



An experimental investigation of the use of nudges embedded in tax filing software to reduce error rates

Miguel A Fonseca
Shaun B Grimshaw
Christos Kotsogiannis

August 2018

EXECUTIVE SUMMARY

HMRC's Making Tax Digital for Business (MTDfB) initiative is aimed at reducing the amount of errors made by customers submitting updates and returns. It contains a novel proposal to present nudges to tax-payers making submissions highlighting potential mistakes and directing them to guidance or support (HMRC, 2017).

This report details results from an experiment designed to investigate the effectiveness of nudges on customers using an example tax filing software.

555 UK residents were recruited to undertake an online tax filing task in which they were paid to minimise the number of errors in their tax submission by having an amount deducted from a maximum possible payment of £10 for each mistake made.

Subjects were randomly assigned to one of several treatments which contained differences in whether and which nudges were displayed in a feedback report presented to them after their initial tax submission.

The presence of nudges in the feedback report led to raised levels of both resubmission and alteration of the tax return by subjects compared to treatments with no nudges, indicating that subjects responded strongly to the presentation of nudges to review and alter their tax submissions.

The presence of a series of responsive nudges in the feedback report, reflecting errors made by the subject, led to a significant increase in the subjects' average score of correctly entered values relative to the treatment with no nudges. A significant difference was observed for the correct re-categorisation of a number of specific items. The majority of the corrections were found to be correlated with the examination of guidance after the presentation of the nudges.

The presence of responsive nudges in the feedback report also led to a significantly higher level of the correction of category total values independent of the correction of the assignment of items to categories, arising from subjects rechecking the calculations within their tax forms.

The presentation of a nudge that provided subjects with a specific example of how the category should not be used was associated with a significant correction of the associated items. The use of a nudge that provided subjects with information that an entry was incorrect but without further information as to what the correct value should be caused a high degree of alteration, but not a significant degree of correction.

The majority of subjects self-reported finding the nudges in the feedback report useful. A number of subjects further commented that any feedback should be designed to directly assist in filing, not simply comment on mistakes.

Increased levels of resubmission and alteration after presentation of the feedback report were also observed for subjects presented with unintended nudges, relating to a correctly assigned or unused category, indicating the potential for unintended consequences associated with use of nudges in the tax form. Furthermore, there is evidence from a number of control treatments that the presentation of unintended nudges led to subjects introducing errors into their tax returns that were not present in their first submission.

The experimental results clearly indicate that subjects responded to the presence of nudges and reviewed their tax submission. The subsequent quality of their resubmission relates to whether the nudge was appropriate and if the nudge contained sufficient information to assist the subject to make an appropriate correction.

Acknowledgements

This work was conducted in the Tax Administration Research Centre at the University of Exeter, jointly funded by the Economic and Social Research Council, HM Revenue & Customs and HM Treasury (grant no. ES/K005944/1); we are very grateful to our funders for their support. The views expressed in this report are the authors' and do not necessarily reflect those of the funders.

1 Introduction

1.1 Overview

Estimates of the 2016-2017 tax gap show that £9.1 billion for the period was lost from avoidable taxpayer errors (HMRC, 2018).

Avoidable errors include items labelled as “Failure to take reasonable care” resulting from a customer's carelessness and/or negligence in adequately recording their transactions and/or in preparing their tax returns (£5.9 billion). The total also includes basic errors in tax tasks where taxpayers were deemed to have taken reasonable care (£3.2 billion).

The stated ambition of HMRC is to become one of the most digitally advanced tax administrations in the world, modernising the tax system to make it more effective, more efficient and easier for customers to comply (HMRC, 2017).

Businesses will use software that compiles their tax data as part of their ordinary day-to-day activity, potentially highlighting any possible errors (for instance, arithmetical mistakes or figures which look out of place) and offering prompts for information that might otherwise be overlooked (HMRC, 2015).

The aim of the study presented in this report is to examine the potential impact of measures presented through tax software after a customer's initial submission, designed to prompt willing taxpayers to be compliant.

The key questions the present study addresses are:

- Do nudges improve compliance/elicit a change in user behaviour?
- Do taxpayers ignore the nudges? If so, why?

We answer these questions by means of a laboratory experiment.

1.2 Tax Experiments

The current experiment adds to a large body of literature using experiments to study individual behaviour in a tax setting.

Experimental economists are generally agreed on a set of principles that govern the conduct of experiments, notably that; i) subjects should be incentivised in a salient manner, ii) there should be no deception in the experiment and iii) subjects should be randomly assigned to treatment conditions to provide statistical validity.

A treatment is a specific set of values for the experimental parameters, and the experimenters are interested in how a change in parameters affects behaviour.

Previous tax experiment studies have examined the effects of classic policy levers, such as audit and fine rates and of other policies, such as amnesties, audit policies and on compliance behaviour (reviewed in Malézieux, 2018).

Recent experimental studies have investigated the impact of the provision of taxpayer service on compliance (Vossler and McKee, 2017; McKee, Sildake and Vossler, 2018); the effect of different content and delivery of guidance (Chen, Grimshaw and Myles, 2017), and the potential outcomes of the use of pre-population and prompts in tax forms (Fonseca and Grimshaw, 2017).

A number of studies have examined the effect of experience of the tax system on compliance in tax experiments with some suggesting there is little difference (Alm, Bloomquist and McKee, 2015)

whereas others suggest that taxpayer samples are more compliant than student ones (Choo, Fonseca and Myles, 2016).

For the purpose of the current study, it was determined that a sample of UK resident non-students who are self-employed or consider themselves as an entrepreneur were more likely to have used or potentially to use the target software.

2 Experiment

2.1 Overview

The experiment required subjects to categorise and complete a digital tax return for a series of items of expenditure given in relation to the profile of a small business owner.

A number of the expenses items were chosen as they were found to be more likely to cause errors in pilot studies and to be in line with the types of errors commonly observed in tax returns.

The experimental design rewards subjects for being as accurate as possible in their return, such that subjects would focus on limiting the number of errors made in their return rather than non-compliance.

The incentive scheme for the experiment reflects this design, in that subjects were charged a fixed fee for each item entered incorrectly, rather than given tax, audit and penalty rates as is typical in many tax experiments. The experiment's incentive scheme is therefore consistent with a taxpayer who wishes to be compliant, but may make mistakes in the filing process. The mechanism of reducing the payment for an incorrect value means there is no incentive to be non-compliant.

The incentive scheme also means that each item in the profile has an equal weighting in the outcome, no matter what the material value given. Therefore subjects should focus on all items, rather than just the ones with the nominally highest values.

2.2 Design

The experiment was run online through a bespoke web site hosted by the University of Exeter. Figure 1 illustrates the steps undertaken by a subject through the experiment.

Upon completion of the experiment, subjects were debriefed and informed of how to withdraw from the study should they have wished to.

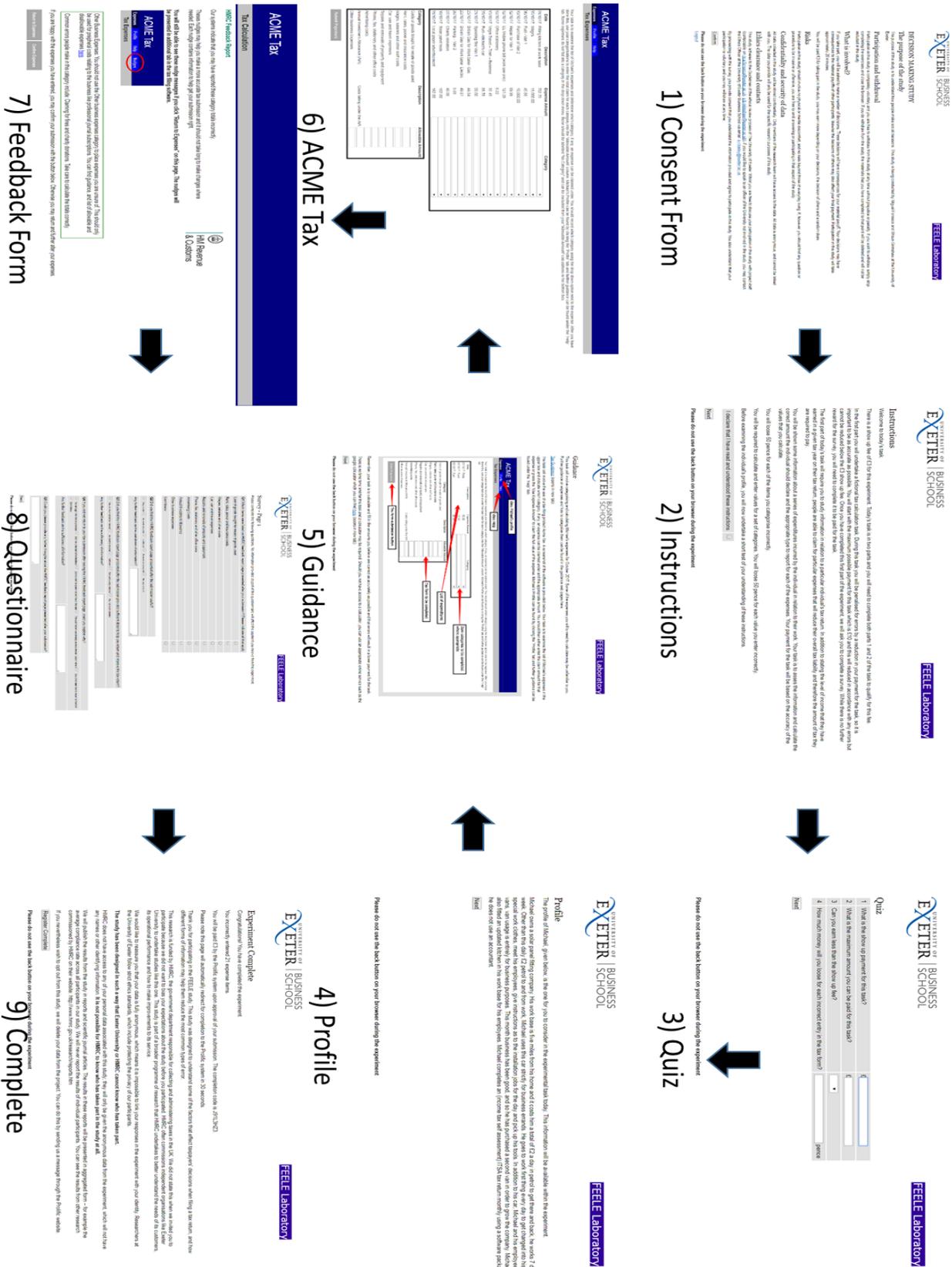


Figure 1: Illustration of experimental workflow

2.2.1 Instructions

The [instructions](#) detailed the primary task of the experiment, which would require subjects to study information in relation to a particular individual's tax return. Payment would be based on the accuracy of the taxpayer's assessment of expenses, in particular which types of expenses should be reported and the appropriate amount to declare.

The maximum possible payment was £10, with a guaranteed minimum of £3. Subjects were informed they would lose money for an incorrect response in a manner detailed in the Task sub-section.

2.2.2 Quiz

Upon confirmation of reading the instructions, subjects were presented with a [quiz](#) containing 4 questions relating to the incentive structure of the experiment, which they could not proceed past until all questions had been answered correctly.

2.2.3 Profile

Subjects were presented with the [profile](#) for Michael, the owner of a solar panel fitting company (see [Appendix](#)). The profile was accessible from within the tax filing software such that subjects could re-check the details (and subjects were informed of this).

2.2.4 Introduction to guidance and tax software

Subjects were informed that their task was to process Michael's expenses for October 2017.

This tax [guidance](#) largely replicated sections of the www.gov.uk HMRC tax guidance, but in a manner that each set of information (by category) was placed on its own web page such that software could track which items were accessed. The experimental software was able to track when subjects accessed guidance, in particular if it was before or after the first submission of their tax form.

Subjects were also shown an annotated screenshot of the ACME Tax software that they would use to complete the task.

2.2.5 Task

The form subjects were presented to make their choices in was designed in the style of an expenses processing App – "[ACME Tax](#)".

