores, see Table 7. Due to the overall low scores in *Familiarity with Passive Sports*, results of that item are excluded to increase readability.

*Table 7 - Correlations between Familiarity and overall score.*

| Condition | Video Games | Tabletop Games | Puzzle Games | Active Sports |
|---|---|---|---|---|
| Overall | 0.17 | -0.01 | 0.06 | 0.11 |
| Plain | 0.36 | 0.17 | 0.26 | 0.22 |
| Rich | 0.27 | 0.12 | 0.20 | -0.11 |
| Gamification | -0.07 | -0.18 | -0.21 | 0.15 |

*Note:* Pearson's *r* was calculated for all correlations. n = 175.

To explore whether *Familiarity with Games* has a moderating effect on both the effects in hypotheses 1 and 2 (see 1.3.5), several moderation analyses were conducted.

For the analysis of moderation of *Familiarity* on the effect of Gamification on overall performance, a total *Familiarity* score was calculated and centered. Furthermore, dummy variables were used for the condition variable with the *plain* condition as reference category.

The results seem to concur with the findings above: *Familiarity* has a significant moderating influence on the effect of the condition on the performance, just not in the direction that was originally hypothesised.

In the *Plain* condition, *Familiarity* is shown to have a significant positive effect on performance, while in the *Gamified* condition, a significant negative effect is visible. In the *Rich* condition, no effect is interpretable.

Regarding tax compliance, *Familiarity* did not have a significant moderating effect. See Figures 6 and 7 for a graphical representation, and Table 8 for an overview of regression scores.

*Table 8 - Overview of Regression results (Familiarity on performance, compliance)*

| | Overall Performance[a] | | |
|---|---|---|---|
| | B | β | p |
| Intercept | 12461.61 | | < .001 *** |
| Mode *Rich* | -64.55 | -0.17 | .86 |
| Mode *Gamified* | -396.25 | -1.13 | .26 |
| Familiarity | 504.03 | 3.46 | < .001 *** |
| Mode *Rich* x Familiarity | -334.04 | -1.64 | .10 |
| Mode *Gamified* x Familiarity | -610.67 | -3.22 | .002 ** |
| | Tax Compliance [b] | | |
| | B | β | p |
| Intercept | 5.63 | | <.001 *** |
| Mode *Rich* | 0.62 | 0.80 | .43 |
| Mode *Gamified* | 1.01 | 1.40 | .16 |
| Familiarity | -0.54 | -1.73 | 0.08 |
| Mode *Rich* x Familiarity | 0.03 | 0.07 | 0.95 |
| Mode *Gamified* x Familiarity | 0.64 | 1.644 | 0.10 |

[a] Total mean score was used as a metric for performance.
[b] The Sum of relative compliance scores was used as a metric for tax compliance.
* $p < 0.05$, ** $p < 0.01$, ** $p < 0.001$.

## 3.3 Exploratory Analysis

### 3.3.1 Perceived realism.

Gamification is used as a means to increase context-richness (see also 1). The amount to which the experiment situation was perceived as *context-rich* can be explored with one item in the “Attitude towards Taxation” block, namely “Perceived Realism” (see also 2.2.3). No significant differences were found among the three conditions regarding this item, $F(2, 172) = 0.98$, $p = .38$.

