Subject confusion and task non-completion: Methodological insights from an artefactual field experiment with adolescents in India☆

Anuvinda Pulickal, Sujoy Chakravarty *

Centre for Economic Studies and Planning, School of Social Sciences, Jawaharlal Nehru University, New Delhi, India

ARTICLE INFO

ABSTRACT

In the behavioral and experimental economics literature, very few experiments measure and analyze subject confusion or the completion of laboratory tasks. In this paper, we examine pre-task quiz performance in our artefactual field experiment with an adolescent subject pool in India and find that females have a lower degree of confusion in comprehending task instructions as compared to males. Furthermore, economic status is negatively associated with subject comprehension and completion of incentivized experimental tasks. Academic ability is negatively associated with subject confusion and positively associated with task completion. Finally, we find a positive association between pre-task quiz performance and the completion of incentivized tasks. This suggests that unincentivized pre-task quiz scores may act as a signal for potential non-completion of incentivized tasks in experiments, particularly with non-standard subject pools.

1. Introduction

The presence of confused experimental subjects who do not fully grasp the institution that is presented in laboratory instructions leads to incorrect or biased responses in incentivized tasks (Guala, 2005). This confusion is seen by Andreoni (1995) to be confounded with social preferences in N-player prisoner's dilemma (NPD) games. In general, if subjects lack a proper understanding of experimental instructions, it can compromise both internal (Ferraro & Vossler, 2010) and external validity (Benz & Meier, 2008; Laury & Taylor, 2008; Ferraro & Vossler, 2010) of experimental results. In economics experiments, subjects often attempt a pre-task quiz that tests their understanding of institutional rules embodied in the instructions (Friedman & Cassar, 2004). Surveying 260 experimental studies published between 2011 and 2016, Freeman et al. (2018) find that 39% of these use pre-task quizzes to reinforce instructions and to detect confusion among subjects. The

authors conduct a laboratory experiment that compares the efficacy of different methods of instruction delivery and reinforcement on reducing subject confusion, and find that 18% of subjects are confused even in their most effective method. Such persistence of subject confusion motivates an examination of the association between subject characteristics and confusion. Bigoni and Dragone (2012) is the only other study known to us that explores the effect of intrinsic subject attributes on subject confusion in an economics experiment. The attributes they explore in their laboratory experiment conducted with students in Europe are education, logical skills, linguistic ability and impulsiveness. They find that logical skills are negatively associated with confusion while impulsiveness is associated with higher confusion in the comprehension of laboratory tasks.

Confusion regarding institutional rules, attitudes regarding the topic of study, lapses in attention or memory and insufficiency of rewards in motivating actions on the part of subjects may lead to missing responses

---

or attrition in experimental data (Gomila & Clark, 2020).[1] Furthermore, administrative issues such as subjects exiting before study completion and incorrect recording of data may also lead to missing data (Gerber & Green, 2012; Gomila & Clark, 2020). Though subject attrition in longitudinal field and natural experiments has been extensively studied (Gerber & Green, 2012), it has received less attention in single-session laboratory or lab-in-the-field experiments which are common in experimental and behavioural economics. In single-session experiments, a type of attrition that is documented by researchers is missing, incomplete or inadmissible responses from subjects in incentivized tasks (Chandler & Kapelner, 2013; Cleave et al., 2013; Slonim et al., 2013; Callen et al., 2014). Inadmissible responses are those which are provided by subjects but cannot be used for analysis. For example, giving more than one response when only one is sought or providing an input which is outside the feasible range given for the experimental task. The problem is generally more pronounced in field experiments where non-standard participants may have different capacities and motivations with respect to task completion than standard subject pools of university students in developed countries (Harrison & List, 2004; Belot et al., 2015). For example, in a single-session artefactual field experiment in Afghanistan, Callen et al. (2014) find that approximately 13% of their subjects do not complete at least one task. Laboratory experiments also encounter task non-completion, albeit to a lower extent (Cleave et al., 2013; Slonim et al., 2013). Inadmissible or missing responses may be systematically correlated with subject attributes, and disregarding these observations may lead to biased sample averages of measured responses and reduce the external validity of experimental results (Rivers, 1985; Gerber & Green, 2012). Task non-completion may also be systematically associated with treatment conditions reducing internal validity of experimental results (Rivers, 1985; Gerber & Green, 2012; Gomila & Clark, 2020).

In incentivized single-session laboratory experiments, wealth has been found to negatively influence the motivation to participate (Cleave et al., 2013; Slonim et al., 2013). Brocas et al. (2014) find in their laboratory experiment that heterogeneity in the ability of subjects to receive and process experimental instructions is associated with deviations from equilibrium choices. Moreover, in the literatures on experimental and survey methodologies, individual attributes have been linked both to study participation and the quality of responses provided. Respondents who have higher motivation to answer (Lenzner, 2012; Schmidt et al., 2020), more education (Alwin & Krosnick, 1991; Schmidt et al., 2020) and higher verbal skills (Lenzner, 2012) are found to give better quality answers in surveys. Longitudinal studies that examine the role of subject attributes on missingness of data find that it is negatively associated with academic ability (Burkam & Lee, 1998; Howell and Peterson, 2006) and socioeconomic status (Fitzgerald et al., 1998; Zagorsky, 1999; Howell and Peterson, 2002). On the other hand, results for gender are mixed (Rivers, 1985; Burkam & Lee, 1998). Parental education, which is a measure of socioeconomic status (Hauser, 1994; Shrewsbury & Wardle, 2008; Aarø et al., 2009), is found to be positively associated with adolescents' perceived self-efficacy and personality (Krapohl et al., 2014), working memory (Hackman et al., 2014), achievement motivation in academics (Acharya & Joshi, 2009) and educational attainment (Hertz et al., 2008).

To our knowledge, there are no studies that examine the association between task completion and subject specific attributes such as academic ability, gender, economic status and parental education, for experiments conducted with a subject pool from a developing country. Moreover, no study has so far examined the link between pre-task quiz performance as a proxy for subject comprehension of experimental

instructions, and experimental task completion. In a related literature, Freeman et al. (2018) find that higher pre-task quiz scores are associated with lower non-money-maximizing behavior in a decision task, while Boosey et al. (2020) find that the inclusion of pre-task quiz questions increases the proportion of free riders in their repeated public goods game.

Another motivating factor for our study is that there are very few, if any, examinations of subject confusion and task non-completion in experiments with children or adolescent subject pools. Experimental studies with children and adolescents are an increasingly important area of research in economics and studies have highlighted the methodological difficulties of conducting experiments on this non-standard subject pool due to uncertainty about their degree of comprehension of rules governing laboratory tasks (Brocas & Carrillo, 2020; List et al., 2021). Task comprehension checks have been conducted in many experiments with children and adolescents (Tymula et al., 2012; Sutter et al., 2019; Bindra et al., 2020; Kiessling et al., 2021). However, as in studies with adults, there has not been a deeper examination of subject attributes that are associated with comprehension or the efficacy of comprehension checks to understand the incidence of non-completion of experimental tasks. In this paper, we examine heterogeneity with regard to gender, socioeconomic status and academic ability in adolescents' competencies to comprehend and complete experimental tasks. We also study the association between their degree of comprehension of task institutions as measured by their pre-task quiz performance and the completion of incentivized tasks.

