Mendel University in Brno

MENDELU Working Papers
in Business and Economics

107/2025

Neuroscience technologies and minimization of administrative costs: An eye-tracking study of the effectiveness of its use in filling out tax forms

Břetislav Andrlík, Stanislav Mokrý, Petr David

MENDELU Working Papers in Business and Economics
Research Centre
Faculty of Business and Economics
Mendel University in Brno
Zemědělská 1, 613 00 Brno
Czech Republic
http://vyzc.pef.mendelu.cz/en
+420 545 132 605

Citation

Andrlík, B., Mokrý, S., David, P. (2025). Neuroscience technologies and minimization of administrative costs: An eye-tracking study of the effectiveness of its use in filling out tax forms. *MENDELU Working Papers in Business and Economics* 107/2025. Mendel University in Brno.

Abstract

Andrlík, B., Mokrý, S., David, P.: **Neuroscience technologies and minimization of administrative costs: An eye-tracking study of the effectiveness of its use in filling out tax forms.**

Research shows that eye-tracking can reveal which parts of documents attract the most attention and which are frequently overlooked. This method thus offers the public sector an opportunity to optimize form design, enhance efficiency, and reduce administrative burden. The integration of neuroscientific tools into public administration could significantly improve user-friendliness and process efficiency. The authors' findings confirm that eye-tracking and pupillometry are effective instruments for quantifying administrative burden during the completion of tax forms. The experiment demonstrated the ability of these technologies to objectively capture differences in time, attention allocation, and mental workload across various sections of the form. Mechanical sections, such as the transcription of identification data, exhibited lower cognitive demands, yet still generated high pupillometric values. Pupillometric analysis identified two peaks in mental load upon entering the calculation section and during the actual tax liability computation.

Key words

eye-tracking, pupillometry, tax forms, cognitive load, neuroscience technology, administrative costs

JEL: H23, H25, M390

Contacts

Břetislav Andrlík, Department of Finance and Accounting, Faculty of Business and Economics, Mendel university in Brno, Zemědělská 1, 613 00 Brno, Czech Republic, e-mail: bretislav.andrlik@mendelu.cz.

Introduction

The theoretical delineation of administrative costs is primarily associated with the evaluation of tax system efficiency. Within tax systems, administrative costs of taxation serve as one of the indicators of tax collection efficiency. The effectiveness of tax collection is among the fundamental principles of taxation. The current state of knowledge in this area is grounded in the classical tax canons formulated by Adam Smith (Smith, 1776), who, in addition to the principles of equity, certainty, and convenience of payment, emphasized the requirement of low collection costs, i.e., the frugality of the tax system. Every tax should be designed and implemented in such a way that the burdened taxpayers incur the lowest possible costs beyond the actual revenue generated by the imposed tax. Identifying ways to minimize administrative costs is thus a key responsibility of the public sector.

To enable efforts aimed at minimizing administrative costs, it is essential to establish reliable methods for measuring them, allowing for comparisons over time, across jurisdictions, or among different taxes within a given tax system. Several tools and methodologies are commonly used to measure administrative costs, each with varying degrees of accuracy in capturing the induced administrative burden. However, each of these methods has its own limitations. Given that induced administrative costs represent a pressing issue for public finance systems and that existing measurement methods are imperfect, there is a clear need to explore innovative and modern approaches, tools, and technologies that can complement the current methodological toolkit.

Key parameters of these methods include not only the ability to quantify induced costs but also the objectivity of the data obtained, the capacity to work with both qualitative and quantitative data, and the ability to identify removable bottlenecks that generate unnecessary administrative burdens.

The aim of this paper is to assess whether eye-tracking technology can be effectively employed to monitor gaze patterns and to identify pupil dilation characteristics, i.e., pupillometry, for the purpose of evaluating the cognitive demands of completing a tax return and verifying the suitability of this tool for measuring administrative costs. This approach builds on studies that have clearly demonstrated a close relationship between pupil dilation and mental effort. The empirical validation will be conducted using a road tax return form under the conditions of the Czech Republic.

1 Literature overview

In the private sector, neuroscience-based methods and tools, such as eye-tracking and pupillometry, are increasingly employed as key instruments for maintaining competitiveness in dynamic and technologically demanding environments (Khaneja and Arora, 2024). In contrast, these methods remain marginal within public administration, despite their potential to significantly contribute to the identification and elimination of inefficient processes, thereby reducing both administrative and other associated costs.

As eye-tracking technology has matured and become more accessible, its use has expanded. The design, implementation, and experimental parameters of studies employing this technology to investigate economic theory have been addressed by Lahey and Douglas (2016).

Eye-tracking has already been applied in research across finance, economics, and accounting. This tool enables researchers to uncover cognitive processes and decision-making strategies. In finance, eye-tracking is typically used to analyse how investors read financial reports and which information they consider most relevant. It can reveal how investors respond to different formats of financial statements and which sections attract the most attention (Lynch and Andiola, 2019). In this context, functional magnetic resonance imaging has also been employed (Barton et al., 2014).

Glimcher et al. (2009) highlight the contribution of neuroeconomics to the study of human economic behaviour, a view supported by Birnberg and Ganguly (2012) in their commentary. In economics, eye-tracking is used to understand economic behaviours such as consumer decision-making or hiring discrimination. It can reveal how individuals search for information and what factors influence their economic choices.

In accounting, eye-tracking has been used to analyse how accountants, depending on their level of seniority, interpret financial statements. It has also been applied to improve the user-friendliness of financial reports and enhance decision-making efficiency (Grigg and Griffin, 2014). Alton et al. (2014) employed eye-tracking to assess the extent to which users engage with form instructions. Their findings demonstrated that most individuals tend to underestimate the need to study the provided instructions.

1.1 Neuroscience technologies and administrative costs

Concerns regarding the validity of eye-tracking experiments have often stemmed from the assumption that the mere presence of eye-tracking equipment may alter participant's behaviour. However, Kee et al. (2021) demonstrated that eye-tracking devices, particularly high-tech systems, can be integrated into experimental designs without significantly affecting participants' economic behaviour.

A comprehensive review of studies conducted between 1980 and 2002 on the measurement of administrative costs was provided by Evans (2003), encompassing over 60 studies. Historically, administrative costs related to taxation were first assessed using survey-based methods (Haig, 1935). While these early studies focused primarily on compliance costs, a key insight was the inverse relationship between direct and induced administrative costs, an observation later confirmed by Chittenden, et al. (2010).

A more recent alternative for quantifying induced costs imposed by central government regulation is the Standard Cost Model (SCM). This methodology aims to quantify the activities that businesses must undertake to comply with mandatory information obligations (OECD, 2003). The SCM framework calculates costs as a product of three parameters: price, time, and quantity SCM Network, 2006). Price includes wages and overheads (for internal activities) or hourly rates (for outsourced services). Time reflects the duration required to complete the activity, and quantity denotes the number of obligated entities and the frequency of the activity per year. SCM's strength lies in its granularity, allowing for precise cost estimation at the level of individual administrative actions. However, the time parameter often relies on subjective estimates provided by surveyed firms.

Between 2006 and 2019, the World Bank and PricewaterhouseCoopers published annual comparative studies of tax systems across 191 economies (PwC, 2020). These studies aimed not only to benchmark tax systems but also to support global learning and reform. A key metric was the "ease of paying taxes", which included the number of payments required and the total time (in hours) needed to fulfil tax obligations.

A modern neuromarketing method with potential for measuring time-related aspects of tax compliance is the eye-tracking experiment. Eye-trackers, devices that monitor gaze behaviour, are widely used in fields such as neuroscience, psychology, industrial engineering, marketing, and computer science (Duchowski, 2003). In the context of administrative burden, eye-tracking has parallels with UX (User eXperience) research on web forms (Tan, 2009). In accounting and financial reporting, Grigg and Griffin (2014) explored how eye-tracking can inform the design of user-centered financial and business reports.

The use of emerging technologies in behavioural accounting experiments has been discussed by Rotaru et al. (2018), who introduced the concept of neuroaccounting, the application of neuroscientific tools to accounting research. Eye-tracking is particularly relevant for studies on attention and pupil response (pupillometry). Similarly, Tank and Farrell (2022) reviewed the role of neuroscience in accounting research, categorizing studies into decision-facilitating (how individuals process information) and decision-influencing (how individuals respond to controls). The present study aligns with the former, focusing on how individuals process information and complete tax forms.

1.2 Measuring Cognitive Load Using Eye-Tracking and Pupillometry

Cognitive load measurement via eye-tracking and pupillometry is based on the premise that cognitive processing demands manifest physiological eye responses. Modern eye-tracking systems can simultaneously record gaze trajectories and pupil diameter during task performance. Experimental designs typically manipulate task difficulty or information density to elicit varying levels of cognitive load. Pupil size is continuously monitored and compared to a baseline value recorded prior to the task or during low-demand control condition. The most common metric is the task-evoked pupillary response (TEPR), the change in pupil diameter relative to baseline (Krejtz et al., 2018).

To accurately capture cognitively driven changes, external conditions must be tightly controlled, especially lighting and visual stimulus brightness, as these can induce reflexive pupil changes unrelated to cognitive load (Duchowski et al., 2018; Mathôt and Vilotijević, 2022). Eye-tracking data are preprocessed to remove artefacts such as blinks, and pupil data are normalized to baseline to account for individual differences. While no universal analysis protocol exists, recent methodological frameworks have improved standardization (Mathôt and Vilotijević, 2022).