*Figure 6 - Familiarity with Games as a moderator on Performance*

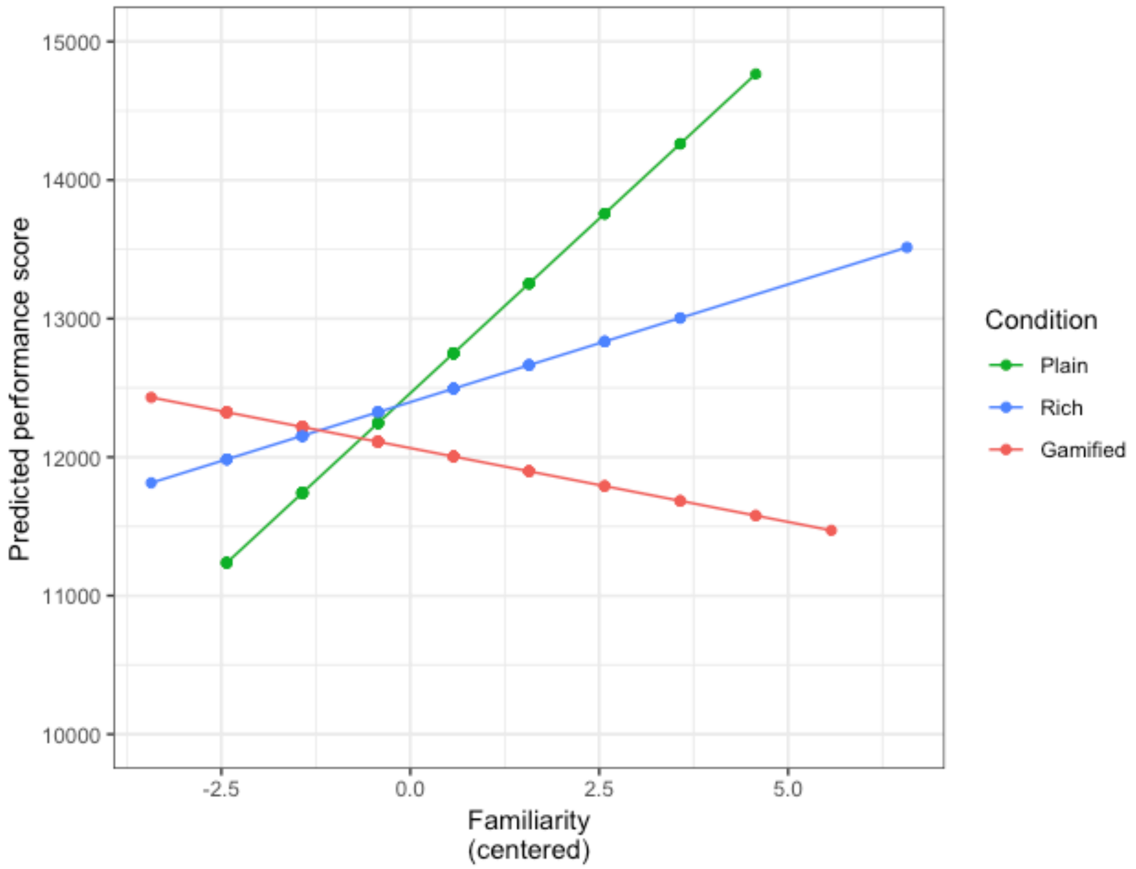
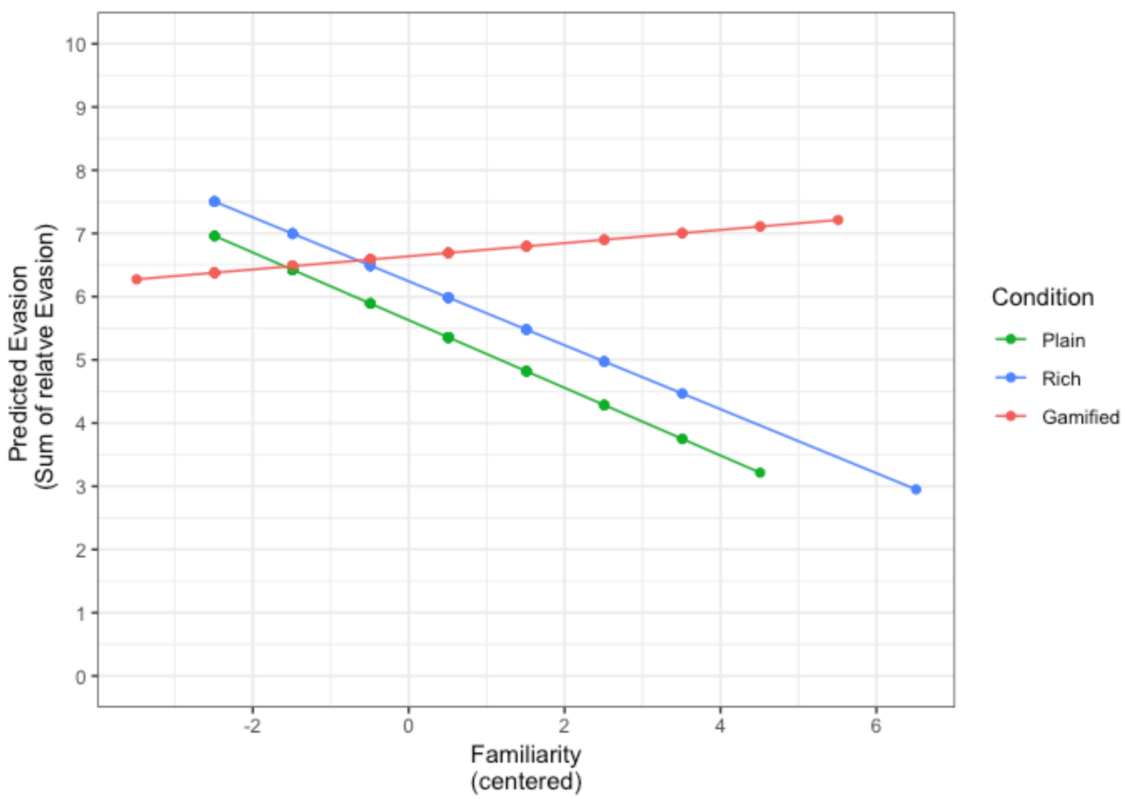