In our artefactual field experiment conducted with 1025 adolescents who are 11 and 12th grade students in Delhi, India, we employ unincentivized pre-task quizzes to detect subject confusion in the comprehension of institutional rules pertaining to our tournament, dictator and trust games. We use a subject's quiz performance to define a count variable that measures the degree of confusion. We explore the association between this variable and academic ability, gender, economic status and parental education, controlling for religion, caste, and session-related heterogeneity. Next, we define task non-completion as the count of incorrectly attempted or unattempted incentivized experimental tasks by a subject. We examine the association of confusion, academic ability, gender, economic status, and parental education with this non-completion variable, controlling for religion, caste, and session-related heterogeneity.

We find that females, poorer adolescents, and those with higher academic ability display significantly lower confusion. Additionally, higher ability individuals display significantly higher task completion, while subjects who have higher economic status have significantly lower completion of incentivized tasks. Finally, subjects displaying higher confusion in pre-task quizzes have significantly lower completion of incentivized tasks.

2. Data and methodology

2.1. Sample and procedure

We invite 1039 adolescents studying in the 11th and 12th grades in six government schools and four private schools in Delhi to participate in our research on economic preferences and discriminatory behavior displayed by adolescents when paired with someone of similar or different economic status. Out of the 1039 invited to participate, 1025 (98.65%) participate in our experimental sessions.[2] Tables 1 and 2 describe various attributes of our subjects. Other categorical variables pertaining to our subjects and experimental conditions are specified in Table ST.4 in the supplementary material. Our subject pool is between the ages of 14 and 19, approximately gender balanced (44% female),

---

[1] To detect inattentive or confused subjects in experimental studies, psychologists often use "catch trials" where trivial choices are presented and subject behavior is recorded. These are rare in economics but have been employed by Brocas et al. (2019) and Charness et al. (2014).

[2] No data is used in this study from 14 invited individuals who choose to opt out of participation either before or during the session.

Table 1

Descriptive statistics for continuous variables.

Variable	Obs	Mean	Std. Dev.	Min.	Max.
Economic status index	1025	0.001	2.59	−4.15	15.37
Score in grade 10 exam (%)	1004	75.02	15.77	33	99
Sum of years of parents' education	954	23.73	8.44	0	38
Number incorrect or unanswered in pre-task quiz	1025	0.47	1.57	0	18
Number of incentivized tasks not completed	1025	0.08	0.44	0	6
Number incorrect in pre-task quiz	1025	0.24	0.95	0	16
Number unanswered in pre-task quiz	1025	0.22	1.16	0	14

Table 2

Descriptive statistics for categorical demographic variables.

Variable	Obs.	Min.	Max.	Categories	Frequency	Percentage
Gender	1008	0	1	Male	569	56.45
				Female	439	43.55
Caste category	1017	1	5	General Category	726	71.39
				Other Backward Castes	93	9.14
				Scheduled Castes	89	8.75
				Scheduled Tribes	57	5.60
				Don't know or Prefer not to say	52	5.11
Religion	1009	1	8	Hinduism	730	72.35
				Christianity	61	6.05
				Islam	49	4.86
				Jainism	12	1.19
				Sikhism	33	3.27
				Buddhism	20	1.98
				Other religions	9	0.89
				No religion	95	9.42

mostly Hindu (72%) and forward (general) caste (71%). The adolescents in our sample are of varying academic ability. The scores in a standardized exam that all our subjects took at the end of the 10th grade vary in the range of 33% to 99%. To represent the economic status of our subjects, we construct an economic status index employing the method of Principal Components Analysis (PCA) on the data on ownership of 16 consumer durable goods and housing characteristics pertaining to our subjects' households. This data is collected as part of the post-experiment survey. We expect our adolescent subjects to report data on ownership of consumer durable goods and housing characteristics more reliably and accurately than data on their household income or household consumption expenditure. Furthermore, data on ownership of consumer durable goods and housing characteristics will reflect a longer-term economic status of the subject's households as compared to data on their household income or consumption expenditure which are considered more volatile indicators (Desai et al., 2010). In the supplementary material, Table ST.3 shows the statistical summary of variables used for construction of the economic status index. Figure SF.1, which shows the averages for different deciles of this index in our sample, displays substantial variation in economic status. In section S.3 of the supplementary material we describe the construction of our economic status index in detail. Parental education provides another measure of our subjects' socioeconomic status in addition to the asset ownership-based index we construct. Sum of parental education of adolescents in our sample is widely dispersed in the range of 0 to 38 years with a mean of 23.73 years.

Our study comprises 41 sessions of one hour duration and they are conducted in school classrooms during regular school hours.[3] A session includes an experiment and a post-experiment survey and uses printed

response sheets on which subjects mark their responses. No identifying details of subjects are collected and they are identified by a randomly allocated study ID for the purpose of rewarding them. We employ a within-subjects design for our tasks. Accordingly, in each session, subjects are randomly allocated to four game tasks and one risk elicitation task in five separate orders.[4] Our subjects participate in a tournament game (Niederle & Vesterlund, 2007), dictator game (Kahneman et al., 1986), and in two separate roles: once as a proposer (trustor) and once as a respondent (trustee) in trust games (Berg et al., 1995; Fehr & Schmidt, 2006). Responses from trustees are collected using the strategy method.[5] The subjects are randomly assigned to one of four treatments we administer. Subjects in Treatment Higher are informed that their counterparts in the four two-player games are have a higher economic position. In Treatment Similar, the counterparts in the games are described as having a similar economic position and in Treatment Lower, the counterparts in the games are described as having a lower economic position. For subjects in Treatment Baseline, any information about the economic position of the paired players is omitted.

For the four game tasks, subjects complete the corresponding pre-

---

[4] The orders in which these games are presented to our subjects are given in section S.1 of the supplementary material.

[5] The strategy method elicits responses from the trustee for every contingent action of the proposer or trustor. After these are submitted, payoffs are calculated by matching the trustor's actual action with the contingent strategy pre-supplied by the trustee. See Johnson and Mislin (2011) for descriptions of different implementations of trust games in both game and strategy forms and a meta-analysis of behaviour using 162 replications of trust games in various forms.
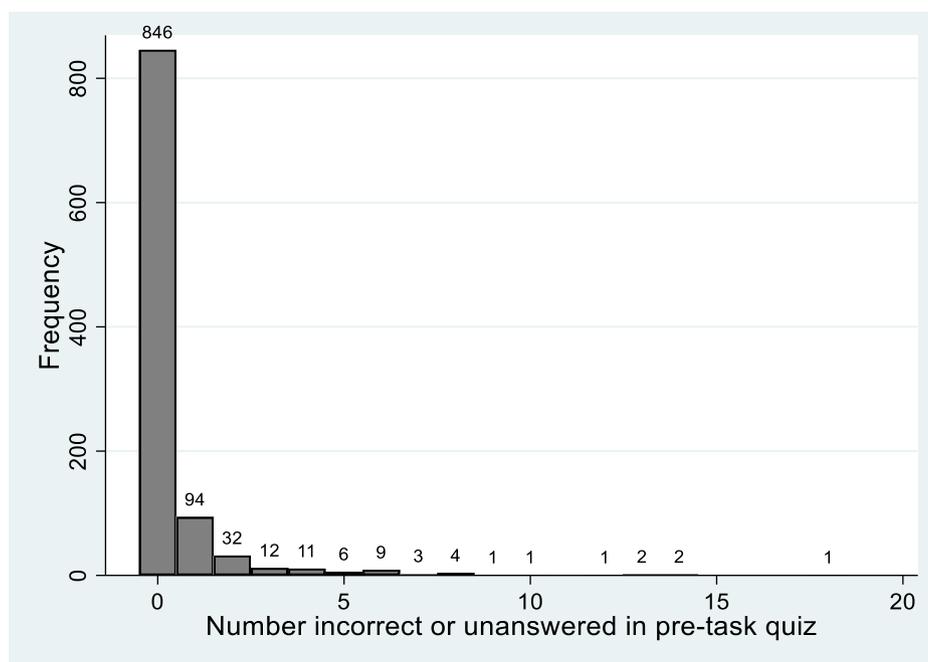
---

[3] Each of these 41 sessions is conducted by a randomly chosen pair of experimenters from a pool of 22 research assistants who are graduate students at Jawaharlal Nehru University.