Beyond pupil dilation, eye-tracking data can reveal additional indicators of cognitive load. For example, blink frequency tends to decrease during intense concentration (Chen and Epps, 2014), and microsaccades, small involuntary eye movements, occur less frequently but with greater amplitude under high mental effort (Krejtz et al., 2018). These supplementary metrics can be triangulated with pupillometry to provide a more comprehensive picture of cognitive load.

Pupillometry offers a non-invasive, real-time method for monitoring mental workload. Unlike subjective self-reports or intrusive dual-task paradigms, eye-tracking operates passively in the background, preserving the user's primary task. The high temporal resolution (hundreds of Hz) enables detection of subtle dynamic changes in cognitive load. Numerous studies have confirmed that increased task demands correlate with pupil dilation, making pupillometry a reliable indicator of cognitive effort (van der Wel and van Steenbergen, 2018). This finding has been replicated across diverse contexts, from mental arithmetic to complex decision-making.

Another advantage is the ability to localize attention and identify moments or interface elements that trigger peak cognitive load, valuable insights for designing more effective interfaces and instructional materials (Chen and Epps, 2014). Cognitive load data can also be used adaptively; for instance, e-learning systems may adjust content delivery based on detected load levels (Yuksel et al., 2016). Eye-tracking and pupillometry thus serve as powerful tools for bridging cognitive load theory with practical applications in human–computer interaction (HCI) and intelligent interfaces.

Despite their strengths, these methods have limitations. Environmental factors, especially lighting, can significantly affect pupil size, potentially confounding cognitive load measurements. Even minor variations in brightness or contrast can trigger pupil reflexes unrelated to mental effort (Duchowski et al., 2018). Therefore, experiments must be conducted under controlled conditions with carefully calibrated stimuli (Mathôt and Vilotijević, 2022). In real-world settings, such as driving, pupil changes may reflect both cognitive load and environmental lighting fluctuations.

Another technical limitation is the pupil foreshortening effect, geometric distortion of the pupil image when gaze deviates from the camera's optical axis, leading to inaccurate diameter estimates (Duchowski et al., 2018). This introduces noise and reduces reliability outside optimal lab conditions. Moreover, pupil dilation reflects general autonomic arousal, which may be influenced by emotional excitement, stress, or surprise, factors not necessarily related to cognitive load. Consequently, pupillometry is often combined with other metrics (e.g., self-reports or performance tests) to isolate cognitive components.

Pupillometry also lacks specificity regarding the type of cognitive load—it indicates overall effort but cannot distinguish between intrinsic task complexity and external distractions. Some studies even show that reducing extraneous load can paradoxically increase pupil size, as freed-up mental capacity is redirected toward deeper processing (Rodemer et al., 2023). Thus, pupillometric data must always be interpreted within the context of the task and theoretical framework.

Finally, practical deployment may be constrained by equipment availability and calibration requirements. High-quality eye-trackers often require precise setup and head stabilization, which may be impractical outside laboratory environments. However, advances in integrated eye-tracking technologies (e.g., in headsets or laptops) are lowering these barriers. Overall, eye-tracking and pupillometry represent powerful tools for cognitive load assessment, but their effective use demands careful experimental design and nuanced interpretation.

.

2 Methodology

This study analysed differences in the time demands of completing individual sections of a tax return form and the distribution of visual attention between the form itself and the corresponding instructions. Below, we describe the participant selection, equipment and materials, data collection procedures, preprocessing of eye-tracking and pupillometric records, and the statistical methods applied.

Czech participants aged 18–65 (N = 60; balanced gender ratio) were divided into two experimental groups. Group A completed a tax return form for three vehicles, while Group B completed the form for only one vehicle. Assignment ensured a balanced distribution of gender and age across both groups. All participants signed informed consent prior to the experiment.

The experiment was conducted under laboratory conditions with controlled lighting and a fixed workstation setup. A 24″ BenQ Zowie XL2430 monitor (1920 × 1080 px, 144 Hz) displayed an online tax return form implemented in Google Sheets. The form was divided into seven core sections: (1) personal information (Part_1), (2) tax authority information (Part_2), (3) tax return details (Part_3), (4) residence and contact information (Part_4), (5) vehicle information (Part_5), (6) tax base calculation (Part_6), and (7) tax liability calculation (Part_7). Each section displayed two screen areas side-by-side: the form area on the left ("AOI Form") and the instruction area with specific values on the right ("AOI Instruction"). After completing a section, participants closed it and proceeded to the next one, maintaining a consistent chronological order for all participants. The entire experiment was designed using SR Research Web Link software, which enables interaction logging in a web-based interface.

Gaze position and pupil diameter were recorded using a desktop-mounted SR Research EyeLink 1000 Plus eye-tracker (1,000 Hz, resolution approx. 0.01°). A standard five-point calibration and validation procedure was conducted at the beginning of the experiment, with an acceptable mean error below 0.5° of visual angle. Before each section, a fixation cross (1,000 ms) was displayed to standardize the initial eye position and serve as a baseline for pupil dilation calculations. Each section was self-paced and ended when the participant closed the browser window.

Eye-tracking data were preprocessed using SR Research Data Viewer. All samples marked as blinks or invalid were removed from the pupillometric metrics. For each section, AOIs were defined as fixed rectangles covering the form and instruction areas. From each AOI, we extracted dwell time (cumulative fixation duration within the AOI) and both average and maximum pupil diameter values. Pupil diameter was measured in arbitrary units, calculated by the eye-tracking device as a pixel-based area metric. The primary outcome variables were: (1) total time spent completing the form in each section (IP_DURATION), (2) cumulative dwell time in AOI Form and AOI Instruction, (3) average pupil diameter change relative to baseline, and (4) maximum pupil dilation during each section.

We then transformed the data into two core datasets. For analysing time demands, we created a data frame where each row represented one respondent and one of the seven sections (with IP_Duration as the dependent variable in seconds; predictors included SectionBase, Group, Experience, Age, and Sex). For analysing attention allocation, we constructed a "long" format dataset where each row described a respondent × section × AOI combination, with Dwell_s indicating the cumulative time spent in the specified AOI (Form or Instruction).

Statistical analyses were conducted as follows:

Total Time Spent in the Form (IP_DURATION)

To assess differences in IP_DURATION across sections and between experimental groups A and B, we estimated a linear mixed-effects model (LMM) with a random intercept for respondent (ID) and fixed effects for SectionBase, Group, Experience, Age, and Sex, including their interactions with SectionBase. The model was specified as:

IP_Duration ~ SectionBase × (Group + Experience + Age + Sex) + (1 | ID)

Main effects and interactions were tested using Type III Wald chi-square tests to determine whether time demands varied across sections and whether these differences depended on group assignment or demographic variables. Post-hoc pairwise comparisons between groups A and B within each section were conducted using estimated marginal means (EMMs) with Benjamini–Hochberg correction for p-values.

Attention Allocation Between Form and Instruction

To analyse how participants' visual attention (measured by cumulative dwell times) was distributed between the form and instruction areas, we estimated a second LMM with Dwell_s (in seconds) as the dependent variable:

Dwell_s ~ AOI × Section × Group + AOI × Section × (Experience + Age + Sex) + (1 | ID)

Here, AOI (Form vs Instruction) tested whether time spent differed between areas in the reference section; the AOI × Section interaction captured how this difference varied across the seven sections; and the AOI × Section × Group interaction assessed whether the Form vs Instruction ratio differed between Group A (three vehicles) and Group B (one vehicle). Additional interactions with Experience, Age, and Sex were included to identify potential demographic effects on attention allocation. Again, Type III Wald chi-square tests were used to assess significance, followed by EMMs to compute proportions and contrasts for each section and group.

Effect of Completion Accuracy

After completing the final section ("Tax Base"), each participant's result was manually verified against the expected value. We focused solely on this section, as the others did not involve calculable outcomes. If the response matched the correct value, the variable Correct was coded as 1; otherwise, as 0. To determine whether correct completion was associated with different time demands, we included Correct as a fixed effect in an LMM alongside demographic variables (Age, Sex), Group (A vs B), and Experience (prior tax return experience). A random intercept for ID accounted for repeated measures. The model was specified as:

Dwell_s (in Part_7) ~ AOI × Section × Group + AOI × Correct + (1 | ID)

Statistical tests involved estimating the model using LMM functions, testing fixed effects via Wald t-tests (Satterthwaite approximation for degrees of freedom), and evaluating the significance of the Correct variable. We reported β coefficients (estimated time differences in seconds between correct and incorrect cases), standard errors, t-values, and p-values to assess statistical significance.

Economies of Scale

To estimate time savings from processing additional vehicles, we compared predicted average times for key sections, "Vehicle Information" (Part_5), "Tax Base Calculation" (Part_6), and "Tax Liability Calculation" (Part_7), between Group A (three vehicles) and Group B (one vehicle). For Group A, total time was divided by three to estimate per-vehicle time. Group B values remained unchanged. Relative savings were calculated as:

Time_B - (Time_A / 3) / Time_B × 100 [%]

Pupillometry

Pupillary data were analysed on two levels: (a) average pupil diameter change from baseline, and (b) maximum dilation during each section. For each metric, we estimated an LMM:

PupilDelta ~ Trial + (1 | ID)

where Trial denoted one of the seven sections. The main effect of Trial was tested using Type III Wald tests. We also examined the interaction Trial × Group to determine whether the A vs B condition influenced pupillary response. A dedicated LMM with a section-level factor and the same random intercept was used to compare average and maximum values. Our aim was to capture cognitive load during the completion of each tax form section. For the purposes of this study, we assumed that participants were not exposed to emotional stress while completing the form, meaning that pupillometric data should primarily reflect cognitive load.

All statistical analyses were conducted in RStudio (Version 2023.06.1+524). All graphs, including box plots and bar charts with 95% confidence intervals, were generated using the ggplot2 package.