We employed two types of form, representing two different technologies:

Form Type	Characteristic	Incentive	Measure
Item	Subjects were required to characterise each of the items in the list of expenses and enter the claim amount alongside	Subjects were informed they would lose £1 per incorrectly entered item	Direct measure of ability to categorise items
CatTotal	Subjects were required to characterise each of the items in the list of expenses and enter the total claim per category in an additional section of the screen below	Subjects were informed that they would lose 50p per item incorrectly categorised and 50p for each category total that was incorrectly entered	Asses ability to categorise items and tests for consistency in netting across categories

Table 1: Details of the two types of tax form used in the experiment

2.2.6 ACME Tax

Subjects completed the task online on a page designed to replicate a simple expenses processing software App - “ACME Tax”. Subjects were informed that expenses that did not fall into a category should be labelled as “No category”. In addition to the expenses tab, a second tab allowed subjects to re-examine Michael’s profile, while a third tab gave links back to the [guidance](#) and [instructions](#).

2.2.7 Feedback Form

The “HMRC Feedback Report” was presented to subjects from within the ACME Tax App upon the submission of their tax return. In all cases this form had two buttons, one whereby subjects could return their tax return and potentially alter the details, and a second button whereby subjects could confirm their tax return on move on to the next state of the experiment.

The key treatment elements of the experiment lie in the contents of feedback report presented after the subjects submit their tax return for the first time. In conditions where nudges are presented, the feedback report states the HMRC systems indicate that the subject may have entered something incorrectly. The feedback form allows for the presentation of a nudge or nudges to the subject in relation to their submission.

Nudge	Detail
No	The feedback form contained no further information
FixOther	The feedback form always contained a nudge relating to the category “Other Business Expenses”
FixAia	The feedback for always contained a nudge relating to the category “Annual Investment Allowance (AIA)”
FixCost	The feedback from always contained a nudge relating to the category “Cost of Goods”
Res	The feedback form contained up to four nudges based on the differences between the category totals entered and the expected value, with the nudges placed in order of the greatest difference between the entered and appropriate value first

Table 2: Details of the nudge conditions used in the experiment

Table 2 lists the set of nudge conditions used in the experiment. The case of No (nudge) allow for a comparison on the effects where a nudge or nudges are presented in the HMRC Feedback report to where no nudges are presented.

In conditions with nudges, subjects were informed that the information would be also available to them during any subsequent tax filing. Subjects were told that the nudges would be available under a fourth, additional, tab in the ACME Tax software if they chose to return and alter their tax return (See Panel 7 of Figure 1 for further detail).

Subjects were able to return and alter their tax return as many times as they wished, but the “HMRC Feedback Report” only presented nudges after the initial submission and not in response to any further submissions.

Nudge	Detail
Annual Investment Allowance	<p>You should review what can apply for the acquisition of assets under the Annual Investment Allowance. You can find specific guidance here.</p> <p>Common errors people make in this category include: Claiming for renovation of business premises.</p>
Other Business Expenses	<p>You should not use the Other business expenses category to place expenses you are unsure of. This should only be used for peripheral costs relating to the business like professional journal subscriptions. You can find guidance, and a list of disallowable expenses here.</p> <p>Common errors people make in this category include: Claiming for fines and charity donations.</p>
Cost of goods bought for resale or goods used	<p>You should not include all goods bought, this section should only be used for goods that are a direct cost of production for goods sold. You can find specific guidance here.</p> <p>Common errors people make in the category include: Claiming for tools.</p>

Table 3: Details of key nudges used in the experiment

Links to guidance provided within nudges opened the same guidance page as was available before they submitted their return. Importantly, we were able to detect if the page was opened through the link in the nudge.

2.2.8 Questionnaire

Upon final submission of the tax form, subjects were presented with a questionnaire.

Subjects were informed that the answers they gave to the questionnaire would not affect the amount of their payoff, but that they would need to complete the questionnaire to receive payment.

Subjects who had received nudges were first presented with a series of questions about the nudges.

For details of all items in the questionnaire, see the [Appendix](#).

2.3 Prolific Sample

Subjects were recruited through the Prolific system: <https://prolific.ac> (Palan and Schitter, 2018).

The Prolific system contains a database of individuals who have signed up to partake in paid scientific studies. Further details of the system and recruitment are given in the [Appendix](#). All subjects recruited had declared themselves to be UK residents.

The majority of recruitment was based on two pools of potential subjects who had previously responded to questions in the Prolific system that could be used as pre-filters. Table 4 details the number of potential subjects associated with the two filters, and of a third pool of additional subjects.

Pool	Filter	Number of Potential subjects	% Male	% Female
1	Self-employed	1,045	25.6	61.8
2	Entrepreneur	1,520	30.2	65.8
3	No additional filter	14,870	30.0	64.8

Table 4: Details of the Prolific sample base

Upon completion of the experiment, Prolific offers a file of information related to the subjects used in the study – analysis of elements of this information (including counts of previous experimental outcomes on prolific) is included in the section relating to the [questionnaire](#) responses.

2.4 Treatments

Table 5 summarises the treatments and the number of subjects who completed each of the treatments

Treatment	Detail	Subjects	Nudges
ItemNo	Set categories only, no nudges	83	83 None
ItemFixOther	Set categories only, always display “Other Business Expenses” nudge	80	45 Responsive 35 Unintended
ItemFixAia	Set categories only, always display “AIA” nudge	84	71 Responsive 13 Unintended
ItemFixCost	Set categories only, always display “Cost of Goods” nudge	78	31 Responsive 47 Unintended
CatTotalNo	Set categories and total, no nudges	80	80 None
CatTotalFixOther	Set categories and total, always display “Other Business Expenses” nudge	74	50 Responsive 24 Unintended
CatTotalRes	Set categories and total, display up to four nudges based on totals entered	76	75 Responsive 1 None

Table 5: Details of the experimental treatments. Nudges column shows the number of subjects presented with nudges as i) None – no nudges due to treatment or fully correct initial submission ii) Responsive – reflecting incorrect entries in their tax return iii) Unintended – reflecting errors not made in their initial submission

Allocation of subjects to treatments was done randomly on a by arrival basis within each pool.

3 Results

555 subjects completed the experiment for an average payment of £6.07 for a median time of 18 minutes and 11 seconds (equivalent to £20.03 per hour).

78.2% of the sample answered the 4 questions of understanding on the first attempt and 93.3% by the second attempt.

The results focus on two key measures. The first is the change in the number of correctly entered values, where a positive value indicates an entry in the tax form that a subject has corrected, such as the reassignment of an item to a correct category or the correction of the total value entered for a specific category. The second measure is the change in the number of items assigned to the correct category.

3.1 Responsive Nudges

- **Result 1:** *The presence of responsive nudges in the feedback report lead to a significant increase in the subjects’ average score of correctly entered values.*

The average change in the number of correctly entered values within the final submission of the form in the CatTotal treatment without nudges was 0.05 compared to 1.07 for the treatment with nudges (T-test: $p < 0.001$). Subjects were both significantly more likely to resubmit their tax form in the CatTotalRes treatment (77.6%) compared to the CatTotalNo treatment (7.5%) and to make an alteration to a value (61.8% to 3.8%).

In the responsive nudges treatment, CatTotalRes, subjects received up to four nudges based on the errors in their initial tax submission. Table 6 shows that a small proportion of subjects performed very well in the initial submission and received a small number of nudges, while the majority of subjects received at least 3 nudges.

Number of Nudges	Percentage of Treatment
0	1%
1	4%
2	11%
3	17%
4	67%

Table 6: Proportion of subjects in CatTotalRes treatment receiving the number of nudges

Figure 2 illustrates which nudges were most commonly shown to subjects, corresponding largely to the set of errors in the assignment of categories discussed in more detail in a later [section](#) of the report. Figure 2 also shows that only a small proportion of subjects clicked on the link for “extra guidance” given in the nudge.

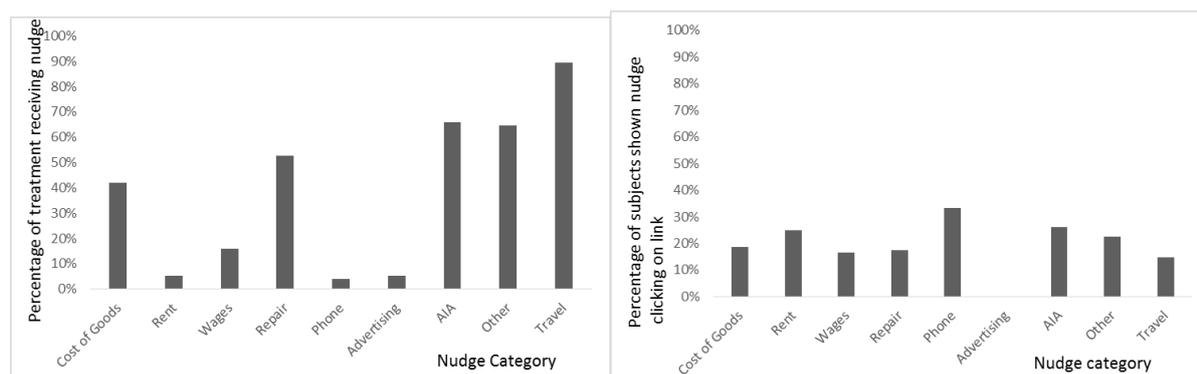


Figure 2: (Left) Proportion of CatTotalRes treatment being shown a nudge corresponding to a particular category; (Right) Proportion of subjects in CatTotalRes treatment shown a particular nudge that clicked on the provided link

3.2 Change in the assignment of items to categories

- **Result 2:** The presence of nudges in the feedback report lead to the reassignment of items to the correct category.

The left panel of Figure 3 shows that a considerable larger portion of subjects altered the assignment of a category for an item which was wrong in the initial submission compared to those who altered an assignment for an item that was correct in the CatTotalRes treatment. The right panel of Figure 3 highlights that a high proportion of those who altered the category of an item that was incorrectly assigned in the initial submission did so in a manner which corrected the assignment.

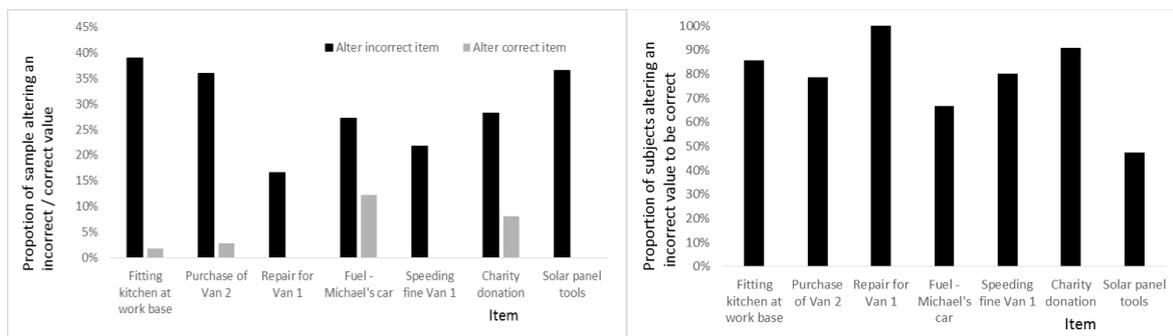


Figure 3: (Left) Proportion of CatTotalRes treatment sample who altered an item with an incorrect initial assignment, and percentage of CatTotalRes treatment sample who altered an item with a correct initial assignment; (Right) Proportion of subjects who altered an item with an incorrect initial assignment in the CatTotalRes treatment to be a correct assignment

Table 7 examines the effect of the CatTotalRes treatment by a comparison of the rate of reassignment of an item to the correct category in their final submission among subjects who had assigned the item incorrectly in their initial submission against the corresponding rate for the CatTotalNo treatment.

Table 7 further separates the comparison of correcting reassignment rates to subjects who had examined guidance prior to their initial submission and those that had not.

	Overall			No Prior Guidance			Prior Guidance		
	No	Res	Diff	No	Res	Diff	No	Res	Diff
Fitting kitchen at work base	0.0	35.0	***	0.0	10.0		0.0	60.0	***
Purchase of van 2	0.0	27.5	***	0.0	25.9	**	0.0	30.7	**
Repair for van 1	0.0	16.7	***	0.0	0.0		0.0	31.6	***
Fuel for Michael's car	0.0	18.2		0.0	0.0		0.0	25.0	
Speeding fine van 1	3.3	17.4		0.0	23.5	*	6.3	0.0	
Charity donation	0.0	25.6	***	0.0	10.5		0.0	40.0	***
Solar panel tools	0.0	17.3	***	0.0	18.8	**	0.0	15.0	*

Table 7: Proportion of subjects who reassigned to the correct category in their final submission having assigned the item to the wrong category in their initial submission by treatment (CatTotalNo and CatTotalRes), and by i) all relevant subjects and then separated by ii) those who had not assessed guidance prior to the initial submission and iii) those that had assessed guidance prior to the initial submission; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The majority of the correcting effect is found among subjects who accessed guidance prior to their initial submission, suggesting that the nudges worked to cause a proportion of such subjects to reassess their initial assignments, with a positive effect also found for the two AIA category items for subjects who had not accessed guidance prior to their initial submission.