Fig. 1. Distribution of Confusion.

task quizzes.[6] They are incentivized with a flat participation reward of 20 'study tokens' equivalent to 10 Indian Rupees (INR) and further rewards based on their decisions and effort.[7] On average, subjects earn 200 study tokens or INR 100 (US$ 4.70) in our hour-long study.[8] This is higher than the minimum hourly wage for skilled labor in Delhi (INR 82, Government of Delhi, 2018) and a substantial reward for a school student. Experiment instructions including pre-task quizzes are given in section S.1 of the supplementary material.

2.2. Measuring subject confusion and task non-completion

Subjects in our experiment participate in four game decision tasks after doing the corresponding pre-task quizzes. They choose between non-competitive and competitive schemes in a two-player tournament game and make dictator, trustor, and trustee decisions in separate games. Subjects make the tournament entry decision by ticking one out of two boxes; proposer and trustor decisions by ticking one out of five boxes; and four decisions as a trustee using the strategy method, for four different trustor transfers. Non-completion of an incentivized experimental task occurs if a subject chooses more than one option in the competition, dictator, and trust games, reciprocates more than the maximum possible amount in any of the four trustor transfers or does not provide a response for a task. We define task non-completion as a count variable in the interval [0, 7], indicating the number of instances in which a subject provides an incomplete or unusable response.

We employ four unincentivized pre-task quizzes, one before each

decision task, comprising 28 questions in total, to measure subjects' understanding of task rules as given in the instructions. These pertain to concepts regarding rules of the tournament choice game, dictator game and trustor and trustee roles in trust games. The experimenters read aloud and explain the instructions to the subjects. After this, subjects answer the quiz questions by choosing between two options. On completion, experimenters explain the rules again and present correct answers to the quiz. We represent a subject's degree of confusion over the four quizzes by the count of the total number of incorrect or unanswered questions in the interval [0, 28].

Out of the 1025 subjects who participate, we find that 179 (17.46%) subjects have some degree of confusion, and 49 (4.78%) do not complete one or more incentivized tasks in the correct way. We present the distribution of counts of unanswered or wrongly answered pre-task quiz questions and task non-completion in Figs. 1 and 2, respectively. Of our subjects, 80.39% have no incorrect answers in the quizzes and complete all corresponding tasks. As evident from Fig. 2, all subjects whose data is used in our analysis completed at least one task correctly.

In the supplementary material, Table ST.1 shows confusion and non-completion over different experimental tasks, and Table ST.2 shows the joint distribution of confusion and task non-completion in our subject pool. We examine the bivariate relationship between our confusion and non-completion measures visually in Fig. 3, which presents the line of best fit for the linear association between the two. The shaded area represents the 95% confidence interval for the predictor. The Pearson pairwise correlation between confusion and non-completion R, is 0.17 and this is significant at the 1% level (p-value = 0.0000). Thus, a higher degree of confusion displayed in the pre-task quiz is significantly associated with lower completion of incentivized tasks.

The two panels in Fig. 4 compare the average confusion and non-completion for the two gender categories, and provide 95% confidence intervals for these averages. We observe that confusion and task non-completion are both lower for females and the differences are significant at the 1% level for confusion (t-test p-value = 0.0059) and at the 5% level for task non-completion (t-test p-value = 0.0496).

Fig. 5 presents over six panels, the relationship of confusion and non-completion separately with our three demographic covariates of

---

[6] The risk preference elicitation task does not have a pre-task quiz and hence we do not explore it in this study. In addition, subjects also have to perform two rounds of simple arithmetic tasks pertaining to the tournament game. These tasks like the risk elicitation tasks do not have pre-task quizzes and are not explored in this study.

[7] The rewards are premium stationery items worth the amount of money a subject earns in tokens as the school administrations forbade monetary payments to students.

[8] We use the average PPP exchange rate between INR and US$ in 2019. See https://data.worldbank.org/indicator/PA.NUS.PPP (accessed 22/05/2021).
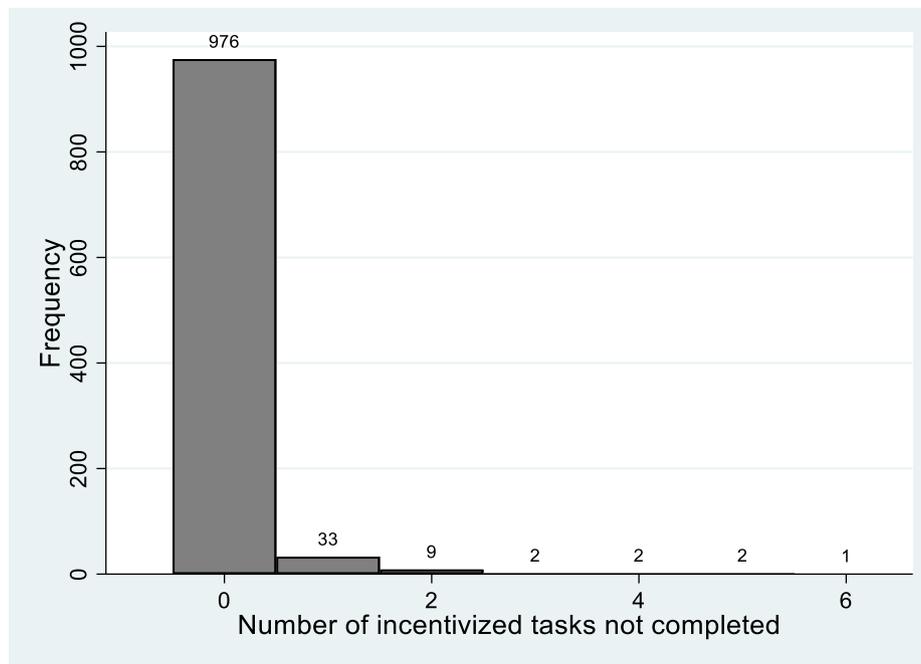
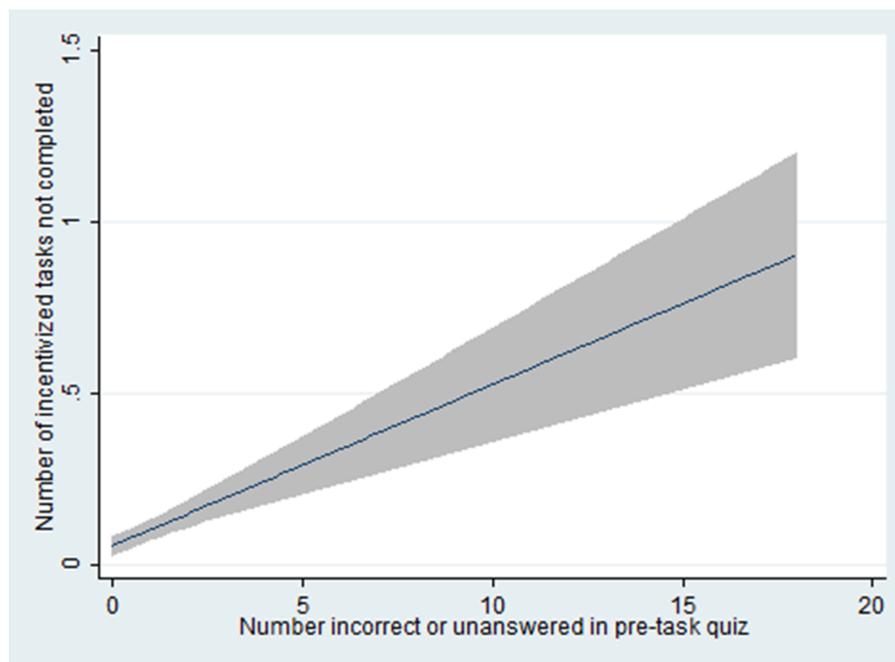**Fig. 2.** Distribution of Non-completion.