3 Results

The data analysis focused on testing the hypothesis of scale economies, specifically, whether the average time required to complete a tax return decreases with the number of vehicles declared. Additionally, the study aimed to identify the most cognitively demanding sections of the tax form, examine attention allocation to instructional content, and explore interactions between time-related variables and sociodemographic characteristics.

3.1 Total Time Required for Completing Individual Form Sections

This subsection presents a quantitative analysis of the time burden associated with completing different sections of the tax return. A linear mixed-effects model (LMM) was employed to compare the duration of task completion across the main sections of the form. The model also tested the effects of task version (Group A: three vehicles vs. Group B: one vehicle), prior experience with tax returns, age, and gender.

The model enabled the identification of form sections that represent significant administrative bottlenecks and assessed whether these differ by group or user experience. Results from the Type III Wald chi-square tests (see Table 1) revealed that total completion time varied significantly across form sections (χ^2 (6) = 317.17; $p < .001$), confirming that some sections impose a substantially higher administrative burden than others.

The main effect of task version (Group A vs. B) was not statistically significant (χ^2 = 0.05; p = .822), nor were the effects of prior experience (χ^2 = 0.01; p = .922), age (χ^2 = 0.00; p = .980), or gender (χ^2 = 0.21; p = .644). These factors, in isolation, did not influence the average completion time.

However, significant interaction effects were observed:

FormSection × Group: χ^2 (6) = 293.45; $p < .001$, indicating that the time burden differed between groups in specific sections.

FormSection × Experience: χ^2 (6) = 33.79; $p < .001$, suggesting that experienced and inexperienced participants performed differently depending on the section.

FormSection × Age: $p = .042$, indicating a modest age-related effect on time spent in certain sections.

These findings highlight the importance of considering contextual and demographic interactions when evaluating administrative burden and form usability.

Tab. 1: Type III Wald χ^2 Tests for Global Effects in the Mixed-Effects Model

	Chisq	Df	Pr(>Chisq)
(Intercept)	5.553	1	0.018
FormSection	317.171	6	0.000
Group	0.050	1	0.822
Experience	0.010	1	0.922
Age	0.001	1	0.980
Sex	0.214	1	0.644
FormSection:Group	293.445	6	0.000
FormSection:Experience	33.790	6	0.000
FormSection:Age	13.045	6	0.042
FormSection:Sex	9.342	6	0.155

As shown in Table 2, the final section, Part_7: Tax Liability Calculation, represents the primary bottleneck in terms of administrative burden. Participants in Group A (three vehicles) required 581 ± 17 seconds, whereas those in Group B (one vehicle) needed only 426 ± 17 seconds, resulting in a statistically significant contrast of 425 seconds ($p < .001$).

Tab. 2: Estimated Marginal Means

FormSection	Group A	Group B	Diff	p
Part_1	93 ± 17.0	88 ± 17.1	5	0.8256
Part_2	70 ± 17.0	65 ± 17.1	6	0.8185
Part_3	126 ± 17.0	118 ± 17.1	8	0.7376
Part_4	119 ± 17.0	130 ± 17.1	-11	0.6610

Part_5	132 ± 17.0	80 ± 17.1	52	0.0321
Part_6	83 ± 17.0	50 ± 17.1	33	0.1685
Part_7	851 ± 17.0	426 ± 17.1	425	<0.001

Differences between individual sections of the form, in the context of group-level variation, are also clearly illustrated in Figure 1.

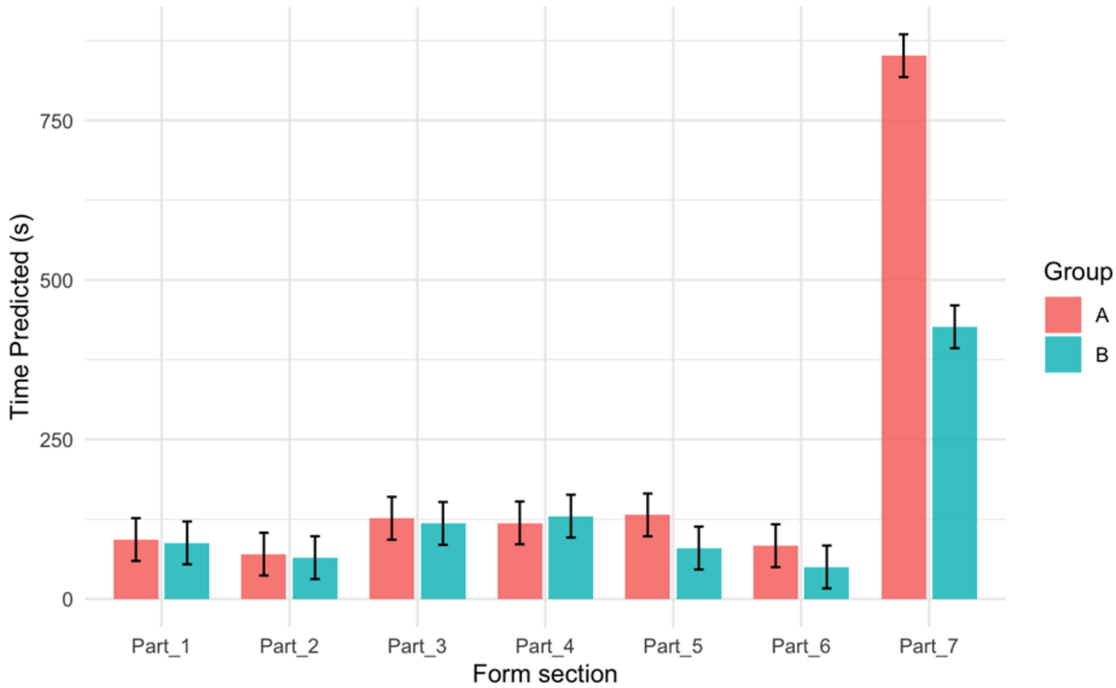

Fig. 1: Time for sections x groups

The Section × Group interaction revealed that the difference between task variants A and B was significant only in the tax liability calculation section, i.e., Part_7. Participants in Group B (one vehicle) required approximately 420 seconds less than those in Group A (three vehicles), see Table 3. Similarly, prior experience reduced completion time by approximately 158 seconds, but again, this effect was observed only in this most demanding section. Age and gender had minimal impact on total completion time, older participants spent slightly more time in the tax liability section (by approximately 3 seconds), while gender was not a significant factor in any section.

Tab. 3: Key Coefficients for the Tax Liability Calculation Section (Part 7)

Term	est_ci	p.value
SD (sk. A vs. refer.)	705 s (± 50.7)	0.000

SD × Group B	-420 s (± 31.6)	0.000
SD × Experience	-158 s (± 32.9)	0.000
SD × Age (1 year)	3 s (± 1.2)	0.009

Note: This table reports only statistically significant effects ($p < .05$).

3.2 Allocation of Visual Attention Between the Form and Instructions

To determine where participants directed their gaze during the task, we modelled dwell time (in seconds) spent by each respondent on the Form area (AOI = "Form") and the Instruction area (AOI = "Instruction") across all seven sections of the tax return. A linear mixed-effects model (with a random intercept for each respondent) was used to control experimental group (A/B), prior experience, age, and gender.

Global tests (see Table 4) confirmed that the distribution of visual attention depends on the section type and differs between experimental groups only in the calculation section. The main effect of AOI (Form vs. Instruction) was not significant in the reference section ($\chi^2 = 0.38$, $p = .54$), but the AOI × Section interaction was highly significant ($\chi^2 = 160.1$, df = 6, $p < .001$), indicating that the proportion of time spent on instructions versus the form varies substantially across sections.

The main effect of Group (A/B) was not significant ($\chi^2 = 0.02$), but both the Section × Group interaction ($\chi^2 = 443.5$, $p < .001$) and the three-way interaction AOI × Section × Group ($\chi^2 = 145.1$, $p < .001$) showed that group-level differences emerged only in the final tax liability calculation section (Part_7).

The effects of prior experience, age, and gender were observed at the section level (e.g., Section × Experience, $\chi^2 = 82.1$), but none of these factors significantly influenced the Form/Instruction ratio (all $p > .79$ for interactions with AOI).

In summary, in routine sections of the form, participants spent approximately one-quarter of their total dwell time reading instructions. However, in the final section (Part_7), this share dropped below 15%, and the time gap between Groups A and B widened considerably.

Tab. 4: Type III Wald χ^2 Tests for the Model Dwell_s ~ AOI × Section × Group

Effect	Chisq	Df	Pr(>Chisq)
(Intercept)	4.887	1	0.027
AOI	0.381	1	0.537

Section	390.534	6	0
Group	0.016	1	0.9
Experience	0	1	0.995
Age	0.162	1	0.688
Sex	0.392	1	0.531
AOI:Section	160.128	6	0
AOI:Group	0.005	1	0.941
Section:Group	443.478	6	0
AOI:Experience	0	1	0.992
AOI:Age	0.004	1	0.95
AOI:Sex	0.011	1	0.917
Section:Experience	82.115	6	0
Section:Age	13.3	6	0.039
Section:Sex	24.699	6	0
AOI:Section:Group	145.136	6	0
AOI:Section:Experience	45.696	6	0
AOI:Section:Age	3.18	6	0.786
AOI:Section:Sex	11.298	6	0.08

A closer examination of the time metrics reveals that in the initial administrative sections (Part 1–6), instructions accounted for 21–42% of total dwell time (see Table 5 and Figure 2). In contrast, during the final tax liability calculation section (Part_7), users focused almost exclusively on the form itself, 88% of dwell time in Group A and 93% in Group B.