*Figure 7 - Familiarity with Games as a moderator on tax evasion*



# 4. Discussion

## 4.1 The impact of Gamification

### 4.1.1 Gamification and performance.

Looking at the results, it is obvious that the implementation of game elements did not have a significant effect on participant performance. While the *Gamification* condition was rated significantly more favorably than the *Plain* condition (although not significantly higher than the *Rich* condition), neither the overall performance, nor the performance progression differed significantly. Several possible explanations are available:

First, methodological failure: The most obvious explanation is that the current study failed to implement game elements in a way that are conducive to participants' motivation. This explanation would entail that the game elements were implemented incorrectly, or not sufficient enough to make a significant difference in participant behavior.

Since the participants rated the *Gamification* condition favorably concerning visual design and motivational aspects, this however seems unlikely. If the game elements were implemented incorrectly, the differences in favorability rating between the *Gamification* and *Rich* conditions would be more apparent. Also, the game elements were implemented following the best practices of Basten (2017) and Hamari and Eranti (2011).

Second, the limited fffects of short-term Gamification: Since a classic tax experiment is usually  completed within a short time span, any game elements have to fit this time span as well. This excludes long-term game elements such as a level system, which would reward participants for repeated participation. For the achievement badges, this also means that all badges must be earn-able in a single session, thus excluding long-term achievement badges.

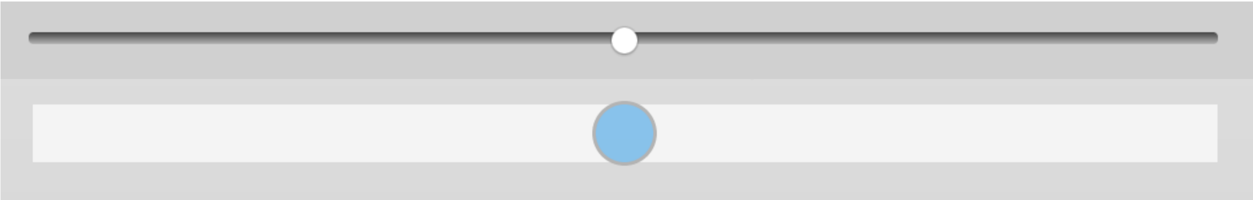
Many platforms that have implemented game elements (for examples, see 1.2) have features that facilitate long-term motivation (such as “leveling up” through experience points, or a large array of different achievements, many of which can not be completed at the first visit). It is entirely possible that the motivational benefits of short-term gamification are negligible, and the main factor for intrinsic motivation stems from long term gamification.

Third, confusing visual design: From the inception of the study, measures were taken to isolate the effect of visual design on overall motivation and impression. Since the *Plain* condition had marginally (though not significantly) better results than both the *Rich* and *Gamification* conditions, a possible explanation is that the visual design of the *Rich* and *Gamification* conditions was detrimental to participant performances.

When examining the slider tasks, the differences between the *Plain* condition and the other two are unlikely to be important: For all conditions, the slider elements had the same widths and margins. That means that each page for each condition had the same height, and all elements had the same placements on the pages. Also, all displays in the laboratory had the same dimensions and resolutions. Therefore, all participants had to traverse the same amount of scroll distance to complete their slider tasks. The sliders in the *Rich* and *Gamification* conditions had a different visual design than those in the *Plain* condition. For a direct comparison, see Figure 8. The *Rich* slider was designed with a larger clickable surface and a color scheme that fit the overall color scheme of the study. Again, the results make it appear unlikely that one design yields a significant advantage over the other.

Fourth, the “Chocolate Covered Broccoli”: Researchers and User Interface designers have become somewhat wary of applications with badly implemented game

*Figure 8 - Comparison between the* Plain *slider (above) and the* Rich/Gamification *slider (below)*



elements. The term "chocolate covered broccoli" (for example, Prince, 2013) is used to describe these applications: A tedious application does not become more motivating for its users if game elements are added artificially (just like broccoli does not become more appealing by adding chocolate sauce).