Fig. 3. Confusion against Non-completion.

interest. We find that ability proxied by 10th grade examination score is negatively associated with both confusion (pairwise $R = -0.20$; p-value $=0.0000$) and non-completion (pairwise $R = -0.09$; p-value $=0.0041$). On the other hand, sum of years of parents' education is negatively correlated with confusion (pairwise $R = -0.15$; p-value $= 0.0000$) but not significantly with non-completion (pairwise $R = -0.04$; p-value $= 0.1656$). For economic status index, the relationship with both confusion (pairwise $R = -0.02$; p-value $= 0.5848$) and non-completion (pairwise $R = 0.05$; p-value $=0.1087$) are insignificant. Thus, we observe that most of our subject-specific attributes examined separately, affect confusion and/or task completion. To estimate the effects of these variables in a more precise and robust way, we use count regressions in

the next sub-section.

2.3. Conceptual framework and econometric specifications

In this analysis, we study if confusion as proxied by the count of incorrect or incomplete responses to pre-task quizzes is systematically associated with individual specific attributes: gender, economic status, academic ability and parental education, controlling for religion, caste category, and session level heterogeneity. The other important relationship we investigate is how the count of non-completion of incentivized tasks is associated with confusion and the four individual specific attributes used above to determine confusion, controlling for caste
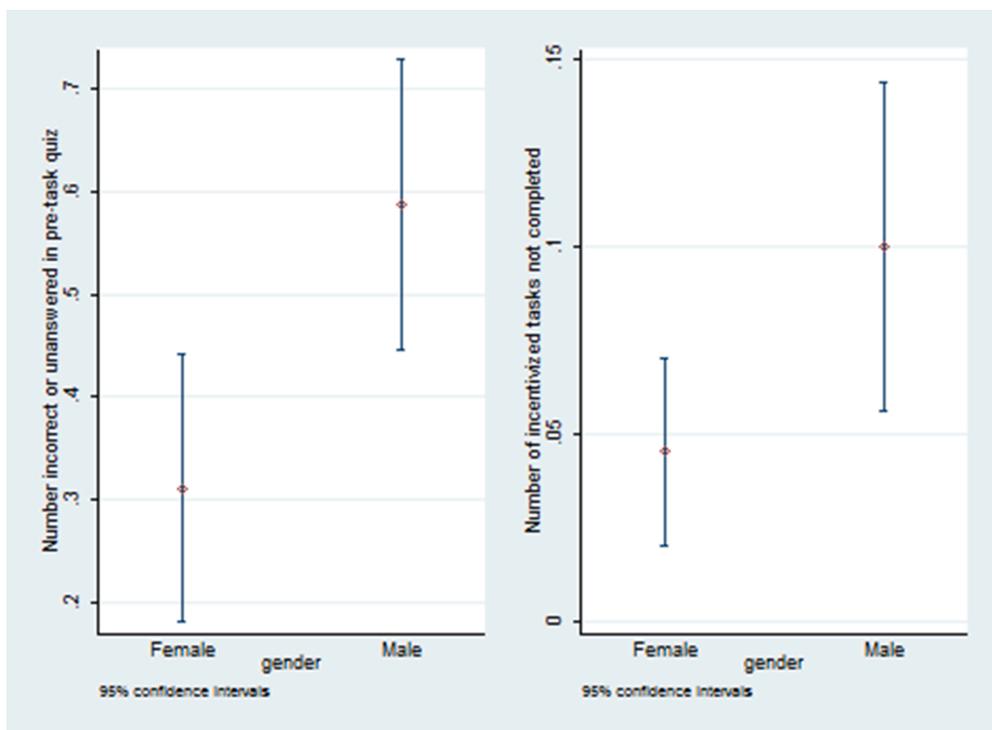
Fig. 4. Confusion and Non-completion against gender.

category, religion, and session level heterogeneity.[9]

To model determinants of subject confusion, we estimate a regression equation of the form given in (1) with the count of wrongly answered or unanswered quiz questions as the dependent variable against two groups of regressors: Individual attributes (*X*) comprising academic ability, gender, economic status index and parental education, and fixed effects (*F*) given by categories corresponding to caste, religion, and session. These are described in Table 3.

$$\ln(Subject\ Confusion) = \beta_0 + \beta'_1 X + \beta'_2 F \qquad (1)$$

To obtain determinants of task non-completion, the full model we estimate is specified by (2) with the count of unattempted or wrongly attempted incentivized tasks (task non-completion or Task_NC in short) as the dependent variable and the *X* and *F* groups of variables as regressors as in (1). In addition, we use the count of subject confusion which is the dependent variable in (1) as an independent variable in this regression.[10]

$$\ln(Task\ NC) = \beta_0 + \beta'_1 X + \beta'_2 F + \beta_3(SubjectConfusion) \qquad (2)$$

We use a Negative Binomial model for regressions (1) and (2). The

Negative Binomial is a generalization of Poisson model for count data that allows for over-dispersion, i.e. – its conditional mean may differ from the conditional variance (Wooldridge, 2012; Ryan et al., 2021).[11]

To evaluate the robustness of our results, we also run three variants of (2): a. without subject confusion, using only the *X* and *F* groups of variables as regressors, b. with subject confusion and only *F* group variables as regressors and c. with subject confusion and only *X* group variables as regressors.

3. Results

3.1. Main results

Table 4 presents coefficients (columns 1 and 3) and average marginal effects (AMEs, columns 2 and 4) from our estimations. We present estimates from specification (1) in columns 1 and 2, and variant a. of specification (2) in columns 3 and 4. The latter regresses Task NC on the *X* and *F* group of regressors but does not include the Subject Confusion variable. In Negative Binomial regressions, the coefficient can be interpreted as a semi elasticity (Greene, 2008, 2009; Wooldridge, 2012). Thus, the value of the coefficient is the difference in the logs of the

---

[9] We have 41 sessions in our experiment. Fitting 40 separate session-specific fixed effects accounts for heterogeneity in treatments given, order of presentation of treatments, school, academic stream and the pair of experimenters who ran sessions. Alternatively, in regression analysis presented in Table ST.6 in the supplementary material, we use separate school, task order, treatment, and stream level fixed effects instead of session fixed effects. These have been defined in Table ST.5 and described in Table ST.4. We find that results on our independent variables of interest are robust to either sets of controls.

[10] As equation (2) uses as regressors both the dependent (confusion) as well as covariates and categorical variables from estimations presented in columns 1 and 2 of Table 4, we check for multicollinearity among our independent variables for this specification. The variance inflation factors (VIFs) for all our independent variables remain less than 4 and the Mean VIF is 1.79. This low level of variance inflation indicates that multicollinearity issues are negligible in equation 2. See O'brien (2007) for more on multicollinearity, variance inflation and conventions associated with them.