Tab. 5: Estimated Times and Instruction Share in Each Section

Section	Group	Form	Instruction	Instr %
Part_1	A	40 s	22 s	35 %

Part_1	B	38 s	21 s	36 %
Part_2	A	35 s	13 s	26 %
Part_2	B	30 s	13 s	30 %
Part_3	A	81 s	21 s	21 %
Part_3	B	71 s	22 s	23 %
Part_4	A	55 s	25 s	31 %
Part_4	B	60 s	25 s	29 %
Part_5	A	54 s	40 s	42 %
Part_5	B	35 s	24 s	40 %
Part_6	A	49 s	16 s	24 %
Part_6	B	30 s	10 s	25 %
Part_7	A	572 s	75 s	12 %
Part_7	B	306 s	23 s	7 %

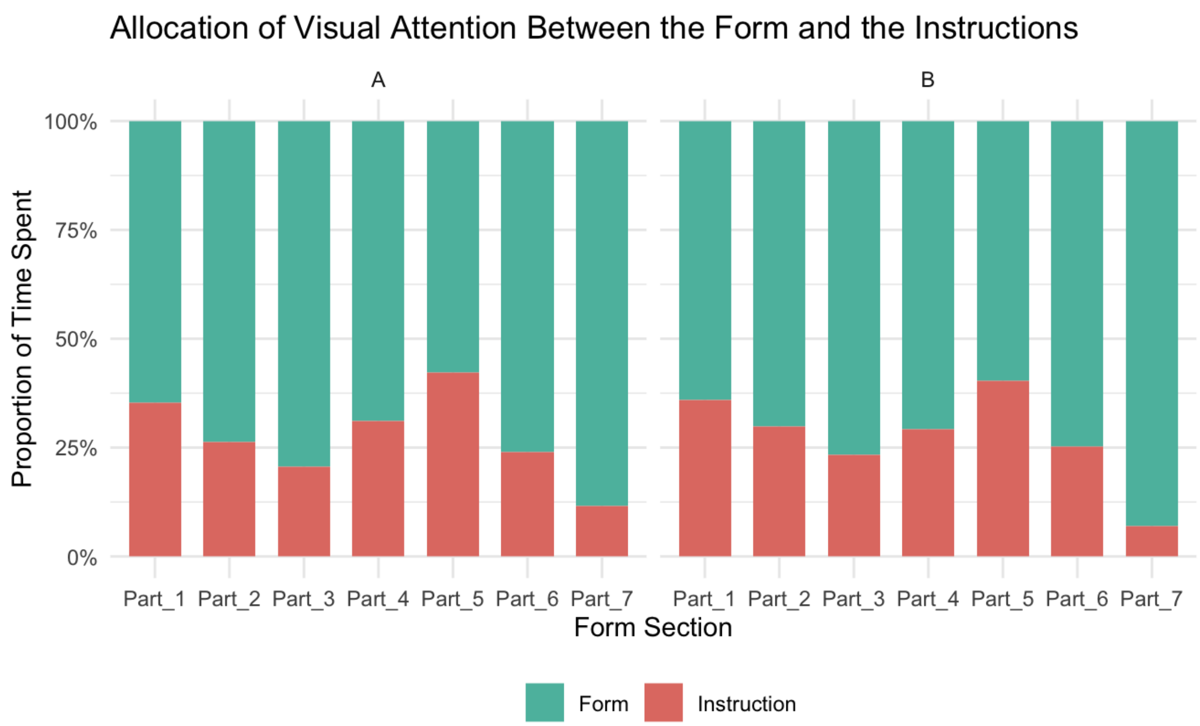

Fig. 2: Estimated Times and Instruction Share in Each Section

A comparison of completion times across form sections and participant groups is further illustrated in the boxplot figures provided in the appendix:

Figure S1 – Form_1: Personal Information
Figure S2 – Form_2: Tax Authority Information
Figure S3 – Form_3: Tax Return Details
Figure S4 – Form_4: Address and Contact Information
Figure S5 – Form_5: Vehicle Information
Figure S6 – Form_6: Tax Base Calculation
Figure S7 – Form_7: Tax Liability Calculation

These visualizations clearly demonstrate the disproportionate distribution of attention between the Instruction and Form areas. In all cases, participants spent more time interacting with the form itself than reading the accompanying instructions.

3.3 Accuracy of Completion: Tax Base – Road Tax Section

In the final stage of the road tax return process (Part_7), we conducted a detailed analysis of completion accuracy. As previously noted, this section was among the most complex, requiring participants to perform an actual tax liability calculation based on monthly records.

After completion, each submission was manually verified for correctness. The outcome was coded using the binary variable Correct (1 = correct, 0 = incorrect). This variable, along with age, gender, group assignment, and prior experience, served as predictors in a linear regression model (see Table 6).

Tab. 6: Linear Regression Model for Tax Liability Calculation - Correct effect

Predictor	β (s)	SE	t	p	Sig.
(Intercept)	547.745	68.460	8.001	< .001	***
Age	2.067	1.592	1.298	.200	
Sex (male)	55.663	40.463	1.376	.175	
Group B (1 vehicle)	-318.237	40.396	-7.878	< .001	***

Experience	-114.722	42.114	-2.724	.009	**
Correct (yes)	135.469	40.671	3.331	.002	**
N = 60; R^2 = 0.613; adj. R^2 = .578; SE residua = 156.3 s; F (5,54) = 17.13; p < .001					

The linear model predicting the total time spent on this section (in seconds) revealed that the most significant predictors were group assignment, prior experience with tax form completion, and the correctness of the solution. Participants in Group B, who interacted with a differently structured layout, spent on average 318 seconds less in this section than participants in Group A ($p < 0.001$). This may indicate a lower level of cognitive load or more efficient navigation within the interface. Prior experience with completing a tax return led to significantly shorter completion times, approximately 115 seconds less on average ($p = 0.009$), which confirms the positive impact of prior experience on the efficiency of solving complex form-based tasks. Conversely, participants who completed the section correctly spent substantially more time on it than those who made errors, on average about 135 seconds longer ($p = 0.002$). This result suggests that accurate completion of a complex calculation requires greater cognitive effort, potentially reflecting more thorough reading, increased time spent verifying results, or repeated recalculations.

Figure 3 illustrates the differences in completion times for the section involving the tax liability calculation. The figure clearly shows that participants who completed the section correctly spent significantly more time on it, both in Group A and Group B, which supports the findings of the linear model.

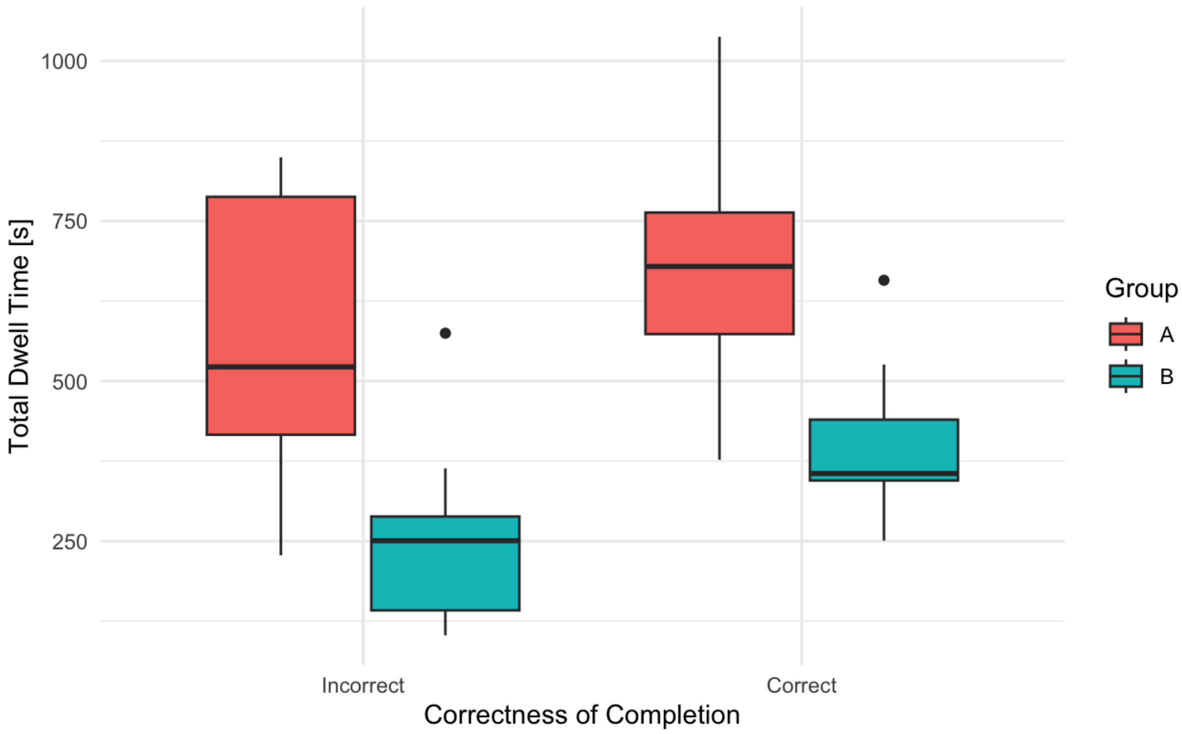

Fig. 3: Box-plot Correct x Incorrect – Form_7 Tax Base

3.4 Economies of Scale – Completion Time by Number of Vehicles

In relation to potential economies of scale, the analysis of per-vehicle completion times (see Table 7) shows that the marginal time increment is substantially lower than the time required for the first vehicle, yet overall leads to only limited time savings. In the mechanically oriented section ("Vehicle Information" – Part_5), the average time per vehicle decreased from approximately 80 seconds to about 44 seconds, representing a reduction of roughly 45%.