It is possible that adding game elements to an "incompatible" application not only has no significant positive impact on participant behavior, but actually a negative impact.

The *Gamification* condition was overall rated as the most interesting (though not significantly so; for details, see 3.2.4).

Another possible explanation is that participants who score high in familiarity with games were disappointed or annoyed by the game elements due to their perceived quality. A high score in "Familiarity with Video Games" and "Familiarity with Competitive Sports" was correlated positively with a high overall game score, but only in the *Plain* and *Rich* condition. This could indicate that participants who are familiar with games (and, consequently, more likely to be motivated by game elements) did not respond positively to the presented game elements.

Fifth, different levels of abstraction: Finally, the results may be explained simply by the differing levels of abstraction among the conditions. While the *Plain* and *Rich* condition offered very abstract information, and little feedback to the participants about their overall performance, the *Gamification* condition was designed to be more responsive (by awarding achievement badges) and more realistic (by coding the points as money).

#### 4.1.2 Gamification and tax evasion.

While the overall amount of tax evasion did not differ significantly among the three conditions, the pattern of progression shows some interesting effects (if not in a way that was originally predicted). In some rounds, participants in the *Gamified* condition tended to report their earnings more honestly.

A possible explanation is that in the *Gamified* condition, the 10% probability of a tax audit was much more salient: While participants in the *Plain* and *Rich* conditions had to simply push a button to progress (and possibly be audited), participants in the *Gamification* condition had to “draw a card”, symbolizing the 10% chance by clicking on one virtual card out of ten. This method of display was chosen to be more game-like. While the overall probability was the same in any condition (in fact, even the “card draw” mechanic in the *Gamification* condition would use the same random number method as the “button press” in the other two conditions), the fact that participant behavior differed is certainly interesting.

One possible explanation would be that due to the coding of the earned points as “currency”, a stronger motivation to preserve earned income was created (and therefore, avoid a fine by correctly reporting their earnings). This explanation is compliant with the ideas of Gamification, and implies that by defining a virtual score as currency, participants are more likely to ascribe value to it (akin to a currency in a game they obtained through playing).

Alternatively, the effect may be a result of the different perception of audit probability and process transparency in the different conditions. The probability of “drawing an ‘audit’ card” may be perceived as smaller than simply triggering an audit with a button press (Fox & Clemen, 2005). Also, the “card draw” mechanic may appear more transparent and visually salient (even though, again, the probability calculation was the same for all conditions).

#### 4.1.3 Familiarity with games.

The results show that participants who are familiar with video games and active sports tend to score higher in the *Plain* and *Rich* conditions, while there is no such effect in the *Gamification* condition. This effect may seem counterintuitive to the original supposition, which states that participants who were more familiar with games would perform better in the *Gamification* condition.

Based on the data, it is possible that participants who are familiar with games tend to perform better in the abstract, while no such effect is observable in realistic situations.

### 4.2 Limitations

#### 4.2.1 Limitations of the sample.

Regarding the sample, the sample consisted exclusively of undergraduate psychology students, who were required by their program to participate in a number of studies. This has several ramifications for the sample’s representativeness.

The relatively young mean and median ages of the sample ( 21, and 20 years respectively) would suggest that many participants have little experience earning money, and paying taxes. Therefore, their behavior in a tax declaration situation may not be representative of a general population.

Another limitation is that participants scored relatively low on the Familiarity with Games questionnaire. Any interpretation of the effect of familiarity with games on either the performance or the risk taking behavior is therefore somewhat fraught.

#### 4.2.2 Design limitations.

The present tax experiment has a relatively simple design. Among the ten rounds, there is no variation in difficulty, audit probability or penalty rate. These decisions were made in order to create a simple, baseline tax experiment on which to test the Gamification-related hypotheses.

However, these design decisions put the experiment at odds with either video games (many of which raise the difficulty over the duration of their play time to match the increase of player competency) as well as tabletop games (which tend to raise the stakes over the course of a game).

Future research should examine the effects of difficulty and risk (audit probability vs. penalty rate) variation in game and non-game settings.

#### 4.2.3 Long-term vs. short-term gamification.

In the context of a singular psychological experiment, some game elements are not feasible for implementation. For example, an experience system which would reward participants with “level ups” would likely be meaningless in the context of a short psychological experiment. In a literature review by Hamari et al. (2014), these long-term progression systems were found to have a positive impact on user motivation, although impact on test performance and other behavior was not conclusive.