[11] For regression (1) we obtain a non-significant χ^2 test statistic in Deviance goodness-of-fit test (p-value = 0.3902), but the Pearson goodness-of-fit test returns us a χ^2 that is significant (p-value = 0.0000) suggesting that Poisson model is inappropriate. For regression (2) both the Deviance goodness-of-fit test and the Pearson goodness of fit test return us a χ^2 that is insignificant (p-value = 1.0000 for both tests) and hence do not signal significant over-dispersion. These results suggest that the Poisson model is acceptable for regression (2). However we choose the more conservative approach and use the less restrictive Negative Binomial model following Ryan et al. (2021) who establish that the latter yield more stable and generalizable results that routinely and robustly outperform the Poisson model.
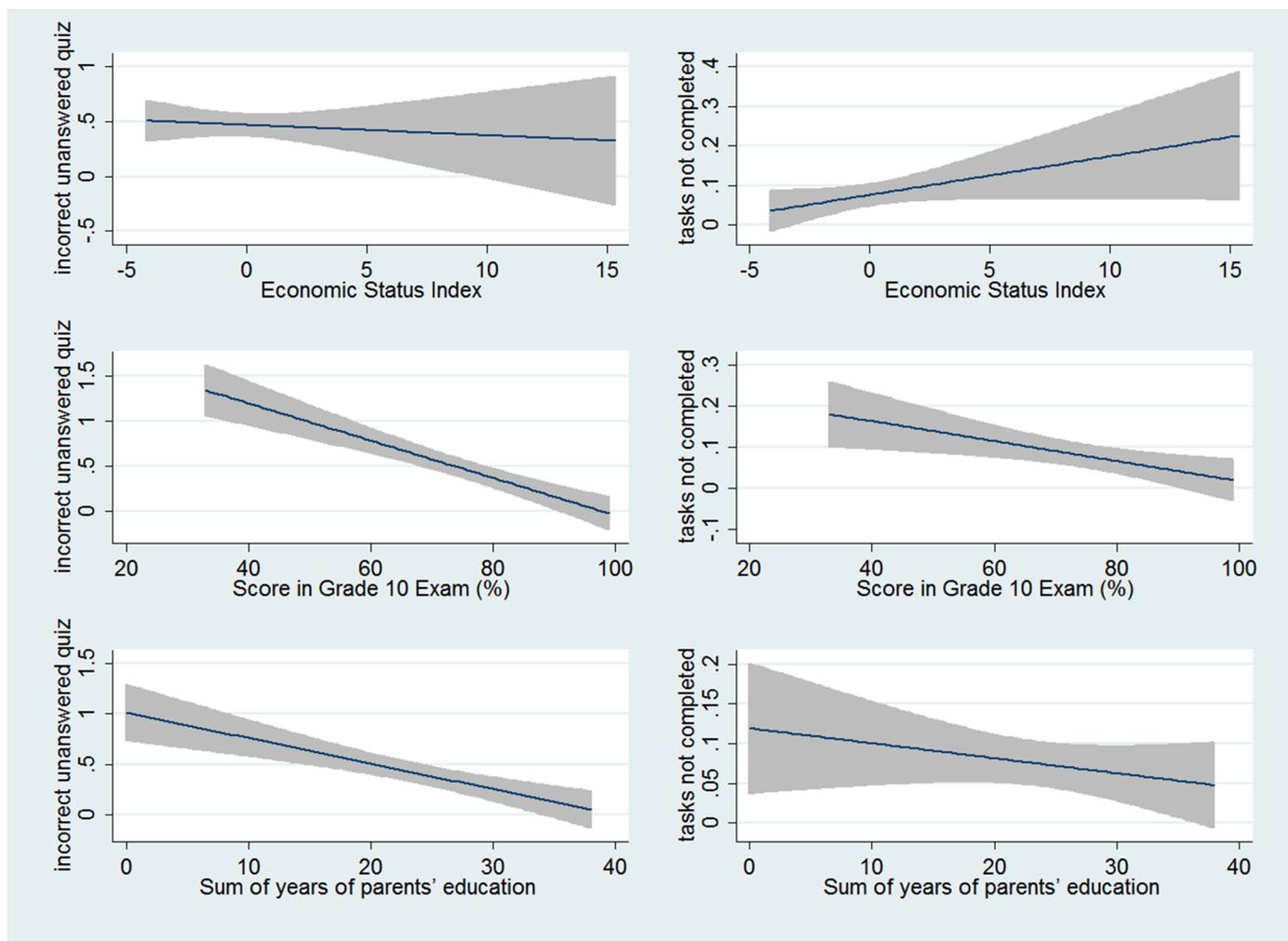
Fig. 5. Confusion and Non-completion against economic status, academic ability and parental education. Linear fit with 95% confidence interval.

Table 3

Description of variables in regression specifications.

Variable (Group)	Description
Subject Confusion$_i$ *(Dep/X)*	Count of wrongly answered or unanswered pre-task quiz questions by a subject
Task NC$_i$ *(Dep)*	Count of wrongly attempted or unattempted incentivized experimental tasks by a subject
Incorrect Quiz *(Dep)*	Count of incorrectly answered pre-task quiz questions by a subject
Unanswered Quiz *(Dep)*	Count of unanswered pre-task quiz questions by a subject
Academic Ability$_i$ *(X)*	Continuous variable that represents the percentage of marks scored by a subject in the Grade 10 School Board Examination.
Female$_i$ *(X)*	Binary variable that takes the value 1 if the subject is female, 0 otherwise.
Economic Status Index$_i$ *(X)*	Continuous variable that represents the economic status of the subject, constructed from data on ownership of consumer durable goods and housing characteristics of his or her household
Parental Education$_i$ *(X)*	Continuous variable that is obtained by adding the years of education of a subject's father and mother
Caste Category$_i$ *(F)*	Categorical variable that denotes the caste category of the subject
Religion$_i$ *(F)*	Categorical variable that denotes the religion of the subject
Session$_i$ *(F)*	Categorical variable that denotes the experimental session in which the subject participated

Table 4

Negative Binomial Regression- Subject Confusion and Task Non-completion.

Independent Variables	Dep. Var: Subject Conf Coefficient (1)	Dep. Var: Subject Conf AME (2)	Dep. Var.: Task_NC Coefficient (3)	Dep. Var.: Task_NC AME (4)
Academic Ability	−0.05***	−0.02***	−0.07***	−0.005***
	(0.009)	(0.005)	(0.02)	(0.002)
Female (ref. cat: Male)	−0.74***	−0.27***	−0.35	−0.02
	(0.22)	(0.08)	(0.43)	(0.03)
Economic Status Index	0.14***	0.05***	0.20***	0.015***
	(0.04)	(0.02)	(0.07)	(0.006)
Parental Education	−0.02	−0.007	−0.015	−0.001
	(0.02)	(0.006)	(0.03)	(0.002)
Caste Category	Yes	Yes	Yes	Yes
Religion	Yes	Yes	Yes	Yes
Session	Yes	Yes	Yes	Yes
Constant	3.46***	–	2.196	–
	(1.05)		(1.92)	
Observations	911	911	911	911
Pseudo R-Squared	0.1905	–		

Robust standard errors in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

expected counts for a unit change in a regressor.[12] The AME we present has a more convenient interpretation, i.e., change in the expected count for confusion or non-completion for a unit change in a regressor, keeping other significant regressors at their mean values. Using this, from column 2, we find that females have a 0.27 lower average count of confusion than males (p-value =0.001).[13] We also see that a unit increase in academic ability is associated with a decrease in average subject confusion by 0.02 (p-value =0.000), and a unit increase in economic status is associated with an increase in subject confusion by 0.05 units (p-value=0.005).