Similarly, in the initial section of the tax base calculation (Part_6), the overall time savings were of a comparable magnitude, with Group A completing one vehicle approximately 45% faster. However, in the actual tax liability calculation ("Tax Base – ROAD TAX" – Part_7), the marginal savings dropped to only about 29.3%, as each additional calculation still required recalculations and rate verification.

Tab. 7: Recalculated Estimated Marginal Means – Relative Time Savings for Form Completion

Form section	Time (Group A, 3 vehicles)	Time (Group B, 1 vehicle)	Per-vehicle time A (s/veh.)	Per-vehicle time B (s/veh.)	Relative savings
Part_5	132 ± 17.0	80 ± 17.1	44	80	45 %

Part_6	83 ± 17.0	50 ± 17.1	27.7	50	44.6 %
Part_7	851 ± 17.0	426 ± 17.1	283.7	426	29.3 %

* Relative savings = (TimeGroupB – (TimeGroupA / number of vehicles)) / TimeGroupB

Repeated form completion does yield time savings; however, their effect may not be substantial enough to provide taxpayers with multiple vehicles a significant time advantage.

3.5 Pupillometry – Average Pupil Dilation

In this part of the analysis, we examined differences in pupil dilation across the individual tasks of the tax return form. To control individual variability, we employed a linear mixed-effects model (LMM) with a random intercept for each respondent. The fifth task (Part_5) served as the reference level, as it exhibited the average level of pupil dilation, see Figure 4.

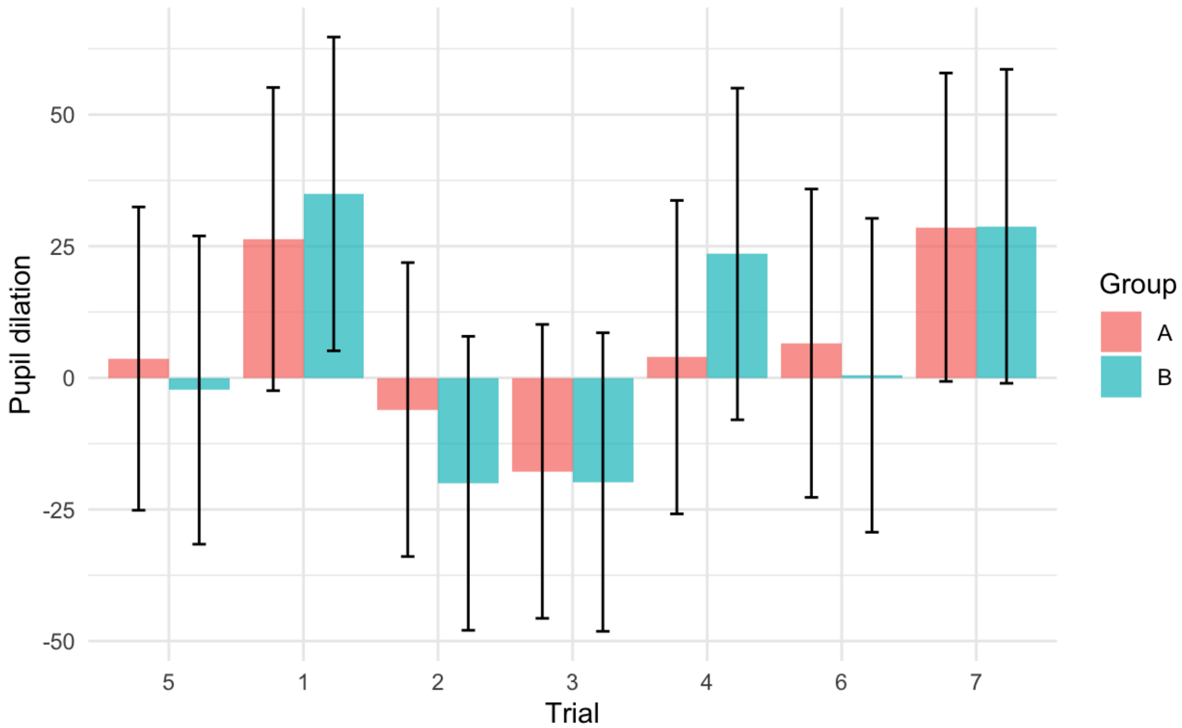

Fig. 4: Mean Pupil Dilation (± 95 % CI) by Trial for Groups A and B

The model revealed statistically significant increases in pupil dilation during the first task (Part_1) (b = 29.70, p = 0.027) and the seventh task (Part_7) (b = 27.94, p = 0.038), compared to the reference task

(Part_5), as shown in Table 8. This may indicate elevated cognitive load during the initial interaction with the form, such as orientation and task structuring, and again in the final phase, where calculations and verification operations may accumulate.

Tab. 8: Linear Mixed Model Results: Comparison of Mean Pupil Dilation Across Parts

Part (vs. ref = 5)	Estimate	Std. Error	df	t-value	p-value
Intercept	0.7483	10.3367	327.68	0.072	0.9423
Part_1	29.6883	13.3530	324.69	2.223	0.0269 *
Part_2	−13.7838	12.9828	322.22	−1.062	0.2892
Part_3	−19.5031	13.0426	322.87	−1.495	0.1358
Part_4	11.2010	13.3642	325.40	0.838	0.4026
Part_6	2.8242	13.4108	324.29	0.211	0.8333
Part_7	27.9320	13.4108	324.29	2.083	0.0381 *
— Random intercept variance —	1118.0	(sd = 33.44)	—	—	—
— Residual variance —	4814.0	(sd = 69.38)	—	—	—

When the model was extended to include interactions with Group A/B, none of the interaction terms proved statistically significant. Similarly, prior experience with completing tax returns did not show a significant main effect or interaction with individual form sections.

Manual review of video recordings revealed that six participants in Group B encountered difficulties when completing the email field in Section 4, which required entering contact details. These participants most struggled with entering the “@” symbol or failed to read the instructions thoroughly, resulting in corrections to previously entered text. However, these participants were not excluded from the analysis and remain included in the dataset.

In the graphical summary of results (Figure 4), a notable increase in pupil dilation is visible for this group specifically in Section 4. Although this spike was not statistically significant, its magnitude may be relevant in the context of participant behaviour. It is therefore reasonable to assume that the increased dilation in this section reflects elevated cognitive load caused by uncertainties or technical difficulties, which were not sufficiently homogeneous across the sample to reach statistical significance.

3.6 Pupillometry – Maximum Pupil Dilation

To complement the analysis, we examined not only average pupil dilation but also the maximum pupil dilation observed during each task, see Figure 5. This metric may more sensitively reflect short-term peaks in cognitive load that could be obscured in average values. As in the previous analysis, we applied a linear mixed-effects model with the same structure.

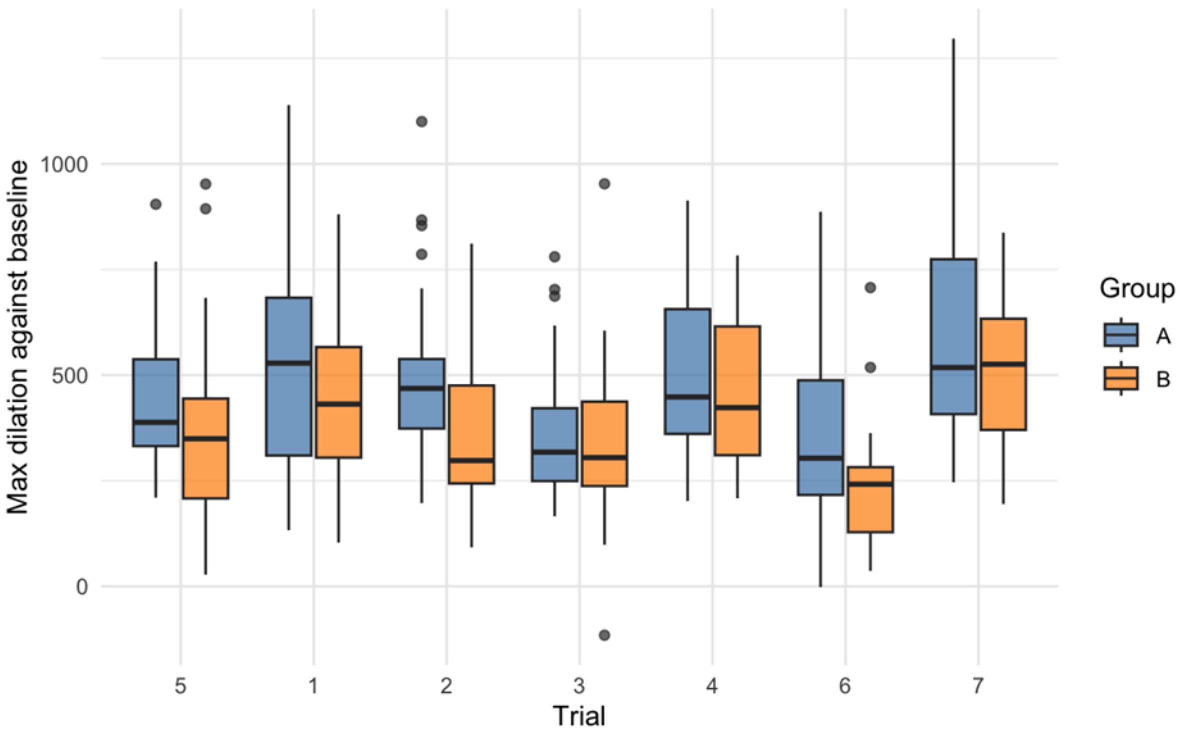

Fig. 5: Distribution of Maximum Pupil Dilation by Trial and Group (Boxplot)

The model showed that maximum pupil dilation differed significantly across several tasks compared to the reference task (Part_5). Statistically significant increases were observed in the first task (b = 89.09, p = 0.006), second task (b = 62.85, p = 0.048), fourth task (b = 64.73, p = 0.042), and seventh task (b = 184.39, p < 0.001). In contrast, the third and sixth tasks exhibited significantly lower maximum dilation (b = –70.42, p = 0.027, and b = –78.31, p = 0.017, respectively), as shown in Table 9.