Interestingly for the item “speeding fine for van 1”, the correction effect was only found for subjects who had not accessed guidance prior to their initial submission, though the effect is weak as the number of subjects under this condition is small.

A larger proportion of subjects in the CatTotalRes treatment (12.3%) changed their assignment of the item “fuel for Michael’s car” from the correct value to an incorrect one compared to the CatTotalNo (0%) treatment ($p < 0.01$).

This would suggest that the treatment is having the effect of causing some subjects to question their assignment of this item, and in most cases of the alteration they switch to no longer claiming for the item, rather than noting that it is the associated claim amount that should be adjusted.

There is mixed evidence on the effect of the individual nudges, as there are a number of confounds in that there is overlap in the set of nudges shown to subjects and differences as to which of the nudges a subject chose to click on.

The most significant factors detected in the reassignment of items to the correct category are shown in Table 8.

	View Guidance Post	Treatment
Fitting kitchen at work base	**	
Purchase of van 2	***	
Repair for van 1		**
Fuel for Michael’s car	**	
Speeding fine van 1		
Charity donation		***
Solar panel tools	***	

Table 8: Significant factors found to be correlated to the reassignment of items to correct categories in the CatTotalRes treatment, assessing if a subject clicked on the link for a piece of associated guidance and the presence of the associated nudge; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

For the subjects that were observed reassigning the item fuel for Michael’s car to an incorrect category, this behaviour correlates to the treatment – suggesting that some people see the nudge and take it to mean that they should remove the claim for fuel, rather than adjust it.

3.3 Category Totals

- **Result 3:** *The presence of nudges lead the correction of category total values independently of changes to the reassignment of items to categories.*

A regression of the overall score change on the number of correctly reassigned category items between the CatTotalNo and CatTotalRes treatments shows a significant treatment effect, indicating a decrease in the number of errors in the subjects upon resubmission independent of the category changes (see table [A9](#)).

This result is due to a small number of specific corrections of total values made by subjects across a variety of category fields.

A particular case is the field advertising, where a number of subjects (3) noted that they had failed to fill in the total value for the field in their first submission (having correctly assigned the local paper advertisement item to the Advertising category) and proceeded to correct the value.

The experimental design allows us to examine for any inconsistency between the values the subject entered for the totals per category and total value of the items they assigned to the category. Details for the analysis of the initial submission can be found later in the [report](#).

There is no evidence for a change in the overall degree of inconsistency in the category totals entered by subjects CatTotalRes treatment after the nudges.

This is in part because as well as corrections, further errors were made where subjects changed categories, and a whole new form of error, whereby a subject changes the assignment of a category of an object and includes the value in the new category but forgets to remove the value from the old category is introduced.

The corrections of total values in result 3 therefore do not reflect an overall increase in consistency, as new inconsistency is introduced elsewhere.

It is unclear from the results as to whether this effect is directly related to the component of the nudge that suggests subjects should check their totals, or a simple effect of the nudges causing subjects to review their returns in such a way that errors of this form are identified and corrected.

3.4 Categorisation of Items

3.4.1 Initial Submission

- **Result 4:** *The majority of subjects were able to assign the correct category to at least 10 items. Several items were observed to be more likely to be categorised incorrectly.*

The average number of correct categories assigned was 12.64 (of 16), the lowest number of correct categories assigned to items by subjects in the initial submission was 4, whereas the highest was 16.

97.7% of the sample assigned at least 10 items to the correct category in their initial submission.

Access to guidance prior to the initial submission was found to be a key factor in the number of items assigned correctly in initial submission (see Table [A2](#)).

Figure 4 shows that there is no significant difference in the average number of items assigned to the correct category by subjects who did not access the guidance in the Item and CatTotal treatments.

Figure 4 also shows that there is a significant increase in the average number of items assigned to the correct category in the Item treatments by subjects who did access the guidance, and the average number of items assigned to the correct category in the CatTotal treatments by subjects who accessed the guidance is significantly higher again.

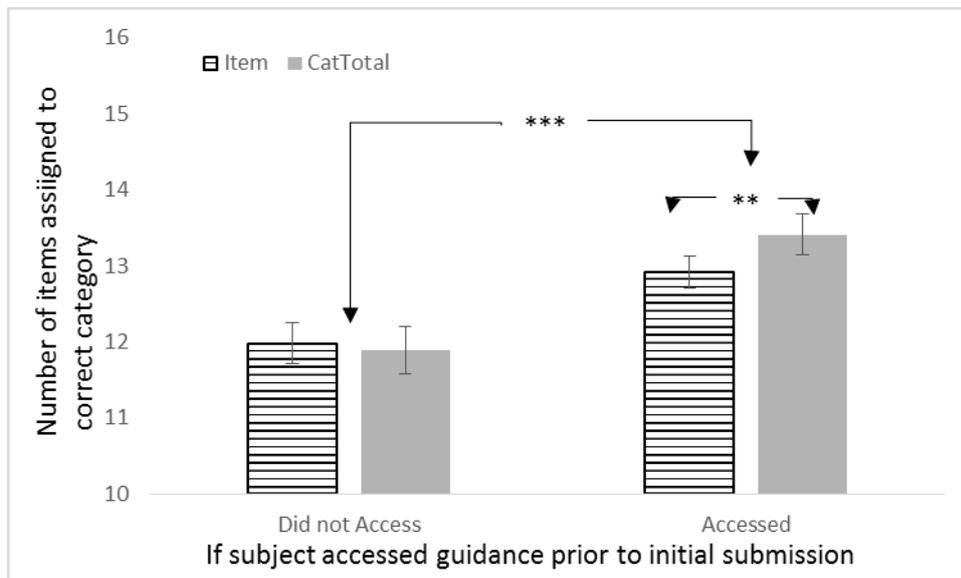


Figure 4: Average number of items assigned to the correct category in the initial submission split by i) if the subject accessed guidance prior to the submission and ii) Item and CatTotal treatments; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

While there is no significant overall difference in the proportion of subjects reading guidance before their initial submission between the CatTotal and Item treatments, there are differences for some particular categories.

In the majority of cases (for Car, van and travel expenses; Rent, rates, power and insurance costs; Wages, salaries and other staff costs; Advertising costs and Phone, fax, stationary and other office

costs) larger proportions of subjects read the guidance in the Item treatments compared to the CatTotal treatments.

However, a significantly higher proportion of subjects were observed to have read the AIA guidance in the CatTotal treatments (48.3%) to the Item treatments (27.4%).

The higher degree of access to the guidance for AIA in the CatTotal treatments may be due to the fact the additional component of the screen shows AIA as a category total that must be completed by the subjects.

Table 9 illustrates that there is a significant correlation between accessing the guidance and the correct characterisation of particular expenses items.

	Item			CatTotal		
	Not Accessed	Accessed	Diff	Not Accessed	Accessed	Diff
Fitting of kitchen at work base	72.0	77.5		72.5	78.0	
Wages	97.6	94.5		100.0	99.2	
Fuel – van 1	97.6	99.5		98.9	99.2	
Purchase of van 2	17.6	38.5	***	28.6	64.4	***
Repair for van 1	52.8	55.0		42.8	54.6	*
EE mobile phone	93.6	98.5	**	92.9	99.2	**
Office Stationary	96.8	97.5		97.9	98.5	
Virgin train travel	92.8	98.5	**	94.9	98.5	
Fuel for Michael’s car	89.6	89.5		84.7	82.6	
Speeding fine van 1	68.8	73.0		54.1	74.2	***
British Gas – Gas	86.4	94.5	**	95.9	98.5	
British Gas – Electric	86.4	94.5	**	95.9	98.5	
Parking van 2	82.6	94.0	***	84.7	97.0	***
Charity donation	64.0	63.5		43.9	53.0	
Solar panel tools	4.8	27.0	***	4.1	46.2	***
Local paper advert	95.2	96.0		96.9	99.2	

Table 9: Proportion of subjects correctly assigning the category for each item in the Item and CatTotal treatments in their initial submission based on whether or not they accessed guidance prior to the submission; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.4.2 Assignment of category errors in initial submission

Figure 5 summarises the data shown in Table 9, highlighting that subjects found a number of the items harder to categorise than others, in accordance with the pilot testing.

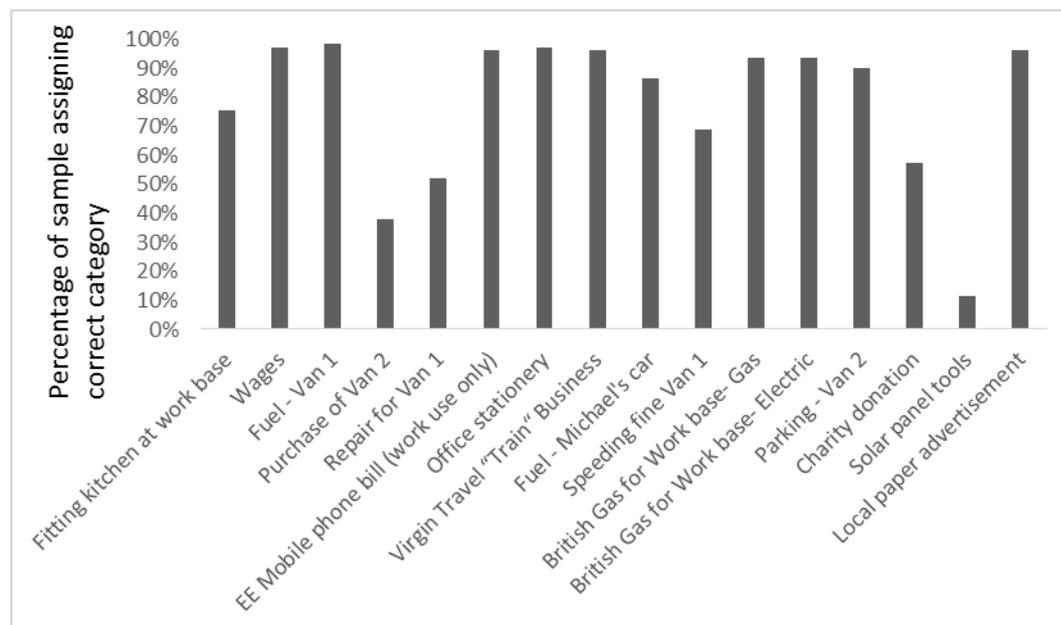


Figure 5: Percentage of sample assigning correct category to each item in the initial submission

	Adver	AIA	Travel	Cost	NoCat	Other	Phone	Rent	Repairs	Wages
Fitting kitchen at workbase		27%		18%	24%	20%	4%			5%
Purchase of van 2			60%	9%	12%	11%			9%	
Repair for van 1			93%		3%	2%				
Fuel – Michel’s car					82%	7%				5%
Speeding fine van 1			70%			22%				5%
Charity donation		3%				95%				
Solar panel tools				43%	4%	10%		3%	27%	

Table 10: Percentage of total incorrect assignments of an item assigned to a specific (incorrect) category (for items with correct assignment rates below 90%) – green shading shows category applicable for item

Table 10 shows the frequency of the incorrect categories used by subject for the most commonly incorrectly assigned items.

Table 10 highlights two categorisation errors that were very common, the assignment of the charity donation to the Other business expenses category and the assignment of the repair for van 1 to Car, van and travel expenses.

The item “speeding fine for van 1” lead to two common incorrect assignments to categories, to Car, van and travel expenses or to Other business expenses.

The item “purchase of van 2” leads to a common incorrect assignment to the category Car, van and travel expenses, but also to a number of other incorrect category assignments

The item “fitting kitchen at work base” leads to a wide spread of different categories being assigned, with no particular mistake the dominant one.

The common incorrect categorisation of the item “fuel for Michael’s car” arises from subjects deciding not to claim the item due to not being able to claim for travel to and from the place of work – as this is the only item where a partial amount is claimable, this item is discussed further later in the report.