From column 4, we see that a unit increase in academic ability is associated with a decrease in non-completion by 0.005 units (p-value =0.001). Also, a unit increase in economic status is associated with an increase in non-completion by 0.015 units (p-value = 0.001). Unlike for subject confusion, gender is not significantly associated with task non-completion. Finally, we find that parental education is not significantly associated with subject confusion or task non-completion in our regression analysis.

Columns 1- 4 of Table 5 present coefficients and AME for variants b. and c. of specification (2). Here our Task NC variable is regressed against Subject Confusion and only variables in the *F* group (columns 1 and 2), and only variables in the *X* group (columns 3 and 4). In columns 5 and 6 we present the full specification (2).

We find that our measure for subject confusion is positively and significantly associated with task non-completion regardless of whether we include or exclude other covariates and fixed effects, with p-values = 0.000, 0.000 and 0.036 for regressions in columns 1, 3 and 5 respectively. The coefficient value for the full model in column 5 represents a $100(e^{0.15} -1)$ or 16% increase in task non-completion for a unit increase

in confusion. The corresponding AMEs are between 0.01 and 0.02 increase in the average count of Task NC for a unit increase in confusion. Of these effects, the ones in column 2 and column 4 are significant at the 5% level (p-values = 0.040 and 0.023, respectively). The AME for confusion in column 6, of 0.01 (p-value = 0.099) is marginally significant. As with regressions on Task Non-completion in Table 4, parental education and gender are never significant in Table 5. Confirming our earlier analysis, coefficients on academic ability from columns 3 and 5 are negative and statistically significant at the 1% level (p-value = 0.000 for both). The marginal effects on academic ability are significant at the 5% (p-value = 0.012) and 1% (p-value = 0.002) levels in columns 4 and 6 respectively. The coefficients on economic status are positive and significant at the 1% level in columns 3 and 5 (p-values equal to 0.000 and 0.003, respectively). The AMEs for economic status are significant at the 5% (p-value = 0.011) and 1% (p-value = 0.009) levels in columns 4 and 6 respectively.

3.2. Distinguishing by type of subject confusion

In Table 6, we separately assess the determinants of incorrectly answered pre-task quiz questions and unanswered pre-task quiz questions. From the AME in columns 2 and 4, we find that compared to males, females have a 0.08 lower average count of incorrect quiz answers (p-value =0.079) and 0.17 lower average count of unanswered quizzes (p-value =0.002). We also see that a unit increase in academic ability is associated with a decrease in average count of both incorrect quiz answers and unanswered quizzes by 0.01 (p-values = 0.000 and 0.003, respectively). Finally, a unit increase in economic status is associated with an increase in the count of incorrect pre-task quiz answers by 0.03 units (p-value=0.006), and a unit increase in parental education is associated with a decrease in the count of unanswered pre-task quizzes by 0.01 (p-value=0.058).

3.3. Robustness checks

We do further analysis to examine the robustness of our results by changing our experimental controls. In Table ST.6 of the supplementary material, we find that our main results in Table 5 are robust to controlling for task order, treatment, stream, and school level heterogeneities instead of controlling for session level heterogeneity. In columns (1) and (2) of Table ST.7, we explicitly present the order effects from the corresponding columns (1) and (2) of Table ST.6.

The last two columns of Table ST.7 redo the regressions presented on columns (1) and (3) of Table 6. As noted earlier, these deconstruct our

---

[12] As computed in Wooldridge (2012, pg. 605), we can obtain the proportionate change in the expected count for a unit change in the value of a regressor from a coefficient $(\widehat{\beta_k})$ of a Poisson or Negative Binomial regression as $[\exp{(\widehat{\beta_k})} – 1]$. Thus, in (1) of Table 4, a coefficient of −0.05 on academic ability maybe interpreted as a proportionate change in the expected count of mistakes in the pre-task quiz of −0.0488 or a decrease of 4.88 %, for a unit increase in academic score.

[13] We conduct two-tailed t-tests and find that compared to males, females in our sample of adolescents enrolled in grades 11 and 12 in Delhi schools come from more educated families (p-value = 0.0000), have higher household economic status (p-value = 0.0068) and have higher academic ability (p-value = 0.0002). Therefore, in all our regression analyses where we test for and report the gender effect on subject confusion and task non-completion, we control for parental education, household economic status, and own academic ability.

Table 5

Negative Binomial Regression - Task Non-completion.

Independent Variables	Dep. Var.: Task_NC Coefficient (1)	Dep. Var.: Task_NC AME (2)	Dep. Var.: Task_NC Coefficient (3)	Dep. Var.: Task_NC AME (4)	Dep. Var.: Task_NC Coefficient (5)	Dep. Var.: Task_NC AME (6)
Academic Ability	–	–	−0.06***	−0.006**	−0.07***	−0.005***
			(0.01)	(0.002)	(0.018)	(0.002)
Female (ref. cat: Male)	–	–	−0.45	−0.04	−0.08	−0.007
			(0.38)	(0.03)	(0.44)	(0.03)
Economic Status Index	–	–	0.20***	0.02**	0.22***	0.02***
			(0.06)	(0.008)	(0.07)	(0.007)
Parental Education	–	–	−0.005	−0.0005	−0.01	−0.001
			(0.02)	(0.002)	(0.03)	(0.002)
Subject Confusion	0.24***	0.02**	0.22***	0.02**	0.15**	0.01*
	(0.07)	(0.01)	(0.06)	(0.009)	(0.07)	(0.007)
Caste Category	Yes	Yes	–	–	Yes	Yes
Religion	Yes	Yes	–	–	Yes	Yes
Session	Yes	Yes	–	–	Yes	Yes
Constant	−4.27***	–	1.55*	–	1.26	–
	(0.94)		(0.85)		(1.88)	
Observations	1006	1006	924	924	911	911
Pseudo R-Squared	0.2391	–	0.0783	–	0.2882	–

Robust standard errors in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6

Negative Binomial Regression - Incorrect Pre-task Quiz, Unanswered Pre-task Quiz.

Independent Variables	Dep. Var: Incorrect Quiz Coefficient (1)	Dep. Var: Incorrect Quiz AME (2)	Dep. Var.: Unanswered Quiz Coefficient (3)	Dep. Var.: Unanswered Quiz AME (4)
Academic Ability	−0.05***	−0.01***	−0.05***	−0.01***
	(0.01)	(0.003)	(0.01)	(0.003)
Female (ref. cat: Male)	−0.43*	−0.08*	−1.10***	−0.17***
	(0.25)	(0.05)	(0.33)	(0.06)
Economic Status Index	0.015***	0.03***	0.09	0.02
	(0.05)	(0.01)	(0.06)	(0.01)
Parental Education	−0.007	−0.001	−0.05**	−0.01*
	(0.02)	(0.003)	(0.02)	(0.005)
Caste Category	Yes	Yes	Yes	Yes
Religion	Yes	Yes	Yes	Yes
Session	Yes	Yes	Yes	Yes
Constant	2.41**	–	3.01*	–
	(1.16)		(1.54)	
Observations	911	911	911	911
Pseudo R-Squared	0.2247	-	0.2271	-

Robust standard errors in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

original confusion variable into two manifestations of confusion, i.e. - the count of incorrect quiz answers (column 3) and the count of unanswered questions (column 4) and separately obtain their determinants. As in Table ST.6, this time instead of session fixed effects we use task order, treatment, stream, and school level fixed effects. We find that using these different but related controls results in the ESI variable becoming insignificant for the count of incorrect answers. On the other hand, for the count of unanswered quiz questions, the ESI variable which is insignificant with session fixed effects becomes significant at the 5% level. For the rest of variables that are significant at the 5% level, the signs and significance remain robust to the change in these controls. From both ST.6 and ST.7, we observe that overall subject confusion, types of subject confusion, and task non-completion differ significantly for certain orders of task presentation.