Also, creating a meaningful narrative element is difficult in a short-term setting. While possible (there are countless narrative video games which can be completed in under an hour), no attempt was made in this study to include a narrative element.

To explore the motivational impact of long-term game elements, future studies could embrace a repeated-measure design, and compare the progression of participants over multiple "sessions".

#### 4.2.4 Monetary incentives.

Partly, this study was conducted to explore whether Gamification is a valid method to avoid monetary incentivization. Apparently, no such conclusion can be drawn from this study, since there were no monetary incentives in any of the conditions. This decision was made to comply with the University of Vienna's guidelines for psychological experiments.

It remains to be seen if and how the effects observed in this study will vary once monetary incentives are introduced.

### 4.3 Conclusions and Future Research

#### 4.3.1 Gamification in psychological experiments.

Overall, the introduction of game elements in a standard tax experiment has not led to a significant difference in participant behavior. On the other hand, participant feedback on the game elements was positive. Therefore, Gamification appears to be a valid instrument to design appealing tax experiments.

### 4.3.2 The current state of Gamification.

Due to the increased attention on Gamification in recent years, a number of critical reports have been published, both in an academic and economic context. Deterding (2010) criticized the often meaningless rewards, specifically the overuse of achievement badges in various services.

Social media platform Foursquare (see also: 1.2, 1.2.6) decided to remove many of their game elements from the main application (Foursquare, 2013, after Rimon, 2014), instead opting to use the game elements in a separate application. The company stated that the “game elements started to break down” with an increase in users, but failed to provide exact information what lead them to remove game elements.

### 4.3.3 Future research.

Future studies may address the limitations laid out above, or further the research of gamification in other settings.

## References

Alm, J., McClelland, G. H., Schulze, W. D. (1992). Why do people pay taxes? *Journal of Public Economics 48* (1), 21-38.

Basten, D. (2017). Gamification. *IEEE Software 17,* 76-81.

Chan, E. (2018). Exposure to National Flags Reduces Tax Evasion: Evidence from the United States, Australia, and Britain. *European Journal of Social Psychology 49* (2), 300-312. https://doi.org/10.1002/ejsp.2388

Cheong, C., Cheong, F. Filippou, J. (2013). Quick Quiz: A Gamified Approach for Enhancing Learning. *Proceedings of Pacific Asia Conference on Information Systems,* 18-22.

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). Retrieved from http://tarc.exeter.ac.uk/media/universityofexeter/businessschool/documents/centres/tarc/publications/discussionpapers/Choo_Fonseca_Myles_2014_v2.pdf

de Araujo, F. A., Carbone, E., Conell-Price, L., Dunietz, M. W., Jaroszewics, A., Landsman, R. et al. (2015). The Effect of Incentives of Real Effort: Evidence from the Slider Task. *CESifo Working Paper 5372*. https://ssrn.com/abstract=2615704

Deterding, S. (2010). Pawned. Gamification and its Discontents [Powerpoint Slides] Retrieved from https://www.slideshare.net/dings/pawned-gamification-and-its-discontents

Eriksen, K., Fallan, L. (1996). Tax knowledge and attitudes towards taxation; A report on a quasi-experiment. *Journal of Economic Psychology, 17,* 387-402.

Falk, A., Heckman, J. J. (2009). Lab Experiments Are a Major Source of Knowledge in the Social Sciences. *Science, 326*, 535-538.

Fox, C. R., Clemen, R. T. (2005). Subjective Probability Assessment in Decision Analysis: Partition Dependence and Bias Toward the Ignorance Prior. *Management Science 51*(9). https://doi.org/10.1287/mnsc.1050.0409

Gill, D., Prowse, V. (2012). A Structural Analysis of Disappointment Aversion in a Real Effort Competition. *American Economic Review, 102,* 469-503.

Hamari, J., Eranti, V. (2011). Framework for Designing and Evaluating Game Achievements. *Proceedings of DiGRA 2011 Conference: Think Design Play*.

Hamari, J. (2012). Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic Commerce Research and Applications, 12*(4), 236-245.

Hamari, J., Koivisto, J., Sarsa, H. (2014). Does Gamification work? — A Literature Review of Empirical Studies on Gamification. 4*7th Hawaii International Conference on System Science.* DOI 10.1109/HICSS.2014.377

Hamari, J. (2015). Do badges increase user activity? A field experiment on the effects of gamification. *Computers in Human Behavior (71)*, 467-478.