Figure 6 shows a reasonable proportion of people (43.4%) simply claimed the total amount spent as car, van and travel expenses for the item “fuel for Michael’s car”.

A small proportion (11.6%) did not claim for the fuel item, whereas 14.1% entered the correct value.

The remainder either entered the wrong category (1%) or entered an incorrect value, where the most commonly entered values can be seen in Figure 6.

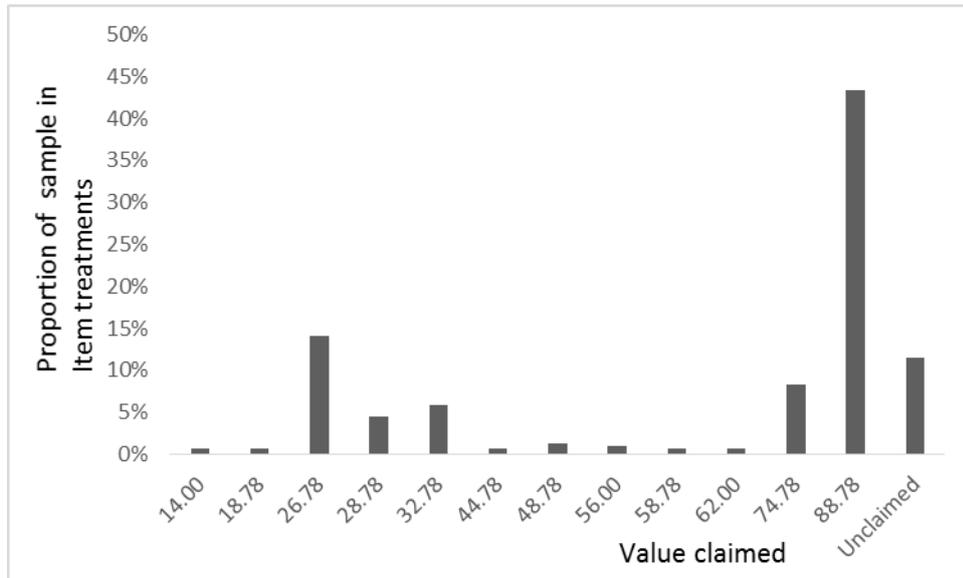


Figure 6: Proportion of sample in Item treatments claiming particular values (where more than 1% of subjects entered a specific value) for Car, van and travel expenses for the item “fuel for Michael’s car”

3.4.5 Final Submission

- **Result 5:** *The presence of nudges in the feedback report lead to raised levels of both resubmission and alteration of the tax return.*

62.9% of subjects resubmitted their tax form, with a significantly higher rate observed for treatments with nudges compared to those without, as illustrated in Figure 7.

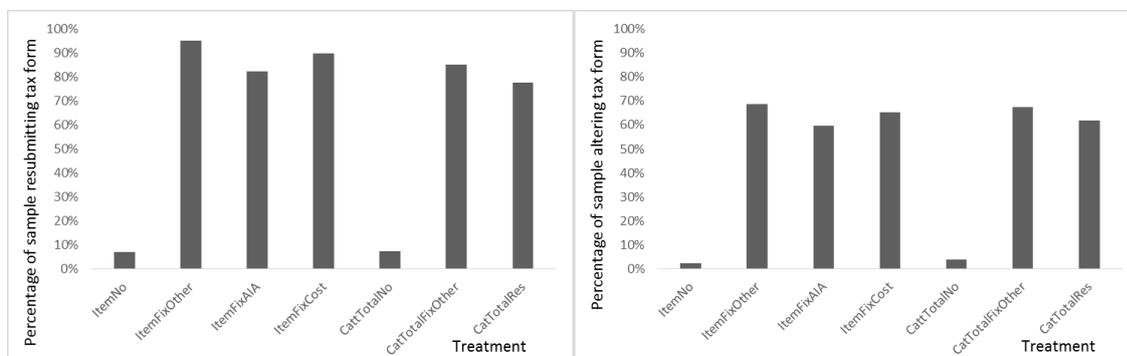


Figure 7: Proportion of subjects in a treatment who (left) resubmitted and (right) altered a category in their tax form by treatment

The alteration rates are lower than the resubmission rates, indicating that some subjects reviewed their tax returns but did not make any changes.

Figure 8 shows the net number of correct changes to the assigned category of items between the initial and final submissions.

Positive values in Figure 8 indicate that there are subjects who altered the (balance of) assignment of the category of items from an incorrect value to a correct one.

Negative values in Figure 8 indicate that there are subjects who altered the (balance of) assignment of the category of items from a correct value to an incorrect one.

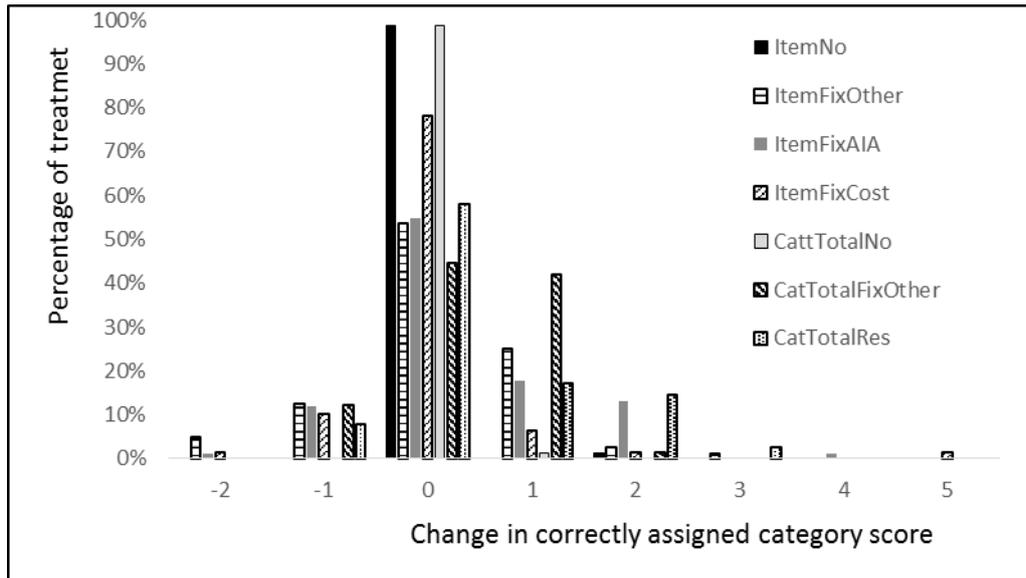


Figure 8: Percentage of treatment observed for a value in the change in the number of correctly assigned items

3.5 Item Treatments

- **Result 6:** *There is evidence that appropriate nudges may be associated with the correct reassignment of categories, whereas unintended nudges may be associated with switches from a correct assignment to an incorrect one. Not all nudges are equally effective.*

The treatments [ItemFixOther](#) and [ItemFixCost](#) generated nudges that could be deemed appropriate (in the sense that the nudge was presented to subjects who had erroneously categorised an item with the corresponding category) or unintended (in that the nudge was presented to subjects who had, correctly, not used the corresponding category).

The treatment [ItemFixAIA](#) is more complicated as there are multiple items for which AIA is the correct category, such that subjects may fail to assign the category, assign the category incorrectly or assign the category correctly.

Analysis of the categorisation changes should therefore depend upon the whether a subjects used the related category in their initial submission.

3.5.1 Other Business Expenses

Figure 9 illustrates that for subjects who did incorrectly assign an item to the Other business expenses category undertook a significant degree of correction in the ItemFixOther treatment compared to the ItemNo treatment (see Table A4).

Figure 9 also illustrates that for subjects who did not incorrectly assign an item to the Other business expenses category introduced a significant degree of error in the ItemFixOther treatment that contained a nudge, compared to the ItemNo treatment that did not.

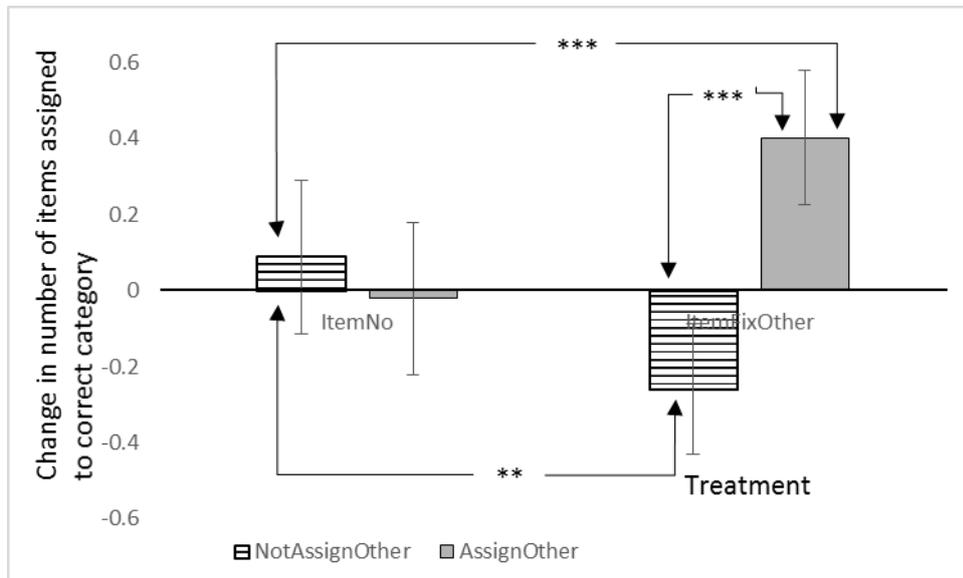


Figure 9: Change in number of items assigned to correct category between ItemNo and ItemFixOther treatments for subjects who did assign an item to the category Other business expenses in their initial submission and those who did not; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.5.2 Cost of Goods

Figure 10 illustrates there is no difference in the number of correctly assigned items between the ItemFixCost treatment and the ItemNo treatment for subjects who did assign an item to the category Cost of Goods and subjects who did not (see Table A5).

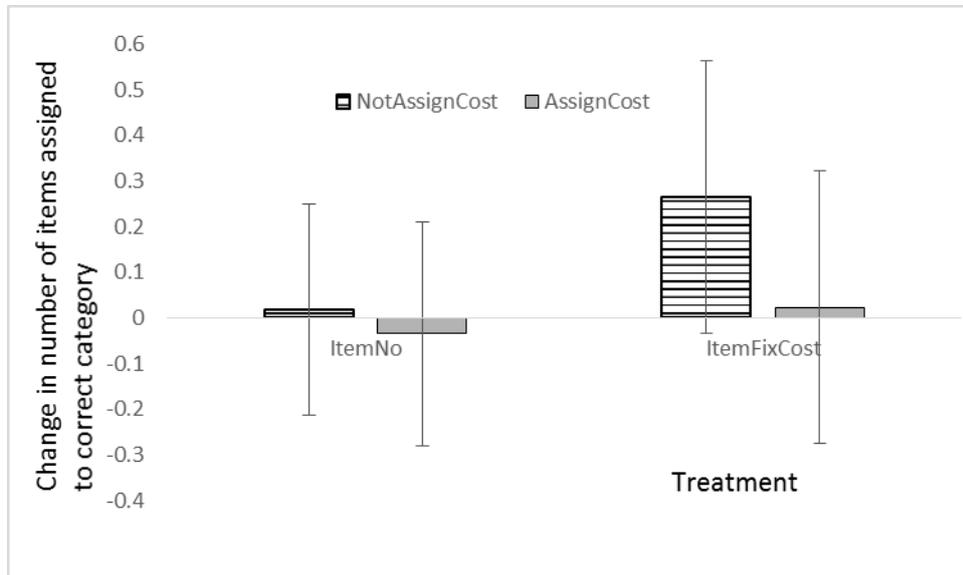


Figure 10: Change in number of items assigned to correct category between ItemNo and ItemFixCost treatments for subjects who did assign an item to the category Cost of goods in their initial submission and those who did not; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.5.3 AIA

Overall, there is a significantly higher increase in the number of correctly assigned items between the ItemFixAIA treatment and the ItemNo treatment by 0.321 points (T-test: $p=0.004$).

Figure 11 illustrates the pattern of how the change in the number of correctly assigned items differs in the ItemNo and ItemFixAIA treatments for subjects who used the AIA category in different ways in their initial submission (see Table A6); the corresponding sample proportions are detailed in Table 11.