Tab. 9 Effect of Task (Trial) on Maximum Pupil Dilation (vs. Part_5)

Part (vs. ref = 5)	Estimate	Std. Error	t value	p value	Significance
Part_1	89.09	32.23	2.765	0.00603	**
Part_2	62.85	31.68	1.984	0.04817	*

Part_3	-70.42	31.68	-2.223	0.02695	*
Part_4	64.73	31.68	2.043	0.0419	*
Part_6	-78.31	32.58	-2.404	0.01679	*
Part_7	184.39	32.58	5.661	3.38e-08	***
— Random intercept variance —	27384	(sd = 165.5)	—	—	—
— Residual variance —	14401	(sd = 120.0)	—	—	—

These findings confirm that the level of cognitive load varied significantly across different sections of the form. The maximum pupil dilation was particularly pronounced in the final task, which may be attributed to increased mental effort during final verification or the completion of calculations. Similar to the results for average dilation, no statistically significant differences were observed between Groups A and B, nor were there any significant interactions between group and task.

Conclusions and discussion

Our findings confirm that eye-tracking and pupillometry are effective and valid tools for quantifying administrative burden during the completion of tax forms. The experiment demonstrated that these methods can objectively capture differences in completion time, attention allocation, and mental workload intensity across various sections of the form. The results suggest that these tools can be effectively used to identify bottlenecks that generate unnecessary induced administrative costs. The insights gained can then be applied to optimize tax return forms and reduce administrative burden by addressing the identified "pain points." Improving the interface and instructions in these critical sections can significantly reduce mental load.

Our data clearly show that when evaluating economies of scale, the primary time bottleneck is the "Tax Base – Road Tax" section (Part_7), where the taxpayer performs the actual calculation of total tax and its allocation across the months of the reporting period. In this section, participants in Group B (one vehicle) spent on average 318 seconds less than those in Group A (three vehicles), with total time variability exceeding six minutes—more than 40% of the entire form completion process. In contrast, Parts 5 and 6 primarily involved mechanical transcription of vehicle identification data; the difference between groups here reflected only the number of entries, not cognitive complexity. These sections generally required the most time and also generated the highest pupillometric load, especially in correctly completed cases.

Marginal analysis revealed that while each additional vehicle increased total completion time, the increase was far from linear. In mechanical sections (e.g., transcription of identification data), the average time per vehicle decreased from approximately 80 seconds to 44 seconds (a 45% reduction), whereas in the actual tax calculation, the reduction was only about 29%. Thus, relative savings ranged from 20–50%, depending on task complexity. Nevertheless, taxpayers with multiple vehicles still spent significantly more time overall. This indicates that while repetition of identical fields may yield a "learning effect," fixed cognitive costs associated with verification and recalculation of rates remain high. In practice, simply duplicating field blocks will not lead to dramatic time savings; automation is necessary (e.g., bulk import of technical parameters from vehicle registries, prefilling rates based on engine size and weight, real-time validation), possibly supported by a step-by-step guided interface.

The linear mixed-effects model explained 61% of the variance; key predictors included:

(i) the complexity of the tax calculation layout,

(ii) prior experience, and

(iii) accuracy of the result—participants who calculated the tax correctly paradoxically required approximately 135 seconds more, likely due to thorough verification (a classic speed–accuracy trade-off).

Statistical tests further showed that prior experience with tax return completion reduced average time by approximately 115 seconds, reflecting familiarity with the general structure of similar forms. Age had a positive effect in simpler sections, older respondents worked faster, likely due to long-term familiarity with the form. Gender analysis revealed that men spent more time verifying personal and contact information, suggesting different strategies for validating identification fields.

Pupillometric analysis confirmed two distinct peaks in mental workload. The first occurred immediately upon entering the calculation section, where the average pupil diameter increased by approximately 30 arbitrary units (precisely 29.7) compared to baseline, indicating an immediate perception of increased complexity. The second peak occurred during the actual tax computation, where maximum dilatation reached +185 arbitrary units (precisely 184.39). This pattern aligns with the task-evoked pupillary response model, which posits that pupil dilation increases proportionally with cognitive load.

This study included only two levels of the "number of vehicles" variable (one vs. three) and thus cannot fully describe the curve of scale economies. Future research should incorporate a broader range of values (e.g., one to ten vehicles). Another limitation is the laboratory setting with controlled lighting and a desktop eye-tracker, which enhances internal validity but reduces generalizability to home environment. Field studies in naturalistic settings would be beneficial, though replicating lighting conditions poses challenges for comparability. To obtain a more comprehensive picture, additional technologies such as EEG or EDA could be employed. Pupillometry alone cannot distinguish cognitive effort from emotional stress, an area where EEG could provide valuable differentiation. For the purposes of this study, we assumed that participants were not exposed to emotional stress during form completion.

The combined use of standard administrative cost measurement methods and neuromarketing experiments offers a synergistic effect, enabling the exploration of cross-links between sociodemographic variables and time-based behavioural data.

There is also potential to expand the experiment in terms of both sample size and the range of variables included in the analysis.

Given the capabilities of eye-tracking technology, a promising direction for future research is the continued use of pupillometry to assess cognitive load in relation to demographic characteristics. Another option is the integration of electroencephalography (EEG) to compare cognitive responses based on prior experience with the task domain.

Neuroscience technologies in general and especially eye-tracking and pupillometry, combined with detailed analysis of dwell times and pupil dilation, offer strong potential for identifying specific steps that generate the highest administrative costs and for designing targeted interventions to simplify tax forms.

References

Alton, N., Rinn, C., Summers, K., Straub, K., and IEEE. (2014). Using eye-tracking and form completion data to optimize form instructions. Paper presented at the 2014 IEEE International Professional Communication Conference (IPCC).

Barton, J., Berns, G. S., and Brooks, A. M. (2014). The neuroscience behind the stock market's reaction to corporate earnings news. The Accounting Review, 89(6), 1945–1977. https://doi.org/10.2308/accr-50841

Birnberg, J. G., and Ganguly, A. R. (2012). Is neuroaccounting waiting in the wings? An essay. Accounting, Organizations and Society, 37(1), 1–13. https://doi.org/10.1016/j.aos.2011.11.004

Chen, S., and Epps, J. (2014). Using task-induced pupil diameter and blink rate to infer cognitive load. Human–Computer Interaction, 29(4), 390–413. https://doi.org/10.1080/07370024.2014.892428

Chittenden, F., Foster, H., and Sloan, B. (2010). Taxation and red tape: The cost to British business of complying with UK tax system. The Institute of Economic Affairs.

Duchowski, A. T. (2007). Eye tracking methodology: Theory and practice. Springer Science and Business Media.

Duchowski, A. T., Krejtz, K., Gehringer, D., Biele, C., Niedzielska, A., Kiefer, P., ... and Giannopoulos, I. (2018). The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 282. https://doi.org/10.1145/3173574.3173856

Glimcher, P. W., Camerer, C. F., Fehr, E., and Poldrack, R. A. (2009). Introduction: A brief history of neuroeconomics. In P. W. Glimcher, C. F. Camerer, E. Fehr, and R. A. Poldrack (Eds.), Neuroeconomics: Decision making and the brain (pp. 1–12). Elsevier.

Grigg, L., and Griffin, A. (2014). A role for eye-tracking research in accounting and financial reporting? In M. Horsley, M. Eliot, B. Knight, and R. Reilly (Eds.), Current trends in eye tracking research (pp. 233–248). Springer. https://doi.org/10.1007/978-3-319-02868-2_17

Haig, R. M. (1935). The cost to business concerns of compliance with tax laws. Management Review, 232–233.

Kee, J., Knuth, M., Lahey, J. N., and Palma, M. A. (2021). Does eye-tracking have an effect on economic behavior? PLoS ONE, 16(8), e0254867. https://doi.org/10.1371/journal.pone.0254867

Khaneja, S., and Arora, T. (2024). The potential of neuroscience in transforming business: A meta-analysis. Future Business Journal, 10, 77. https://doi.org/10.1186/s43093-024-00369-7

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., and Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. PLoS ONE, 13(9), e0203629. https://doi.org/10.1371/journal.pone.0203629

Lahey, J. N., and Oxley, D. (2016). The power of eye tracking in economics experiments. American Economic Review, 106(5), 309–313. https://doi.org/10.1257/aer.p20161009

Lynch, E. J., and Andiola, L. M. (2019). If eyes are the window to our soul, what role does eye-tracking play in accounting research? Behavioral Research in Accounting, 31(2), 107–133. https://doi.org/10.2308/bria-52283

Mathôt, S., and Vilotijević, A. (2022). Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis. Behavior Research Methods, 55(6), 3055–3077. https://doi.org/10.3758/s13428-022-01957-7

OECD. (2003). International standard cost model manual. https://www.oecd.org/gov/regulatory-policy/34227698.pdf

PwC. (2020). Paying taxes. https://www.pwc.com/gx/en/services/tax/publications/paying-taxes-2020.html

Rodemer, M., Eitel, A., Scheiter, K., and Schmitt, T. (2023). Pupil dilation as cognitive load measure in instructional videos on complex chemical representations. Frontiers in Education, 8, 1062053. https://doi.org/10.3389/feduc.2023.1062053

Rotaru, K., Schulz, A. K.-D., and Fehrenbacher, D. D. (2018). New technologies for behavioural accounting experiments. In The Routledge companion to behavioural accounting research (pp. 253–272). Routledge. https://doi.org/10.4324/9781315710129-18

SCM Network. (2006). International standard cost model manual. http://www.oecd.org/regreform/regulatory-policy/34227698.pdf

Smith, A. (1776). An inquiry into the nature and causes of the wealth of nations. W. Strahan and T. Cadell.