Furthermore, we find that a subject's higher inclination to maximize earnings from our experiment proxied by their dictator game giving is associated with both lower subject confusion (pairwise $R = 0.12$; p-value $=0.0002$) and lower task non-completion (pairwise $R = 0.08$; p-value $=0.0104$).

Finally, in Tables ST.8 and ST.9 we study the determinants of subject confusion and task non-completion separately in each of our four treatments using regression of the forms given in Table 4. We find that

the signs of coefficients significant at the 5% level or lower from these sub-samples are identical to those in our main results.

4. Conclusion

In our artefactual field experiment run with an adolescent subject pool in Delhi, India, we find that academic ability is negatively associated with both confusion regarding task rules and task non-completion. Our result on academic ability suggests that the level of abstraction often seen in experimental instructions, although beneficial to avoid framing effects, may make them less comprehensible for subjects with lower levels of education or ability (Brocas & Carrillo, 2020). Our result is also in line with findings from the survey and experimental methodology literature that respondent ability is positively associated with the quality of responses (Alwin & Krosnick, 1991; Bigoni & Dragone, 2012; Lenzner, 2012; Schmidt et al., 2020).

We find that compared to males, females have significantly lower number of incorrect and incomplete responses in pre-task quizzes although there is no significant gender difference in completion of incentivised experimental tasks. We speculate that the higher completion rate in pre-task quizzes of female adolescents may result from their lower impatience and risk preference compared to males which is

observed in other contexts (Sutter et al., 2013, 2019). Our result on the association between gender and subject confusion may point to literature summarized in Severiens and ten Dam (2012), which finds that adolescent girls and young college-age women averagely display higher non-cognitive skills such as motivation, discipline, and time management than their male counterparts. Our results on gender appear consistent with the gender gap in conscientiousness (Verbree et al., 2022), and work effort both in level and consistency found among college-age youth in developed countries (Woodfield et al., 2005).

We find that richer adolescents have higher subject confusion and task non-completion. Our result on economic status obtained after controlling for own academic ability and parental education is different from the findings in longitudinal studies that lower economic status is associated with higher missingness of data (Zagorsky, 1999; Howell and Peterson, 2002). However, it is consistent with Slonim et al. (2013), who find that poorer students are more motivated to opt-in for participation in laboratory experiments in economics than their wealthier counterparts. They interpret the negative association between income and willingness to participate as a consequence of lower marginal utility for experimental payments among richer individuals. In the context of our lab-in-the-field experiment, we conjecture that the motivation to invest the effort to comprehend and complete incentivized experimental tasks may similarly be higher among adolescents from lower-income households.

Our results can aid experimental design and interpretation of results from economics experiments, particularly those with younger subjects. The behavioral and experimental economics literature on children and adolescents documents associations in non-convergent directions between gender and dictator game giving (Fehr et al., 2013; Sutter et al., 2019), and economic status and dictator game giving (Chen et al., 2013; Sutter et al., 2019). Studies also find an association between subject confusion and experimental measures of prosociality such as the public goods game and the dictator game (Andreoni, 1995; Pulickal & Chakravarty, 2022). In our subject pool of adolescents, we find all these associations (Pulickal & Chakravarty, 2022). This suggests that if we do not control for subject confusion it may lead to imprecise estimation of effects arising from gender or economic status on prosociality, in non-standard subject pools particularly those with younger subjects. On the other hand, if the aim is to understand the effect of subject confusion on behavioural traits such as generosity, it may be important to control for demographic variables such as gender or economic status. Experimental studies and particularly those done with these non-standard subject pools, may therefore benefit from extensively using and reporting pre-task quiz questions and employing them suitably in data analysis as control variables.

Our finding that decrease in pre-task quiz performance significantly decreases task completion suggests that performance in pre-task quizzes may act as a signal for potential non-completion of incentivized tasks in experiments, particularly with non-standard subject pools. Missing responses in experiments leads to wastage of resources, lower statistical power, and a lower than desired sample diversity (Gerber & Green, 2012; Gomila and Clark, 2020). Experiments with children and adolescents are more prone to have confused subjects and task non-completion (Brocas & Carrillo, 2020; List et al., 2021). However, comprehension checks are reported in only less than half of experimental studies in economics, even with children and adolescents (List et al., 2021). We feel that our evidence on the significant association between performance in non-incentivised pre-task quizzes and task completion will provide further motivation to experimental economists to conduct pre-task quizzes more extensively in order to improve the quality of data and analysis, particularly in non-laboratory settings.

Data availability

Data will be made available on request.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.socec.2023.101986.

References

Aarø, L. E., Flisher, A. J., Kaaya, S., Onya, H., Namisi, F. S., & Wubs, A. (2009). Parental education as an indicator of socioeconomic status: Improving quality of data by requiring consistency across measurement occasions. *Scandinavian Journal of Public Health, 37*(2_suppl), 16–27.

Acharya, N., & Joshi, S. (2009). Influence of parents' education on achievement motivation of adolescents. *Indian Journal Social Science Researches, 6*(1), 72–79.

Alwin, D. F., & Krosnick, J. A. (1991). The reliability of survey attitude measurement: The influence of question and respondent attributes. *Sociological Methods and Research, 20*(1), 139–181.

Andreoni, J. (1995). Cooperation in public-goods experiments: Kindness or confusion? *American Economic Review, 85*(4), 891–904.

Bindra, P. C., Glätzle-Rützler, D., & Lergetporer, P. (2020). Discrimination at young age: Experimental evidence from preschool children. *Journal of Economic Behavior and Organization, 175*, 55–70.

Belot, M., Duch, R., & Miller, L. (2015). A comprehensive comparison of students and non-students in classic experimental games. *Journal of Economic Behavior and Organization, 113*(C), 26–33.

Benz, M., & Meier, S. (2008). Do people behave in experiments as in the field? Evidence from donations. *Experimental Economics, 11*(3), 268–281.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122–142.

Bigoni, M., & Dragone, D. (2012). Effective and efficient experimental instructions. *Economics Letters, 117*, 460–463.

Boosey, L., Isaac, R. M., Norton, D., & Stinn, J. (2020). Cooperation, contributor types, and control questions. *Journal of Behavioral and Experimental Economics, 85*, Article 101489.

Brocas, I., Carrillo, J. D., Wang, S. W., & Camerer, C. F. (2014). Imperfect choice or imperfect attention? Understanding strategic thinking in private information games. *Review of Economic Studies, 81*(3), 944–970.

Brocas, I., Carrillo, J.D., Combs, T.D., Kodaverdian, N. (2019) "Consistency in simple vs. complex choices by younger and older adults," Journal of Economic Behavior & Organization, 157, 580-601.

Brocas, I., & Carrillo, J. D. (2020). Studying decision-making in children: Challenges and opportunities. *Journal of Economic Behavior and Organization, 179*, 777–783.

Burkam, D. T., & Lee, V. E. (1998). Effects of monotone and nonmonotone attrition on parameter estimates in regression models with educational data: Demographic effects on achievement, aspirations, and attitudes. *Journal of Human Resources*, 555–574.