		ItemNo	ItemFixAIA
BothCorrect	Both AIA items correct	10.8%	15.5%
AssignNotCorrect	Assign AIA, but not correctly	21.7%	23.8%
NotAssign	Did not use category AIA	67.5%	60.7%

Table 11: Percentage of treatment assigning the category AIA in different manners

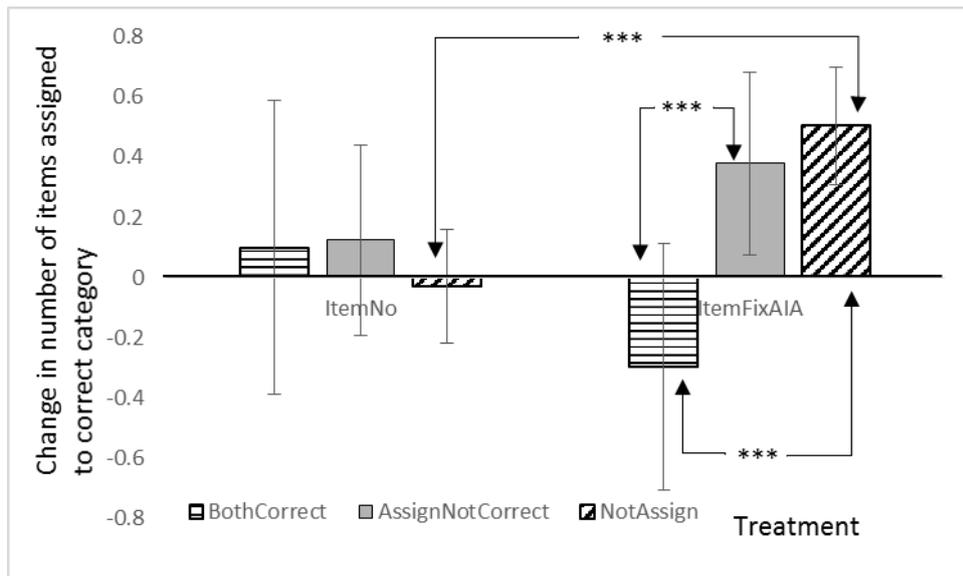


Figure 11: Change in number of items assigned to the correct category items between ItemNo and ItemFixOther treatments for subjects' different patterns of usage of the AIA category in their initial submissions; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 11 shows that the effect of the AIA nudge is complex; overall there is a positive impact, particularly among subjects who had not used the AIA category in their initial submission.

There is evidence for some subjects acting upon an unintended nudge and switching from correct entries to incorrect ones, though the outcome is not significant.

There is also evidence of an unintended consequence whereby some subjects reassign an unrelated item incorrectly to the nudge (AIA) category.

3.6 CatTotal Treatments

3.6.1 Initial Submission Inconsistency

- **Result 7:** A portion of subjects enter values for the category totals that are inconsistent with their assignment of categories to the expenses items.

The values entered by subjects for the category totals in the CatTotal treatments can be assessed for consistency against the predicted totals based on their category assignments. Four subjects who failed to enter any totals for the categories in their initial submissions are omitted from this section of the analysis.

Figure 12 illustrates the percentage of subjects found to be inconsistent for each of the category fields (other than Car, van and other travel expenses, which has a much higher degree of inconsistency arising from the variety of values utilised by subjects for the item fuel for Michael's car).

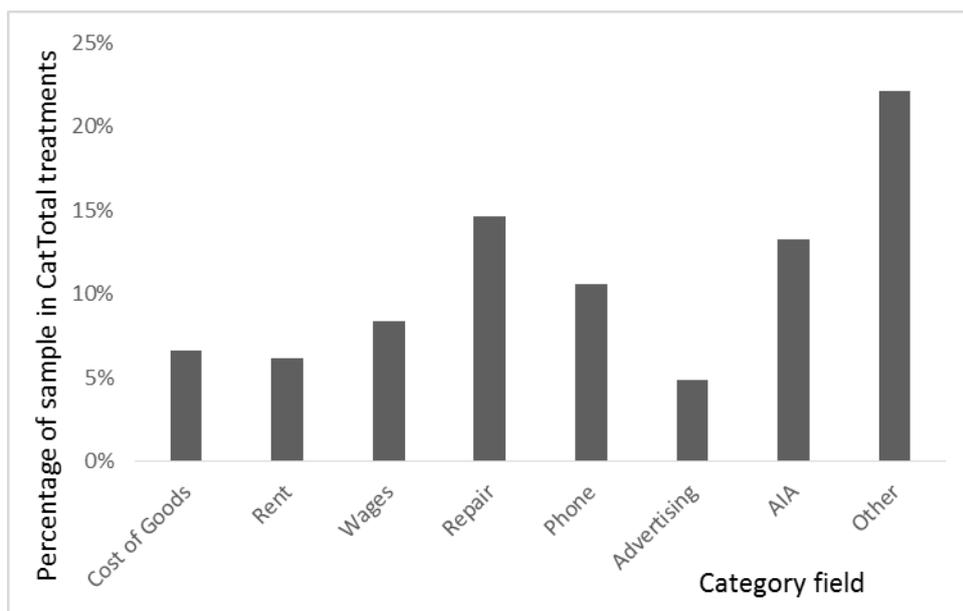


Figure 12: Percentage of sample in CatTotal treatments entering values for the field total inconsistent with the sum of their category assignments

8.4% of subjects entered an inconsistent value for the category field Wages, salaries and other staff costs. As this field only depends on a single value, arising from the item “wages”, it can be used to illustrate the most basic errors made by subjects:

- 58% of subjects entering an inconsistent value for the Wages category entered a value with some form of typing error.
- 32% of subjects failed to enter a value for that field having assigned a value for the category.
- 11% of subjects engaged in some other form of error, such as entering the value appropriate for a different field.

For the category fields with higher inconsistency rates (repair, phone, AIA, other), an additional form of calculation error can be observed, whereby the values entered represent outcomes corresponding to identifiable mathematical operations, typically being out by a unit suggesting an error in addition.

3.7 Questionnaire

Subjects were asked a number of questions about the experiment once they had completed their tax return but before being shown their outcome and payment (Details in [Appendix](#)). In addition, the Prolific system also provided information about subjects which is analysed here.

3.8 Sample characteristics

3.8.1 Prolific Extract Data

88.4% of subjects reported being born in the UK, with the next largest nationality being Poland (at 1.3%).

95.1% of subjects reported English as their first language, with the next highest language recorded being Polish (at 1.1%).

The median age of subjects was 38, with a range reported of 18 to 71.

51.1% of subjects were recorded as being in full-time employment, with 32.1% in part-time employment, 8.3% not in paid work (retired, home-keeping) and 2% unemployed and job seeking (with 6.5% reported as “other”).

67.8% of subjects were recorded as being female in the Prolific system – matching to the 65.3% of subjects who declared a gender in the experiment declaring it as female (though 13.7% of the sample did not express an answer to the gender element of the questionnaire).

Upon completion of a study, researchers have the opportunity to reject the submission of a subject if they believe that they have not made a sufficient effort, for example by failing to give the appropriate answer to an attention check question.

The Prolific system records the number of rejections a subject has received as well as the number of studies they have undertaken.

There was a high degree of variance in the number of previous experiments undertaken by subjects through Prolific with a median of 81 and a mean of 141 over a range of 1 to 1,414. Previous experimental experience was not found to correlate to outcomes in this experiment.

The mean acceptance rate (number of successful studies/total number of studies) for previous studies by subjects was 98.3% (median 99.2%) with a range between 78.6% and 100%.

3.8.2 Questionnaire Data

93.3% of subjects stated that their digital skill level was moderate or above.

81.2% of subjects agreed with the statement “I am comfortable using technology to manage my finances”.

96.2% of subjects reported using the internet more than once a day with an additional 3.1% reporting using the internet more than once a week.

43% of subjects reported working by themselves, 21% in mid-size organisations (between 1 and 500 people) and 11% of subjects reported working in organisations of over 500 people.

35.8% of the sample (46.5% of the subjects who responded) reported working by themselves in an organisation with a turnover of less than £100,000.

51.5% of subjects (66.8 of those that responded) stated working in an organisation with a turnover of less than £100,000.

7% of subjects declined to report whether they previously completed an ITSA tax form, whereas 43% reported they had not and 50% reported they had (7.8% one time; 14.8% 2-5 times; 16.6% 6-10 times; 10.8% more than 10 times).

Pool	% of total sample	% of pool declaring prior ITSA experience
1 (self-employed)	43.0	59.8
2 (entrepreneur)	44.0	47.1
3 (no additional Prolific filter)	13.0	26.4

Table 12: Details of characteristics of the recruitment pools used in the experiment

There are no significant differences between treatments for age, gender, internet usage, ITSA experience, technological comfort with finances or digital skill level, suggesting the randomisation to treatments worked efficiently.

3.8.3 Experimental Feedback

There is no difference between treatments in the reported level of difficulty of the task.

There is a significant decrease in the degree of difficulty reported by subjects who reported having submitted at least one previous ITSA form and with the subject's age.

There is a high degree of correlation between the set of nudges presented to subjects and their recall of them.

There are, however, a significant number of subjects who indicate having viewed a large number of nudges.

For instance, in the Item treatments with only one nudge, 27.7% of subjects reported more than one nudge, whereas 9.5% of subjects did not report observing the nudge.

This observation suggests that a number of subjects answered this question in terms of the guidance they had read, rather than the nudges presented in the feedback report.

A majority of subjects in the CatTotalRes treatment expressed that the nudges were useful, as illustrated in Figure 13.

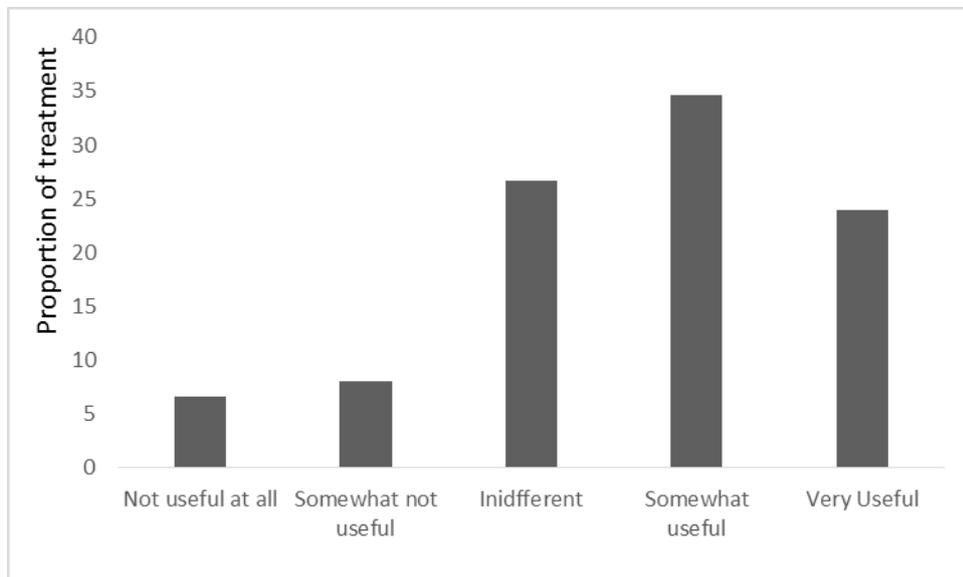


Figure 13: Proportion of CatTotalRes treatment reporting by category on the question "Did you find the HMRC feedback report page presented after the submission useful?"

A majority of subjects in the CatTotalRes treatment reported that the items in the nudge HMRC feedback report provided sufficient detail.