Hessing, D. J., Elffers, H., & Weigel, R. H. (1988). Exploring the limits of self-reports and reasoned action: An investigation of the psychology of tax evasion behavior. *Journal of Personality and Social Psychology, 54*(3), 405-413. http://dx.doi.org/10.1037/0022-3514.54.3.405

Huotari, K., Hamari, J. (2012). Defining Gamification - A Service Marketing Perspective. *Proceedings of the 16th International Academic MindTrek Conference 2012,* 17-23.

Kirchler, E. (2007). *The Economic Psychology of Tax Behaviour*. New York: Cambridge University Press.

Kogler, C., Mittone, L., & Kirchler, E. (2016). Delayed feedback on tax audits affects compliance and fairness perceptions. *Journal of Economic Behavior & Organization, 124*, 81–87. https://doi.org/10.1016/j.jebo.2015.10.014

Loewenstein, G. (1999). Experimental economics from the vantage-point of behavioural economics. *Economic Journal, 109*(453), 25-34.

Mascagni, G. (2017). From the Lab to the Field: A Review of Tax Experiments. *Journal of Economic Services, 32*(2), 273-301.

Montola, M., Nummenmaa, T., Lucerano, A., Boberg, M., & Korhonen, H. (2009). Applying Game Achievement Systems to Enhance User Experience in a Photo Sharing Service. *ek '09: Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era*, 94-97. Tampere, Finland.

Øivind Madsen, D., Stenheim, T. (2015). Experimental methods in economics and psychology: A comparison. *Social and Behavioral Sciences 187,* 113-117.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Prince, J. D. (2013). Gamification. *Journal of Electronic Resources in Medical Libraries (10),* 3, 162-169.

Rimon, G. (2014). Checking in and checking out: what Foursquare's evolution can teach us about Enterprise Gamification. https://www.gameffective.com/what-foursquares-evolution-can-teach-us-about-enterprise-gamification/

Runtastic Team (2015, June 2). *Brand New: Our Running Leaderboard on Runtastic.com*. Retrieved from https://www.runtastic.com/blog/en/brand-new-our-running-leaderboard-on-runtastic-com/

Roth, A.E. (1995) Introduction to experimental economics. In J.H. Kagel and A.E. Roth (eds.), *The Handbook of Experimental Economics*. Princeton, USA, Princeton University Press.

Shannon, S. (2010). The Changing Face of Winning in Video Games. *Museum of Play*. Retrieved from https://www.museumofplay.org/blog/chegheads/2010/11/the-changing-face-of-winning-in-video-games

Wenzel, M. (2004). An analysis of norm processes in tax compliance. *Journal of Economic Psychology (25),* 213-228.

# Appendix

## Abstract in German (Zusammenfassung auf Deutsch)

Das Ziel dieser Studie ist die Erforschung der Auswirkung von Spielelementen auf das Verhalten von Teilnehmer/innen in einem allgemeinen Steuerexperiment. Basierend auf bestehender Literatur wurde erwartet, dass die Implementierung von Spielelementen (im Allgemeinen als *Gamification* bekannt) einen positiven Effekt auf die Leistung und das Steuerhinterziehungsverhalten der Teilnehmer/innen hat. Es wurde erwartet, dass diese Effekte dadurch moderiert werden, wie vertraut Teilnehmer/innen mit verschiedenen Spieltypen sind. Es wurden eine *schlicht gestaltete*, eine *aufwändig gestaltete* und eine *Gamified* Bedingung verwendet, um die Wirkung der visuellen Gestaltung zu kontrollieren. Die drei Bedingungen haben keine signifikanten Unterschiede in der Leistung oder dem Steuerhinterziehungsverhalten bewirkt. Die Vertrautheit mit Spielen hatte nachweislich einen positiven Effekt auf die Leistung in der *schlichten* Bedingung, wogegen sie einen signfikanten negativen Effekt in der *aufwändigen* und *Gamified*-Bedingung hatte. Mögliche Erklärungen für die fehlenden Unterschiede zwischen den Bedingungen sind die beschränkten Effekte von kurzfristiger Gamification verglichen mit langfristiger Gamification, sowohl die Implementierung von ungenießbaren Spielmechaniken.

## Link to application repository

The application is available at the following URL: https://github.com/andreas-Hausberger/TaxGame. The availability and content of the application may be subject to change in the future.