Callen, M., Isaqzadeh, M., Long, J. D., & Sprenger, C. (2014). Violence and risk preference: Experimental evidence from Afghanistan. *American Economic Review, 104*(1), 123–148.

Chandler, D., & Kapelner, A. (2013). Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior and Organization, 90*, 123–133.

Chen, Y., Zhu, L., & Chen, Z. (2013). Family income affects children's altruistic behavior in the dictator game. *PloS one, 8*(11), e80419.

Charness, G., Levin, D., Schmeidler, D. (2014) "A generalized winner's curse: An experimental investigation of complexity and adverse selection". Working Paper.

Cleave, B. L., Nikiforakis, N., & Slonim, R. (2013). Is there selection bias in laboratory experiments? The case of social and risk preferences. *Experimental Economics, 16*(3), 372–382.

Desai, S. B., Dubey, A., Joshi, B. L., Sen, M., Shariff, A., & Vanneman, R. (2010). *Human development in india.* New York: Oxford University Press.

Fehr, E. and K.M. Schmidt (2006) "The economics of fairness, reciprocity and altruism - experimental evidence and new theories," in S. Kolm and Jean Mercier Ythier (ed.), Handbook of the economics of giving, altruism and reciprocity, 1(1): 615–691, Elsevier.

Fehr, E., Glätzle-Rützler, D., & Sutter, M. (2013). The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence. *European Economic Review, 64*, 369–383.

Ferraro, P. J., & Vossler, C. A. (2010). The source and significance of confusion in public goods experiments. *The BE Journal of Economic Analysis and Policy, 10*(1).

Fitzgerald, J., Gottschalk, P., & Moffitt, R. A. (1998). An analysis of sample attrition in panel data: The michigan panel study of income dynamics. *Journal of Human Resources, 33*(2), 251–299.

Freeman, D. J., Kimbrough, E. O., Petersen, G. M., & Tong, H. T. (2018). Instructions. *Journal of the Economic Science Association, 4*(2), 165–179.

Friedman, D., & Cassar, A. (2004). *Economics lab: An intensive course in experimental economics.* Routledge.

Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis and interpretation.* New York and London: W.W. Norton.

Gomila, R., & Clark, C. S. (2020). Missing data in experiments: Challenges and solutions. *Psychological methods*. Advance Online Publication. https://doi.org/10.1037/met0000361

Government of Delhi. (2018) "The Minimum wages (Delhi)" Office of the Labour Commissioner.https://labour.delhi.gov.in/sites/default/files/All-PDF/MWCabinet.pdf (accessed 22/05/2021).

Greene, W. (2008). Functional forms for the negative binomial model for count data. *Economics Letters, 99*, 585–590.

Greene, W. (2009). Discrete Choice Modeling. In T. C. Mills, & K. Patterson (Eds.), *Palgrave handbook of econometrics*. London: Palgrave Macmillan.

Guala, F. (2005). *The methodology of experimental economics*. Cambridge University Press.

Hackman, D. A., Betancourt, L. M., Gallop, R., Romer, D., Brodsky, N. L., Hurt, H., et al. (2014). Mapping the trajectory of socioeconomic disparity in working memory: Parental and neighborhood factors. *Child Development, 85*(4), 1433–1445.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature, 42* (4), 1009–1055.

Hauser, R. M. (1994). Measuring socioeconomic status in studies of child development. *Child Development, 65*(6), 1541–1545.

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N., & Verashchagina, A. (2008). The inheritance of educational inequality: International comparisons and fifty-year trends. *The BE Journal of Economic Analysis & Policy, 7*(2).

Howell, W. G., & Peterson, P. E. (2006). *The education gap: Vouchers and urban schools*. Brookings Institution Press.

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology, 32*, 865–889.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of Business, 4*(2), S285–S300, 59.

Krapohl, E., Rimfeld, K., Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Pingault, J. B., et al. (2014). The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences, 111*(42), 15273–15278.

Kiessling, L., Chowdhury, S. K., Schildberg-Hörisch, H., & Sutter, M. (2021). *Parental paternalism and patience*. Düsseldorf: Heinrich Heine University Düsseldorf, Düsseldorf Institute for Competition Economics (DICE. DICE Discussion Paper, No. 358.

Laury, S. K., & Taylor, L. O. (2008). Altruism spillovers: Are behaviors in context-free experiments predictive of altruism toward a naturally occurring public good? *Journal of Economic Behavior and Organization, 65*(1), 9–29.

Lenzner, T. (2012). Effects of survey question comprehensibility on response quality. *Field methods, 24*(4), 409–428.

List, J. A., Petrie, R., & Samek, A. (2021). How experiments with children inform economics. *National Bureau of Economic Research (NBER)*. Working Paper w28825.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics, 122*(3), 1067–1101.

O'brien, R. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality and Quantity, 41*, 673–690.

Pulickal, A. and S. Chakravarty (2022) "The effect of own and recipient economic status on the dictator game giving of adolescents: Experimental evidence from India," SSRN working Paper," August 8, 2022. doi:10.2139/ssrn.4184279.

Rivers, D. W. (1985). *Attrition in applied social research: A methodological study, dissertations*. Loyola University Chicago, 2363 https://ecommons.luc.edu/luc_diss/2363.

Ryan, W. H., Evers, E. R. K., & Moore, D. A. (2021). Poisson regressions: A little fishy. *Collabra: Psychology, 7*(1). https://doi.org/10.1525/collabra.27242

Schmidt, K., Gummer, T., & Roßmann, J. (2020). Effects of respondent and survey characteristics on the response quality of an open-ended attitude question in Web surveys. *Methods, Data, Analyses, 14*(1), 3–34.

Severiens, S., & ten Dam, G. (2012). Leaving college: A gender comparison in male and female-dominated programs. *Research in Higher Education, 53*, 453–470.

Shrewsbury, V., & Wardle, J. (2008). Socioeconomic status and adiposity in childhood: A systematic review of cross-sectional studies 1990–2005. *Obesity, 16*(2), 275–284.

Slonim, R., Wang, C., Garbarino, E., & Merrett, D. (2013). Opting-in: Participation bias in economic experiments. *Journal of Economic Behavior and Organization, 90*, 43–70.

Sutter, M., Kocher, M. G., Glätzle-Rützler, D., & Trautmann, S. T. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *American Economic Review, 103*(1), 510–531.

Sutter, M., Zoller, C., & Glätzle-Rützler, D. (2019). Economic behavior of children and adolescents–A first survey of experimental economics results. *European Economic Review, 111*, 98–121.

Tymula, A., Belmaker, L. A. R, Roy, A. K., Ruderman, L., Manson, K., Glimcher, P. W., et al. (2012). Adolescents' risk-taking behavior is driven by tolerance to ambiguity. *Proceedings of the National Academy of Sciences, 109*(42), 17135–17140.

Verbree, A. R., Hornstra, L., Maas, L., & Wijngaards-de Meij, L. (2022). Conscientiousness as a predictor of the gender gap in academic achievement. *Research in Higher Education, 16*, 1–22. Aug.

Woodfield, R., Earl-Novell, S., & Solomon, L. (2005). Gender and mode of assessment at university: Should we assume female students are better suited to coursework and males to unseen examinations? *Assessment & Evaluation in Higher Education, 30*(1), 35–50.

Wooldridge, J. (2012). *Introductory econometrics: A modern approach* (5th ed.). South-Western College Publishing.

Zagorsky, J. L. (1999). Young Baby Boomers' wealth. *Review of Income and Wealth, 45*(2), 135–156.