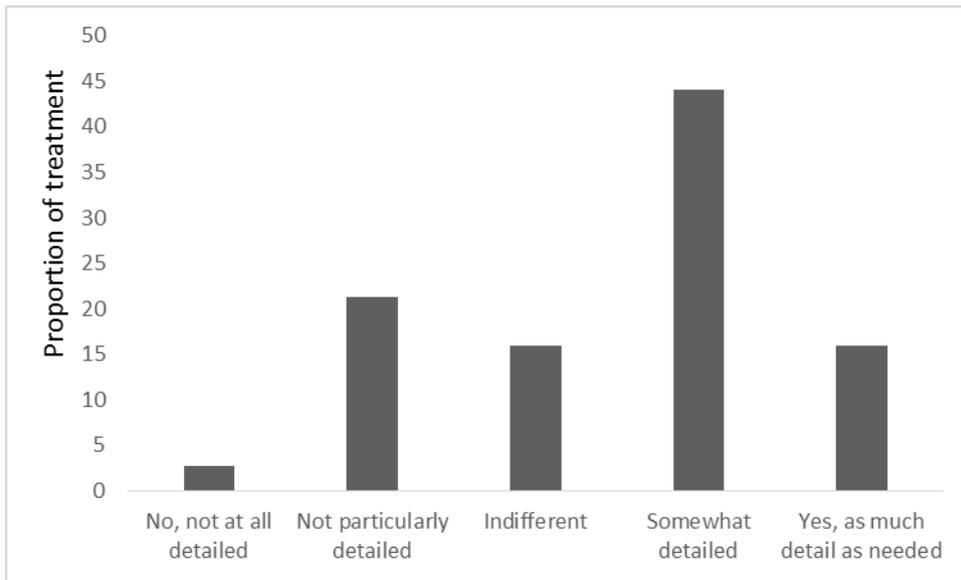


Figure 14: Proportion of CatTotalRes treatment reporting by category on the question “Did you find the HMRC feedback report page presented after the submission provided sufficient detail to help you return and improve the tax return?”

A theme among the comments in relation to the question regarding sufficient detail in the HMRC feedback report is to request for “more specific” information, for “direct detail” and more information as to what the right category should be, rather than just stating that something is wrong.

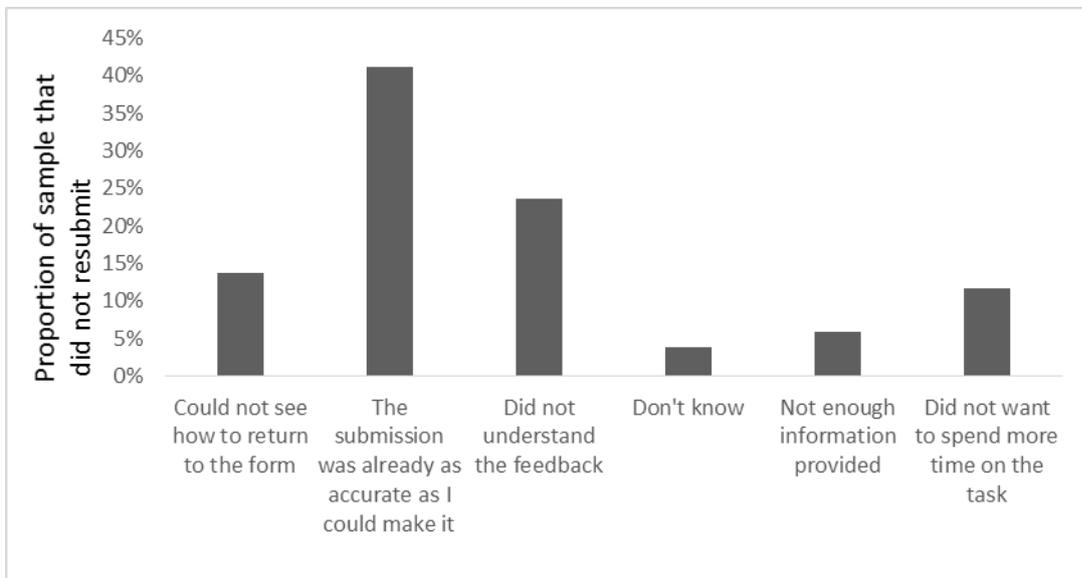


Figure 15: Proportion of sample that did not resubmit responding by category to the question “If you did not return to your tax submission after seeing the HMRC feedback report page, could you explain why?”

The primary reasons given by subjects who did not return the tax form after being presented with the HMRC feedback form were that the submission was already as accurate as possible and that the subject did not understand the feedback.

In open comments about the HMRC feedback report, subjects generally commented that the form was “helpful”, though some commented that the language could be simpler.

Several subjects questioned if the absence of nudges on their second submission meant that they had now corrected their tax return.

4. Discussion

The experimental results make clear that nudges presented in the HMRC feedback lead subjects to review and change their submissions. The impact on the outcome is less clear. Appropriate nudges were found to reduce the level of error in submissions, but unintended nudges were found to increase the level of error. Different individual nudges were found to have different levels of effect.

There is clear evidence that particular nudges presented in the HMRC feedback report (notably relating to Other business expenses and AIA) to subjects who had a related mistake allowed some fraction of the subjects to identify the error and correct it in their second submission.

There is also evidence that unintended nudges, in the sense of a nudge presented to a subject who had not used the corresponding category in an incorrect manner, may have caused some subjects to switch the assignment of an item from a correct value in their initial submission to an incorrect one in their second submission.

The experiment reveals that while a nudge may cause subjects to review their tax report, the subject may be still unable to determine the appropriate correct value from a particular nudge (notably Cost of Goods) and simply switch from one incorrect value to another.

Comments made by subjects in the post experimental questionnaire back up this observation, suggesting that the nudge should contain “direct detail” and be “more specific” to be more helpful.

Furthermore a number of subjects raised questions about the absence of nudges after the second submission of their tax form, in particular questioning if this absence indicates that the changes they had made were correct and that no further alteration was required.

A potential issue with the results presented in this report is that the key behaviour is an experimental artefact. Of particular concern is that the key improvements observed in response to the nudges arise from subjects who were insufficiently attentive during the first submission and then acted upon the nudge. Previous studies have presented evidence of subjects giving responses that are satisfactory for progress rather than reliable ones when responding to questions that require substantial cognitive load (Krosnick, 1991). While there is some evidence for the presence of a small quantity of subjects who were inattentive during the first submission in the sample, indicated by low numbers of correct entries in the initial submission, the results are robust to their exclusion, indicating that the findings are not due to such subjects.

The inclusion of the element of the task that requires subjects to assign categories to items in the CatTotal treatments may be considered artificial in the context of many tax filing applications. It is however a useful abstract step within the experiment as it allows for a clearer understanding of the subjects’ actions. It allows us to clearly separate the process of the assignment of categories from the calculation of totals, which would be impossible to do without the separate capture of the steps.

A second abstraction in the experiment, that the incentive scheme used in the experiment only penalises error and does not offer any reward for non-compliance, means that the outcomes should only be considered in terms of taxpayers who wish to be compliant. This limitation is nonetheless useful as it means that the results are not confounded by non-compliant behaviour, whereby it is not necessarily possible to separate error from evasion.

References

- Alm, J. Bloomquist, K. and McKee, M., 2015, *On the external validity of laboratory tax compliance experiments*, *Economic Inquiry*, 53 (2), 1170-1186
- Chen, J, Grimshaw, S. B. and Myles, G. D., 2017, "Testing and Implementing Digital Tax Administration" in *Digital Revolutions in Public Finance*, Eds Gupta, S., Keen, M., Shah, A. and Verdier, G., 2017, International Monetary Fund, Washington DC
- Choo, L., Fonseca M. A. and Myles, G. D., 2016, *Do students behave like real taxpayers in the lab? Evidence from a real effort tax compliance experiment*, *Journal of Economic Behavior & Organization*, 124, 102-114.
- Fonseca, M. A. and Grimshaw, S. B., 2017, *Do behavioral nudges in pre-populated tax forms affect compliance? Experimental evidence with real taxpayers*, *Journal of Public Policy and Marketing*, 36 (2), 213-226
- Krosnick, J. A., 1991, Response strategies for coping with the cognitive demands of attitude measures in surveys, *Applied Cognitive Psychology*, 5 (3), 213-236
- HMRC, 2018, Measuring tax gaps 2018 edition, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/715742/HMRC-measuring-tax-gaps-2018.pdf, accessed 16-Jul-2018
- HMRC, 2017, Making Tax Digital for business, <https://www.gov.uk/government/publications/making-tax-digital-for-business/making-tax-digital-for-business>, accessed 16-Jul-2018
- HMRC, 2015, Making Tax Digital, <http://webarchive.nationalarchives.gov.uk/20170605162751/https://www.gov.uk/government/publications/making-tax-digital>, accessed 16-Jul-2018
- Malézieux, A., 2018, *A practical guide to setting up your tax experiment*, *Journal of Tax Administration*, 4 (1), 107-127
- Palan, S. and Schitter, C., 2018, *Prolific.ac – A subject pool for online experiments*, *Journal of Behavioural and Experimental Finance*, 17, 22-27
- Vossler, C. A. and McKee, M., 2017. *Efficient tax reporting: The effects of taxpayer liability information services*, *Economic Inquiry*, 55 (2), 920–940
- Vossler, C. A., Siladke, C. A. and McKee, M., 2018. *Behavioral dynamics of tax compliance when taxpayer assistance services are available*, *International Tax and Public Finance*, 25 (3), 722-756

Appendix

Prolific

Subjects are only identified to researchers through their unique code within the Prolific system (Prolific Id), such that responses are anonymous to the experimenter.

Payments for completed studies are made to subjects' bank accounts through Prolific upon completion of studies.

The system currently holds in excess of 50,000 pre-registered subjects; over 27,000 of whom have undertaken at least one of in excess of 8,000 studies conducted on the platform (as of 2017) (Palan and Schitter, 2018).

Upon signing up to the system (or at any point thereafter) users may answer any of a series of questions relating to their personal characteristics, such as age, gender and employment status.

Prolific users are informed that their answers to the questions will enhance their opportunity to undertake studies as researchers will use the responses to filter subjects for their particular requirements.

Researchers advertise studies on Prolific for subjects to sign up to by providing a short description of the nature of the study, basic payment details and the expected time frame.

Researchers can restrict the set of potential subjects that a study is shown to (and therefore who can sign up for) using pre-filters of the previously entered user information.

The study reported here was advertised as "Earn up to £10 through a decision making task".

The Prolific system requires a minimum wage of £5 per hour, based on an expected completion time and a guaranteed payment amount.

The study presented here was presented with a guaranteed payment amount of £3 for an expected time of 30 minutes.

The actual median (and mean) completion times were less than 30 minutes.

Upon completion of the experiment, the software redirects the subject to the Prolific system with a completion code, so that the researcher can verify the subject's submission.

Payment of the guaranteed amount is automatic upon the researcher confirming the submission.

The remainder of the incentive for the experiment, up to the £10 maximum payment (where earned), was made through Prolific in the form of a "bonus" payment.

In the current study, pre-filters for residency in the UK and a student status of No were always applied.

Two further filters were used (separately) to recruit the majority of the subjects, a first for subjects who had reported themselves to be self-employed, a second for those who were entrepreneurs.

There is a bias towards toward females in the potential sample.

This appears to be a bias within Prolific, where 51.6% of potential UK resident subjects report being female, only 23.6% report being male, leaving 24.8% of subjects do not report their gender.

While the proportion reporting their gender increase once the additional filters are applied (see table), there is clear bias towards female Prolific users.

As yet there have been no studies published as to the nature of the Prolific sample. Previous studies have shown that other online recruitment platforms, such as Amazon Mechanical Turk, are more representative of the internet population than student samples (Palan and Schitter, 2018).

Subjects were recruited in batches of 35 over a period of several days, with each batch completing before a new one was launched.

The links to the website were given different parameters for the different pre-filters allowing for identification of the different pools in the results data.

The majority of batches were conducted with the first two pools given in the table (43.1% and 44.0% of subjects respectively), with the final batches using the third, No additional filter, pool undertaken to top up subject numbers.

Prolific Pre-screening Questions

Residency: *In what country do you reside?* United Kingdom

Student: *Are you a student?* No

Additional filters:

Career Status: *Where are you in terms of your career right now?* I am self-employed

Entrepreneurship: *Have you engaged in entrepreneurship/run your own business?* I am currently doing this

Gender: *What sex were you assigned at birth, such as on an original birth certificate?*

Experiment

Consent

DECISION MAKING STUDY

The purpose of the study

The purpose of this study is to understand how people make social decisions. This study is being conducted by Miguel Fonseca and Shaun Grimshaw at the University of Exeter.

Participation and withdrawal

Participation in this study is completely voluntary and you are free to withdraw from this study at any time without prejudice or penalty. If you wish to withdraw, simply stop completing the exercises and close the browser. If you do withdraw from the study, the materials that you have completed to that point will be deleted and will not be included in the study.

What is involved?

If you take part, you will be asked to make a number of decisions. These decisions will have consequences for your material payoff. Participation in this study will take approximately 30 minutes.