Tan, C. C. (2009). Web form design guidelines: An eyetracking study. https://www.cxpartners.co.uk/our-thinking/web_forms_design_guidelines_an_eyetracking_study/

Tank, A. K., and Farrell, A. M. (2022). Is neuroaccounting taking a place on the stage? A review of the influence of neuroscience on accounting research. European Accounting Review, 31, 173–207. https://doi.org/10.1080/09638180.2020.1866634

Yuksel, B. F., Oleson, K. B., Harrison, L., Peck, E. M., Afergan, D., Chang, R., and Jacob, R. J. (2016). Learn piano with BACh: An adaptive learning interface that adjusts task difficulty based on brain state. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 5372–5384. https://doi.org/10.1145/2858036.2858388

van der Wel, P., and van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. Psychonomic Bulletin and Review, 25(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Appendix

Fig: S1 – Part_1 - Personal Information

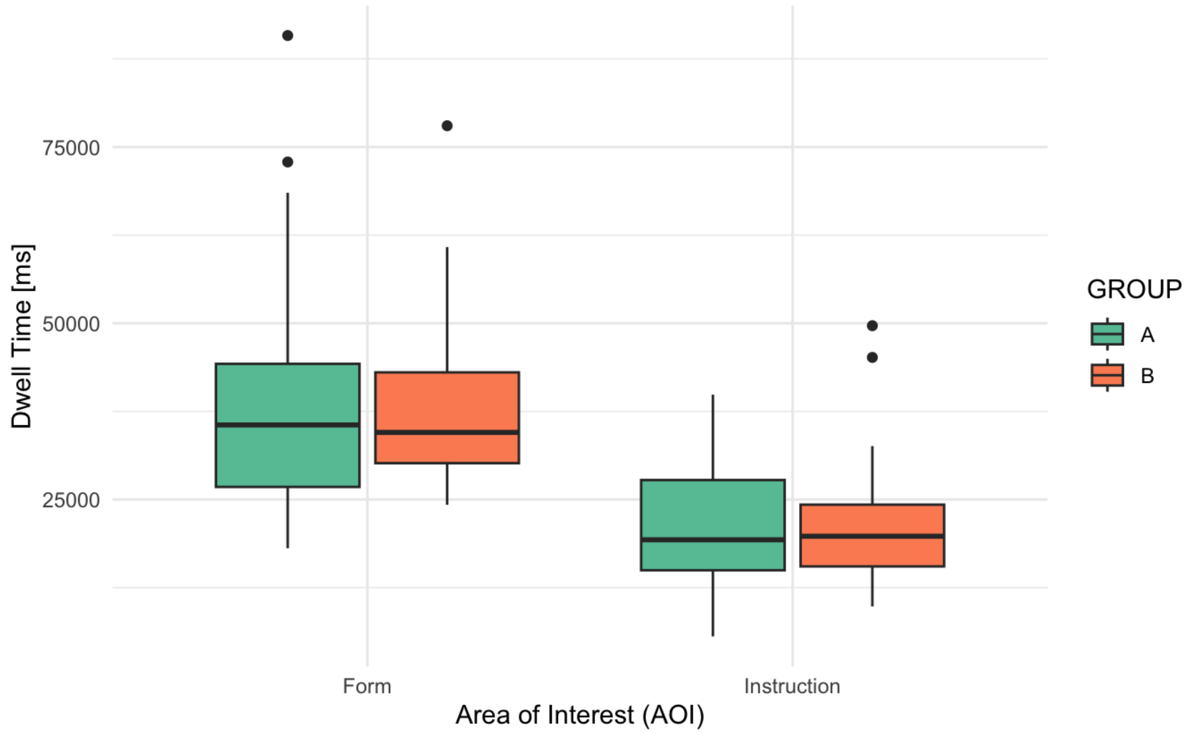

Fig: S2 – Part_2 – Tax Authority Information

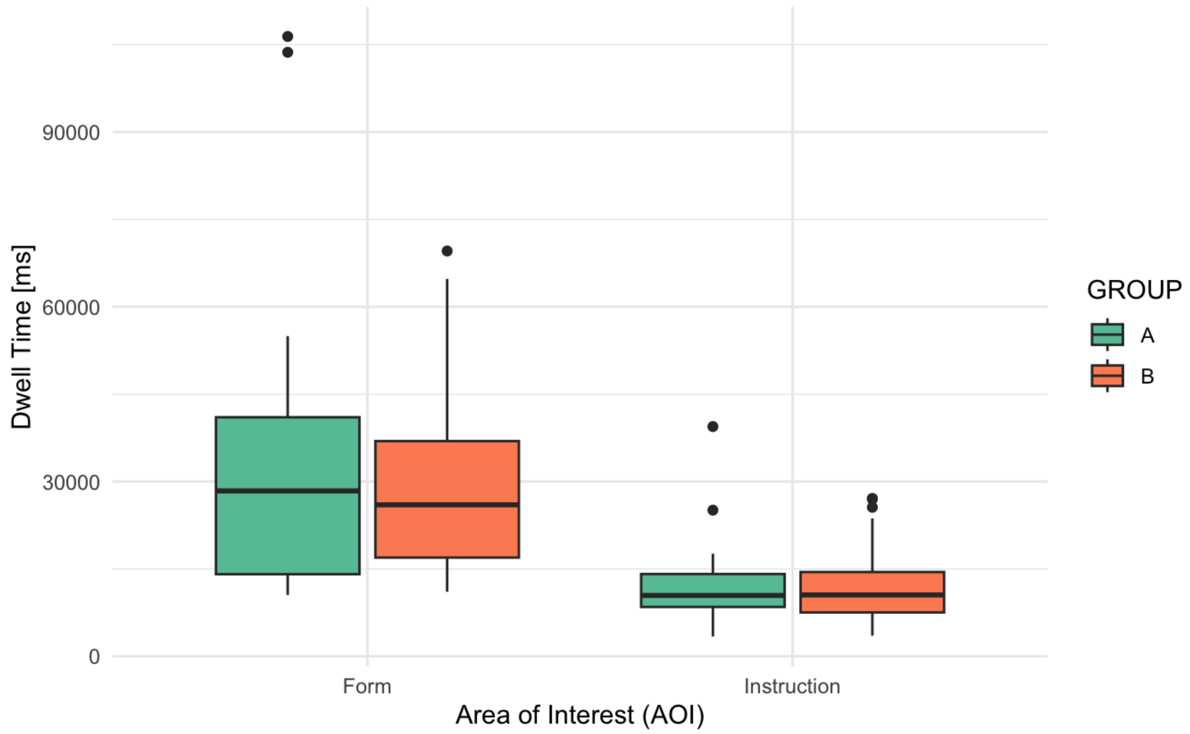

Fig. S3 – Part_3 – Tax Return Details

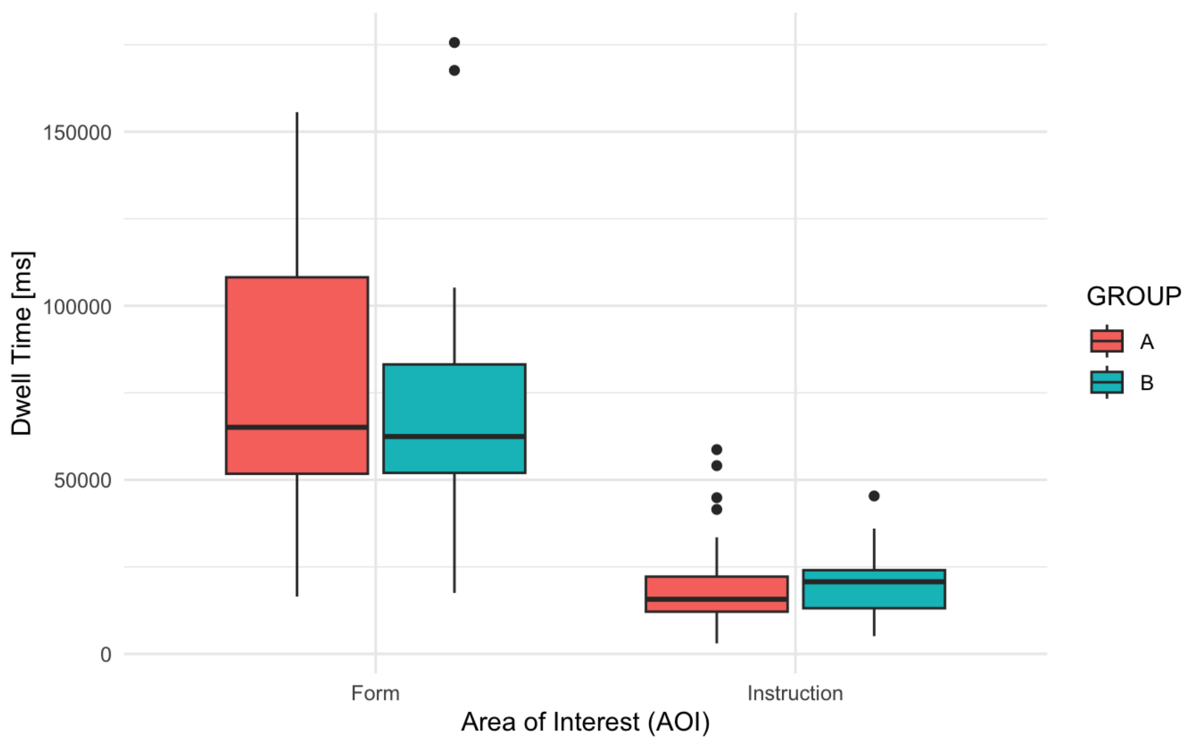

Fig. S4 – Part_4 - Address and Contact Information

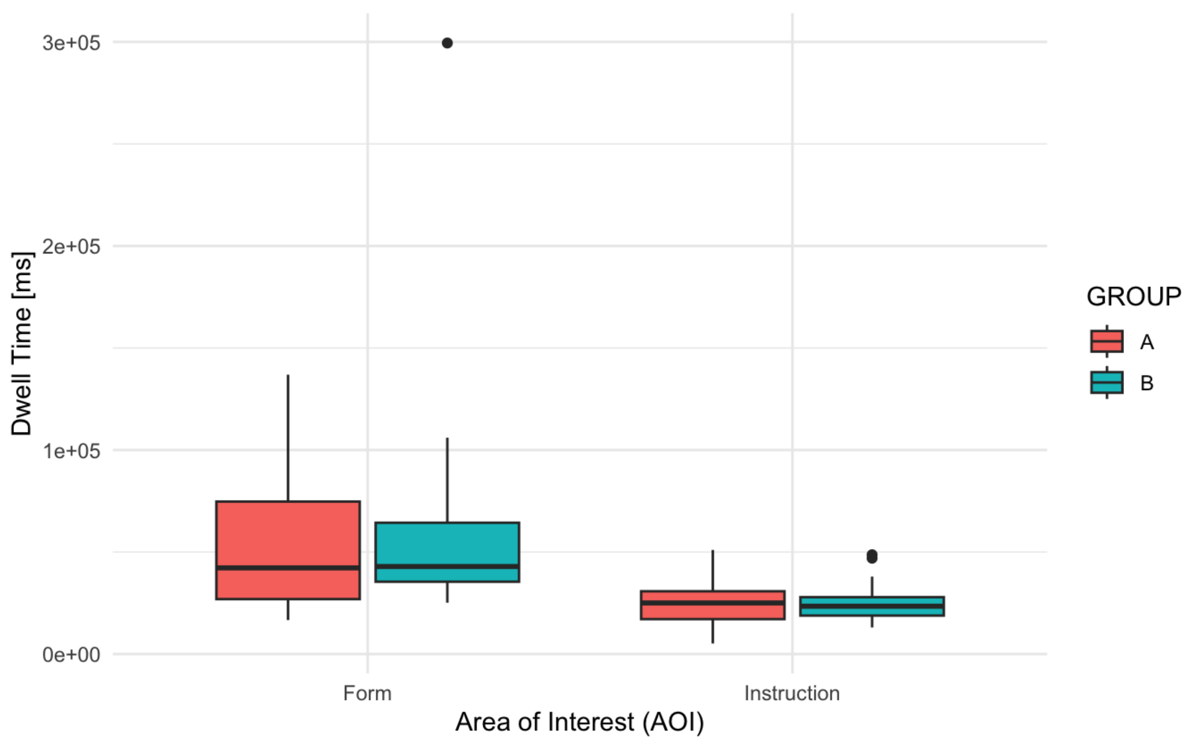

Fig. S5 – Part_5 - Vehicle Information

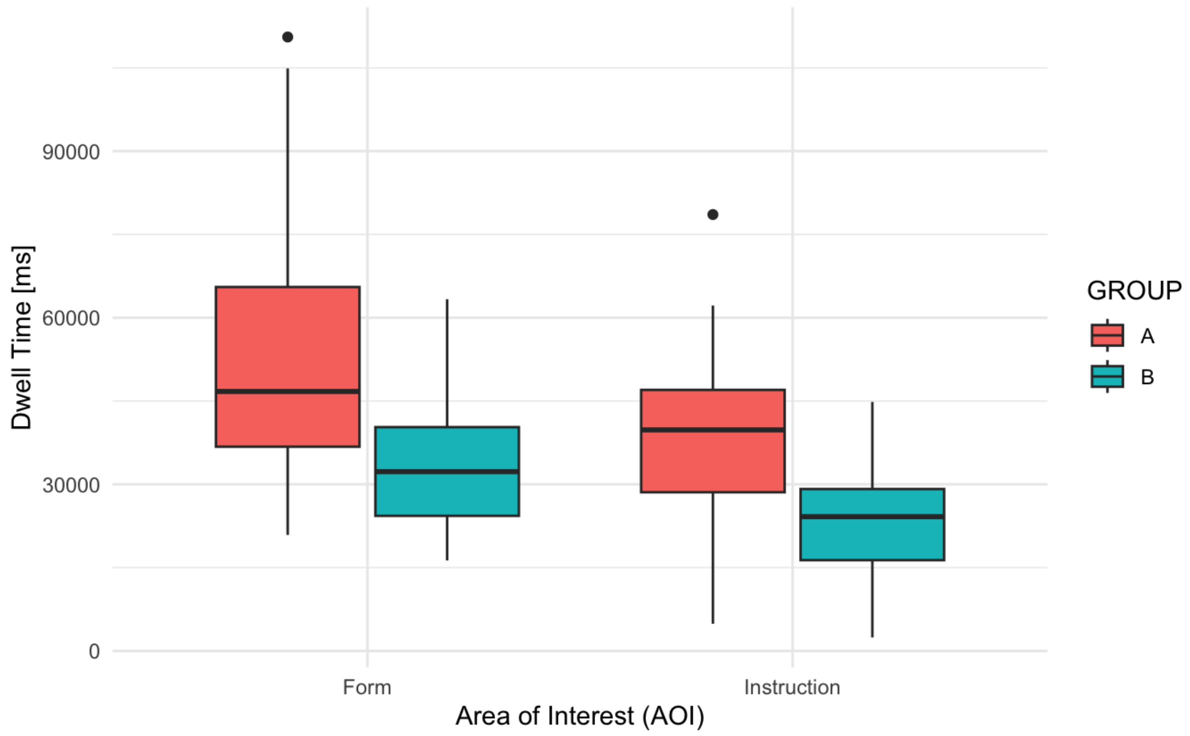

Fig. S6 – Part_6 – Tax Base Calculation

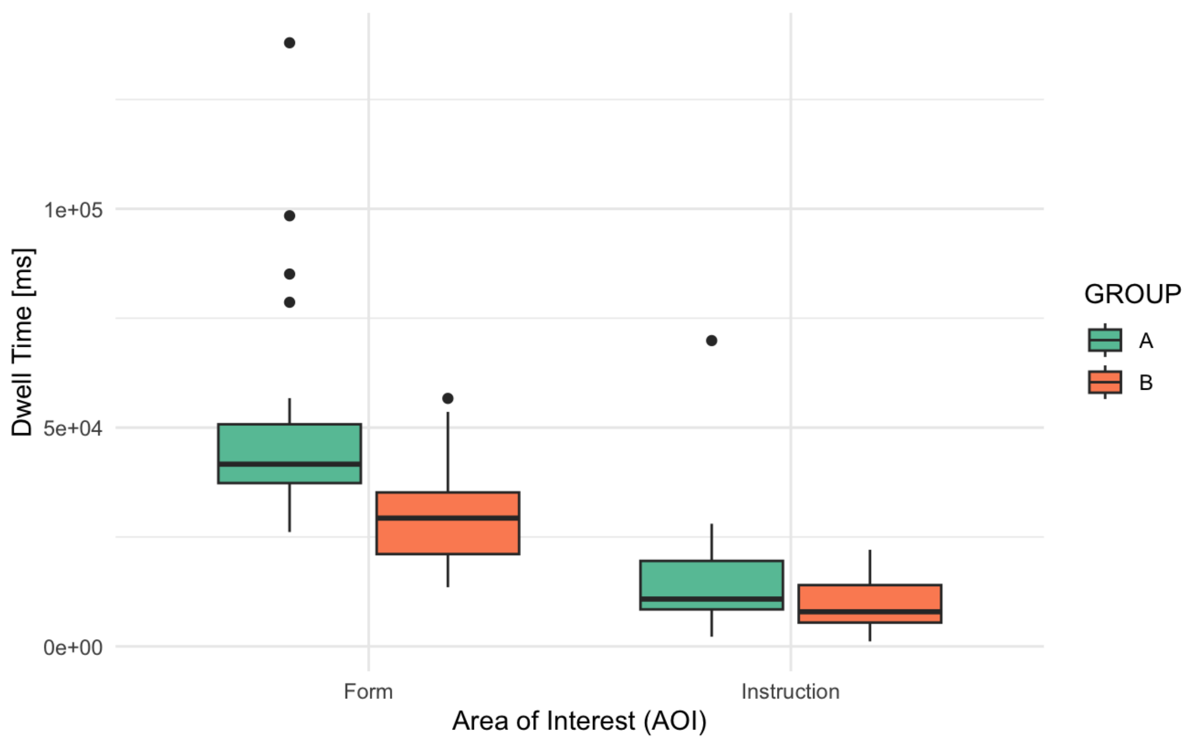

Fig. S7 – Part_7 – Tax Liability Calculation