You will be paid £3 for taking part in the study; you may earn more depending on your decisions.

Risks

Participation in this study should involve no physical or mental discomfort, and no risks beyond those of everyday living. If, however, you should find any question or procedure to be invasive or offensive, you are free to omit answering or participating in that aspect of the study.

Confidentiality and security of data

All data collected in this study will be stored confidentially. Only members of the research team will have access to the data. All data is anonymous, and cannot be linked with you. The data you provide will only be used for the specific research purposes of this study.

Ethics clearance and contacts

This study adheres to the Guidelines of the ethical review process of The University of Exeter. Whilst you are free to discuss your participation in this study with project staff (contactable on m.a.fonseca@exeter.ac.uk; s.b.grimshaw@exeter.ac.uk), if you would like to speak to an officer of the University not involved in the study, you may contact the Ethics Officer at the University of Exeter Business School via email: a.r.bailey@exeter.ac.uk.

By proceeding with the survey, you provide consent that you understand the information provided and agree to participate in this study. You also understand that your participation is voluntary and you may withdraw at any time.

Identification

FEELE: Prolific

Welcome to the FEELE lab.

Please enter your prolific ID in the box below to access the experiment.

Prolific ID	<input type="text" value="59a57efccf"/>
-------------	---

Instructions

Welcome to today's task.

There is a show up fee of £3 for this experiment. Today's task is in two parts and you will need to complete both parts 1 and 2 of the task to qualify for this fee.

In the first part you will undertake a fictional tax calculation task. During this task you will be penalised for errors by a reduction in your payment for the task, so it is important to be as accurate as possible. You will start with the maximum possible payment for this task which is £10 and this will be reduced in accordance with any errors but cannot be reduced below the £3 show up fee. Once you have completed this first part of the experiment, we will ask you to complete a survey. While there is no further reward for the survey, you will need to complete it to be paid for the task.

The first part of today's task will require you to study information in relation to a particular individual's tax return. In addition to stating the level of income that they have earned in a given tax year on their tax return,

people are able to claim for particular expenses that will reduce their overall tax liability and therefore the amount of tax they are required to pay.

You will be shown some information about a series of expenditures incurred by the individual in relation to their work. Your task is to assess the information and calculate the correct amount the individual should declare and the appropriate type to report for each of the expenses. Your payment for the task will be based on the accuracy of the values that you calculate.

You will lose 50 pence for each of the items you categorise incorrectly.

You will be required to calculate and enter values for a set of categories. You will lose 50 pence for each value you enter incorrectly.

Before examining the individual's profile, you will now undertake a short test of your understanding of these instructions.

I declare that I have read and understood these instructions



Quiz

- 1 What is the show up payment for this task?
- 2 What is the maximum amount you can be paid for this task?
- 3 Can you earn less than the show up fee?
- 4 How much money will you lose for each incorrect entry in the tax form?

Profile

The profile of Michael, given below, is the one for you to consider in the experimental task today. This information will be available within the experiment.

Michael owns a solar panel fitting company. His work base is five miles from his home and it costs him a total of £2 a day in petrol to get there and back, he works 7 days a week. Other than this daily £2 petrol to and from work, Michael uses this car strictly for business errands. He goes to work first thing every day to get changed into his special work clothes, meet his employees, give instructions as to the installation jobs for the day and pick up his tools. In addition to his car, Michael and his employees use vans, van usage is entirely for business purposes. This month business has been good, and so he has purchased a second van in order to grow the company. Michael has also fitted an updated kitchen in his work base for his employees. Michael completes an (income tax self assessment) ITSA tax return monthly using a software package, he does not use an accountant.

Guidance

This task will involve categorising and calculating Michael's expenses for October 2017. Some of the expenses you will need to calculate may be unfamiliar to you. Further guidance on expenses and how to categorise them can be found in the guidance web pages here:

[Tax Guidance](#) (opens in new tab)

The task will involve the use of a tax filing product Acme Tax. A screenshot of the software is provided below. Your task is to examine the list of Michael's expenses in the upper box and indicate which category, if any,

an expense can be claimed under and the appropriate amount. You should then either enter the claim amount for that expense or press the "Use Expense Amount" to claim the full value of the expense. Michael's Profile can be found by clicking the "Profile" tab and further guidance can be found under the "Help" tab.

[annotated screen shot of ACME Tax placed here]

Remember, your task is to calculate and fill in the amounts you believe are correct as accurately as possible and that errors will result in a lower payment for the task.

There is no time limit for this task and a calculator may be required. Should you not have access to a calculator, you can use an appropriate online service such as the google calculator which you can find [here](#) (opens in new tab).

Tax Guidance

Homepage

Use these notes to help you fill in your tax return

Business expenses Allowable and disallowable expenses

Expenses can vary from business to business. Some are allowable for tax purposes, others aren't. Allowable expenses include the cost of goods sold or stock used, wages, rent, lighting, postage, phone calls and motor costs such as fuel and insurance or a flat rate amount and any adjustment expenses that arise on a change from cash basis to traditional accounting basis this year.

They don't include:

- your own salary, wages or drawings, National Insurance contributions
- the cost of entertaining
- depreciation
- losses on assets

Some expenses are only partly allowable. For example, you can only claim the business part of the costs of using your own car or using a room in your home as your office.

Please see the links below for specific information on categories of expenses

Car, van and travel expenses

This includes car and van insurance, fuel, parking, hire charges, vehicle licence fees, motoring organisation membership, train, bus, air and taxi fares, hotel room costs and meals on overnight business trips.

Non-business (private) motoring costs, fuel costs directly between home and work, fines, costs of buying vehicles, restricted lease rental expenses and other meals are disallowable expenses.

For example, imagine an individual who travels between home and work costing a total of £30 a month in fuel for his car. Other than this, all fuel costs associated with this car are for business use. If this individuals total fuel spend is £100 for the month, then his allowable claim would be £100 - £30 = £70.

Annual Investment Allowance

You can claim Annual Investment Allowance (AIA). From 1 January 2016 the maximum annual amount of AIA is £200,000. If you use the equipment for both business and private use, you'll need to reduce the Annual Investment Allowance (AIA) you claim by the private use proportion.

Example

1

Gordon buys tools for £5,000 and a van costing £10,000. As the total cost is less than his maximum AIA entitlement for the year, Gordon could claim the full amount as AIA. The tools are for business use only. However, Gordon uses the van 60% for business and 40% for private motoring so he must reduce the amount of AIA he claims on the van to reflect his private use. The AIA he can claim for the van is £6,000 (£10,000 less 40% private use). Gordon's total AIA claim is £11,000 (£5,000 for the tools plus £6,000 for the van). If Gordon claimed AIA for items such as tools or a van and he later sells (disposes of) those items, he may need to pay back part of his allowance. This is a balancing charge.

ACME Tax System

Item	Amount	Category	Allowable
Fitting kitchen at work base	702.75	Repairs and renewals of property and equipment	702.75
Wages	13,000.00	Wages, salaries and other staff costs	13,000.00
Fuel – Van 1	45.88	Car, van and travel expenses	45.88
Purchase of Van 2	10,000.00	Annual Investment Allowance (AIA)	10,000.00
Repair for Van 1	89.99	Repairs and renewals of property and equipment	89.99
EE Mobile Phone Bill (work use only)	121.39	Phone, fax, stationary and other office costs	121.39
Office stationary	8.22	Phone, fax, stationary and other office costs	8.22
Virgin Travel – Train – Business	32.43	Car, van and travel expenses	32.43
Fuel – Michael's Car	88.78	Car, van and travel expenses	26.78
Speeding fine Van 1	50.00	No Category	0.00
British Gas for Work base – Gas	45.54	Rent, rates, power and insurance costs	45.54
British Gas for Work base – Electric	49.81	Rent, rates, power and insurance costs	49.81
Parking – Van 2	5.05	Car, van and travel expenses	5.05
Charity Donation	40.00	No Category	0.00
Solar panel tools	102.00	Annual Investment Allowance (AIA)	102.00
Local paper advertisement	162.00	Advertising costs	162.00

Table A1: Details of the expenses items used in the experiment

Item List and categorisation component

ACME Tax

[Expenses](#)
[Profile](#)
[Help](#)

Tax Expenses

Your task is to examine the list of Michael's expenses and determine which category, if any, an expense can be claimed under. You should record which category using the drop down option next to the expense. After done this, use your categorisations to enter the total claim for each category into the "Allowable Amount" column. Michael's Profile can be found by clicking the "Profile" tab and further guidance can be found under the tab. Some expenses do not fall into a category in the drop down menu, these should be labelled "No Category" and can be excluded from your "Allowable Amount" calculations in the bottom box.

Date	Description	Expense Amount	Category
01/10/17	Fitting kitchen at work base	702.75	<input type="text"/>
05/10/17	Wages	13,000.00	<input type="text"/>
06/10/17	Fuel - Van 1	45.88	<input type="text"/>
10/10/17	Purchase of Van 2	10,000.00	<input type="text"/>
12/10/17	Repair for Van 1	89.99	<input type="text"/>
12/10/17	EE Mobile phone bill (work use only)	121.39	<input type="text"/>
12/10/17	Office stationery	8.22	<input type="text"/>
15/10/17	Virgin Travel - Train - Business	32.43	<input type="text"/>
20/10/17	Fuel - Michael's car	88.78	<input type="text"/>
21/10/17	Speeding fine Van 1	50.00	<input type="text"/>
23/10/17	British Gas for Work base- Gas	44.54	<input type="text"/>
23/10/17	British Gas for Work base- Electric	49.81	<input type="text"/>
26/10/17	Parking - Van 2	5.05	<input type="text"/>
27/10/17	Charity donation	40.00	<input type="text"/>
30/10/17	Solar panel tools	102.00	<input type="text"/>
31/10/17	Local paper advertisement	162.00	<input type="text"/>

Category Description Allowable Amount

Category total entry component

31/10/17	Local paper advertisement	162.00	<input type="text"/>
----------	---------------------------	--------	----------------------

Category	Description	Allowable Amount
Costs of goods bought for resale of goods used		<input type="text"/>
Rent, rates, power and insurance costs		<input type="text"/>
Wages, salaries and other staff costs		<input type="text"/>
Car, van and travel expenses		<input type="text"/>
Repairs and renewals of property and equipment		<input type="text"/>
Phone, fax, stationery and other office costs		<input type="text"/>
Advertising costs		<input type="text"/>
Annual Investment Allowance (AIA)	Costs falling under the AIA	<input type="text"/>
Other business Expenses		<input type="text"/>

Submit Expenses

Feedback

Tax Calculation

HMRC Feedback Report

Our systems indicate that you may have reported these category totals incorrectly.

These nudges may help you make a more accurate tax submission and it should not take long to make changes where needed. Each nudge contains information to help get your submission right.

You will still be able to see these nudge messages if you click "Return to Expenses" on this page. The nudges will be presented in additional tab in the tax filing software.



Cost of goods bought for resale or goods used: You should not include all goods bought, this section should only be used for goods that are a direct cost of production for goods sold. You can find specific guidance [here](#).

Common errors people make in this category include: Claiming for tools.

Car, van and travel expenses: Your car, van and travel expenses must relate entirely to costs as a result of business tasks. Price usage of vehicles or fuel costs between home and work and not allowable. You can find specific guidance [here](#).

Common errors people make in this category include: Failing to subtract non-business fuel spending, inclusion of travel fines and inclusion of vehicle purchases.

Other Business Expenses: You should not use the Other business expenses category to place expenses you are unsure of. This should only be used for peripheral costs relating to the business like professional journal subscriptions. You can find guidance, and list of allowable and disallowable expenses [here](#).

Common errors people make in this category include: Claiming for fines and charity donations. Take care to calculate the totals correctly.

Annual Investment Allowance (AIA): You should review what an apply for the acquisition of assets under the Annual Investment Allowance. You can find specific guidance [here](#).

Common errors people make in this category include: Claiming for renovation of business premises. Take care to calculate the totals correctly.

If you are happy with the expenses you have entered, you may confirm your submission with the button below. Otherwise you may return and further alter your expenses.



Questionnaire

(1) Which items were listed on HMRC feedback report page presented after your submission? Please indicate all that apply

[categories]

(2) Did you find the HMRC feedback report page presented after the submission useful?

TODO

(3) Did you find the HMRC feedback report page presented after the submission provided sufficient detail to help you return and improve the tax return?

[text]

(4) If you did not return to your tax submission after seeing the HMRC feedback report page, could you explain why?

[text]

(5) Could you please provide any further thoughts about the HMRC feedback report page presented after your submission?

[text]

(6) How difficult did you find the task in the experiment?

Quite difficult, Quite easy, Neither difficult nor easy, Very difficult, Very easy

(7) Please explain why you found the task difficult or easy?

[text]

(8) Do you have any other comments you would like to make about any part of the experiment?

[text]

(9) How would you rate your digital skill level? Digital skill level is the extent to which you are able to use and are comfortable with using digital devices such as laptops/smartphones/tables

Very Poor, Quite Poor, Moderate, Quite Good, Excellent

(10) To what extent do you agree with this statement: I am comfortable using technology to manage my finances?

Strongly disagree, Tend to disagree, Neither agree nor disagree, Tend to agree, Strongly agree

(11) Which of the following financial software products are you aware of? Please indicate all that apply

Intuit QuickBooks, Sage Business Cloud Accounting, Xero, Free Agent

(12) On average, how often do you use the internet?

Never, Once a day, More than once a week, More than once a day

(13) What is your gender?

Male, Female

(14) What is your age?

16-24,25-34,35-44,45-54,55-64,65-74,75+

(15) How many times have you filled out a self-assessment tax return on behalf of yourself/your company?

I have never filled out an ITSA form, 1 time, 2-5 times, 6-10 times, More than 10 times

(16) Estimated annual Turnover of business you own or are employed by:

[Various]

(17) Industry sector of the firm you work for or own:

[Various]

(18) Number of people employed by the firm you work for or own:

[Various]

(19) Do you currently use software for processing expenses?

Yes, No

[Experiment Complete](#)

Congratulations! You have completed the experiment.

You incorrectly entered 1 expense item.

You will be paid £3 by the Prolific upon approval of your submission. In addition you will receive a bonus of £6.50.

Please note this page will automatically redirect for completion to the Prolific system in 30 seconds.

Thank you for participating in the FEELE study. This study was designed to understand some of the factors that effect taxpayers' decisions when filing a tax return, and how different forms of information may help them reduce the most common types of error.

This research is funded by HMRC, the government department responsible for collecting and administering taxes in the UK. We did not state this when we invited you to participate because we did not want to bias your expectations about the study before you participated. HMRC often commissions independent organisations like Exeter University to undertake studies like this one. This study is part of a broader programme of research that HMRC undertakes to better understand the needs of its customers, its operational performance and how to make improvements to its service.

We would like to reassure you that your data is fully anonymous, which means it is impossible to link your responses in the experiment with your identity. Researchers at the University of Exeter follow strict ethics standards, which include protecting the privacy of our participants.

The study has been designed in such a way that Exeter University or HMRC cannot know who has taken part.

HMRC does not have access to any of your personal data associated with this study; they will only be given the anonymous data from the experiment, which will not have any names or other identifying information. **It is not possible for HMRC to know who has taken part in the study at all.**

We will publish the results from the study in reports and scientific journal articles. The results in these reports will be presented in aggregated form – for example the average compliance rate across all participants in our study. We will never report the results of individual participants. You can see the results from other research commissioned by HMRC on their website: <http://www.hmrc.gov.uk/research/reports.htm>

If you nevertheless wish to opt out from this study, we will delete your data from the project. You can do this through the Prolific website.

Regressions

Number of items assigned to correct category

	Model 1	Model 2
ReadGuidancePreSubmission	0.939 *** (0.1766)	0.6320 *** (0.2182)
CatTotal	-0.0882 (0.2090)	-0.3333 (0.2508)
ReadGuidancePreSubmission CatTotal	0.5823 ** (0.2718)	0.8052 ** (0.3219)
ITSAExperience		0.3334 ** (0.1635)
Male		-0.2833 * (0.1628)
Age		0.0044 (0.0074)
RejectionRateInProlific		-4.7710 (3.1054)
QuizAttempts		-0.1124 (0.1159)
CompanySizeOne		0.0346 (0.1632)
InitialSubmissionTime		**
Cons	11.9670 *** (0.1386)	11.9860 *** (0.4100)
N	555	427
R ²	13.35%	15.84%

Table A2 OLS regression of factors on the number of items assigned to the correct category in the initial submission; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$ – result is robust to filter of number of items assigned to the correct category in the initial submission is greater than 9 (N=418)

Correctly Assigned Category Item Change

	Model 1	Model 2
ItemFixOther	0.0848 (0.262)	0.1951 (0.1448)
ItemFixAIA	0.34575 *** (0.1248)	0.3290 ** (0.1451)
ItemFixCost	0.1462 (0.1271)	0.2145 (0.1470)
CatTotalNo	0.0265 (0.1264)	-0.0037 (0.1439)
CatTotalFixOther	0.3307 ** (0.1289)	0.3509 ** (0.1460)
CatTotalRes	0.4733 *** (0.1281)	0.4187 *** (0.1500)
NoItemAssignCorrectInit	-0.1077 *** (0.0207)	-0.1453 *** (0.0240)
ITSAExperience		-0.0196 (0.0837)
Male		-0.0488 (0.0837)
Age		0.0003 (0.0037)
RejectionRateInProlific		-4.3648 *** (1.5927)
QuizAttempts		-0.0844 (0.0590)
CompanySizeOne		0.0153 (0.0833)
InitialSubmissionTime		0.0001 (0.0001)
Cons	1.3634 *** (0.2722)	1.9830 *** (0.3626)
N	555	427
R ²	8.25%	9.16%

Table A3 OLS regression of factors on the number of changes in the assignment of items to correct categories between the initial and final submissions; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$ – result is robust to filter of number of items assigned to the correct category in the initial submission is greater than 9 (N=418)

ItemFixOther

	Model 1	Model 2
ItemFixOther	-0.3483 ** (0.1416)	-0.2801 * (0.1450)
IncludeOtherItemInitial	-0.1092 (0.1374)	-0.1749 (0.1376)
ItemFixOther IncludeOtherItemInitial	0.7719 *** (0.1875)	0.7473 *** (0.1952)
NoItemAssignCorrectInit	-0.0379 (0.0288)	-0.0535 * (0.0283)
ITSAExperience		0.0001 (0.0991)
Male		0.1356 (0.0994)
Age		-0.0013 (0.0047)
RejectionRateInProlific		-3.5493 * (1.8923)
QuizAttempts		0.0170 (0.0799)
CompanySizeOne		-0.0516 (0.1013)
InitialSubmissionTime		-0.0000 (0.0001)
Cons	0.5583 (0.3933)	0.8875 * (0.4333)
N	163	127
R ²	16.39%	20.58%

Table A4 OLS regression of factors on the number of changes in the assignment of items to correct categories between the initial and final submissions between the treatments ItemNo and ItemFixOther; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$ – result with regard to ItemFixOther is weakened with filter of number of items assigned to the correct category in the initial submission is greater than 9 (N=123)

ItemFixCost

	Model 1	Model 2
ItemFixCost	0.2460 (0.1705)	0.4282 * (0.2206)
IncludeCostItemInitial	-0.0530 (0.1911)	0.0073 (0.2427)
ItemFixCost IncludeCostItemInitial	-0.1878 (0.2736)	-0.3750 (0.3443)
NoItemAssignCorrectInit	-0.1959 *** (0.0384)	-0.2420 *** (0.0476)
ITSAExperience		-0.1777 (0.1817)
Male		-0.0258 (0.1899)
Age		-0.0070 (0.0088)
RejectionRateInProlific		-2.0826 (3.798)
QuizAttempts		-0.1496 (0.1077)
CompanySizeOne		0.1573 (0.1828)
InitialSubmissionTime		0.0003 (0.0002)
Cons	2.4794 *** (0.4930)	2.7853 *** (0.7269)
N	163	127
R ²	16.39%	20.58%

Table A5 OLS regression of factors on the number of changes in the assignment of items to correct categories between the initial and final submissions between the treatments ItemNo and ItemFixCost; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$ – result with regard to ItemFixCost is robust with filter of number of items assigned to the correct category in the initial submission is greater than 9 (N=123)

ItemFixAIA

ItemFixAIA	-0.3973 (0.2944)	-0.6968 * (0.3896)
AssignNotCorrect	0.0243 (0.2899)	-0.0615 (0.3484)
NotAssign	-0.1285 (0.2747)	-0.2912 (0.3466)
ItemFixAIA AssignNotCorrect	0.6510 * (0.3686)	1.0728 ** (0.4735)
ItemFixAIA NotAssign	0.9294 *** (0.3225)	1.2390 *** (0.4159)
NoItemAssignCorrectInit	-0.0411 (0.0406)	-0.1102 ** (0.0517)
ITSAExperience		-0.0643 (0.1345)
Male		-0.1279 (0.1397)
Age		-0.0025 (0.0064)
RejectionRateInProlific		-1.6307 (3.0864)
QuizAttempts		-0.0868 (0.1193)
CompanySizeOne		0.0824 (0.1364)
InitialSubmissionTime		0.0003 * (0.0002)
Cons	0.6166 (0.6496)	1.6621 * (0.8751)
N	167	127
R ²	15.60%	21.72%

Table A6 OLS regression of factors on the number of changes in the assignment of items to correct categories between the initial and final submissions between the treatments ItemNo and ItemFixAIA; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$ – result with regard to ItemFixAIA is robust with filter of number of items assigned to the correct category in the initial submission is greater than 9 ($N=126$)

Total incorrect score initial submission for CatTotal treatments

CatTotalFixOther	0.1701 (0.1773)	0.2639 (0.1937)
CatTotalRes	0.0465 (0.794)	0.1689 (0.2009)
NoItemAssignCorrectInit	-1.7345 *** (0.0502)	-1.6813 (0.0511)
ITSAExperience		-0.4017 ** (0.1696)
Male		-0.2368 (0.1739)
Age		0.0015 (0.0078)
RejectionRateInProlific		4.3539 (3.1262)
QuizAttempts		0.1376 (0.1459)
CompanySizeOne		0.2087 (0.1690)
InitialSubmissionTime		-0.0000 (0.0001)
Cons	29.3998 *** (0.6596)	28.5700 *** (0.7758)
N	225	178
R ²	84.46%	88.49%

Table A7 OLS regression of factors on incorrect score in initial submission of CatTotal treatments; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

CatTotalFixOther

CatTotalFixOther	-0.5174 ** (0.2004)	-0.5053 ** (0.2368)
IncludeOtherItemInitial	-0.0863 (0.1795)	-0.1116 (0.2052)
CatTotalFixOther IncludeOtherItemInitial	1.6683 *** (0.2530)	1.7513 *** (0.2907)
IncorrectScoreInitial	-0.0119 (0.0229)	0.0017 (0.02764)
ITSAExperience		0.0997 (0.1470)
Male		0.0682 (0.1512)
Age		0.0093 (0.0067)
RejectionRateInProlific		-6.3288 * (3.4292)
QuizAttempts		-0.0519 (0.1384)
CompanySizeOne		-0.0108 (0.1401)
InitialSubmissionTime		-0.0000 (0.0000)
Cons	0.1863 (0.1790)	-0.1315 (0.4083)
N	154	124
R ²	40.55%	46.16%

Table A8 OLS regression of factors on the change in total incorrect score between the initial and final submissions between the treatments CatTotalNo and CatTotalFixOther; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$ – result with regard to CatTotalFixOther is robust with filter of number of items assigned to the correct category in the initial submission is greater than 9 (N=122)

Incorrect score change

Variable	Coefficient	
ItemCorrectCategoryCountChange	1.5743 *** (0.9718)	1.4438 *** (0.0925)
CatTotalRes	0.3103 *** (0.0985)	0.3782 *** (0.1218)
IncorrectScoreInitial	0.0023 (0.0147)	0.0123 (0.0201)
ITSAExperience		0.1731 (0.1210)
Male		-0.0516 (0.1290)
Age		0.0061 (0.0054)
RejectionRateInProlific		0.4696 (2.1847)
QuizAttempts		0.0126 (0.0960)
CompanySizeOne		0.0237 (0.1197)
InitialSubmissionTime		0.0000 (0.001)
Cons	0.0135 (0.1246)	-0.4310 (0.3484)
N	156	118
R ²	79.8%	75.43%

Table A9 – OLS regression of the change in the number of correctly assigned items between first and last submission on the total score change for the CatTotalNo and CatTotalRes treatments