

Mourelatos, Evaggelos; Giannakopoulos, Nicholas; Tzagarakis, Manolis

**Working Paper**

## Productivity outcomes in online labor markets and within-task complexity and difficulty

GLO Discussion Paper, No. 739

**Provided in Cooperation with:**  
Global Labor Organization (GLO)

*Suggested Citation:* Mourelatos, Evaggelos; Giannakopoulos, Nicholas; Tzagarakis, Manolis (2020) : Productivity outcomes in online labor markets and within-task complexity and difficulty, GLO Discussion Paper, No. 739, Global Labor Organization (GLO), Essen

This Version is available at:  
<http://hdl.handle.net/10419/227563>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# **Productivity outcomes in online labor markets and within-task complexity and difficulty**

Evangelos Mourelatos, Nicholas Giannakopoulos\* and Manolis Tzagarakis

Department of Economics, University of Patras

## **Abstract**

We analyze the impact of within-task difficulty and complexity on workers' productivity in online labor markets. Using a randomized control quasi-experiment in AMT we are able to define the difficulty and complexity embodied in requested sub-tasks within a problem-solved task. We find that our productivity measures are negatively related to the difficulty and complexity of a specific sub-task. This finding is robust to several sources of workers' heterogeneity and to different pay schemes.

Keywords: Productivity, Online Labor markets, Task Difficulty and Complexity

JEL codes: J24, D90

\*Address for correspondence: Nicholas Giannakopoulos, Department of Economics, University of Patras, University Campus, Patras, Greece. Email: [ngias@upatras.gr](mailto:ngias@upatras.gr).

## **Introduction**

Why workers in online labor markets where within task heterogeneity is allowed have different amounts of productivity outcomes? While the role of cognitive and non-cognitive skills on productivity outcomes has been examined either on online platforms (Horton 2011, 2017; Dube et al. 2020; Mourelatos et al. 2020) or lab experiments (Cubel et al. 2016), to our knowledge, there is no evidence on how the within task heterogeneity affects these outcomes. This paper fills this gap by imposing within variation regarding the difficulty and complexity of an online task conducted on Amazon Mechanical Turk. Exploiting the features of Ztree (Fischbacher, 2007) we constructed heterogeneous subtasks using an experimental complex arithmetic design based on psychological measures (Lemaire & Arnaud, 2008; Lemaire & Bruna, 2017). The participants had to respond to as many additions of 2-digit random numbers as possible, within a specific time frame. For each of these additions, we constructed three indicators that explicitly define the notion of difficulty (i.e., when the addition of two digit numbers has a sum over 100) and complexity (i.e., whether the first or the second operand of the addition is a rounding of ten). Controlling for a variety of cognitive and non-cognitive correlates, we report robust evidence that our productivity outcomes (i.e., effort, quality, productivity and efficiency) are negatively related to difficulty and complexity. We found that productivity is lower by 20% with difficulty and by 25%-30% with complexity. These differentials remain unaltered when worker's remuneration depends on either a fixed payment scheme or on piece rate arrangements.

## Data and Summary Statistics

The experimental session was divided into two rounds (round 1 corresponds to a fixed payment scheme and round 2<sup>1</sup> corresponds to a piece rate structure of payments) with an obligatory break in the middle. In each round, subjects have to respond to as many additions of 2-digit random numbers as possible, within 150 seconds of each round. Before the online task, subjects filled a questionnaire including questions on demographic, cognitive, personality traits and other socio-economic characteristics. All participants were randomly assigned to five different groups based on the piece rate scheme which activated in round 2.<sup>2</sup>

Table 1 shows summary statistics for the experimental outcomes and the utilized covariates. We observe that subjects' effort increased for complex and difficult subtasks while the remaining outcomes reduced (even slightly). For example, productivity decreases by 0.04\$ and 0.02\$ per second with difficulty and complexity, respectively. It becomes evident that our randomization process ensures that subjects' heterogeneity between complex and difficult subtasks is non-existent.

-- Table 1--

Figures 1 (2) show the average seconds per answer (average share of incorrect answers) by round and pair of the 2-digit random numbers. These patterns confirm the evidence from Table 1 and motivate the forthcoming regression analysis.

-- Figures 1 and 2--

---

<sup>1</sup> Groups are assigned to get for each correct question 0%, 1%, 5%, 10% and 15%.

<sup>2</sup> The total cost of the experiment is 1,482.69\$, i.e. 549\$ flat payment and 686,69\$ of bonus payment and 247\$ the ATM 20%. The experiment was running for 3 hours.

## Empirical Model and Results

We test whether differences in experimental outcomes are due to the differences in subtasks' complexity and difficulty. We estimate the following model:

$$Y_{ij} = \beta_0 + \beta_1 C_{1ij} + \beta_2 C_{2ij} + \beta_3 D_{ij} + \gamma^k X_i^k + \varepsilon_{ij} \quad (1)$$

where,  $Y_{ij}$  refers to a specific outcome for the  $i^{\text{th}}$  subject responding to the  $j^{\text{th}}$  sub-task,  $C_{1ij}$  and  $C_{2ij}$  are dummy indicators denoting the complexity of the 1<sup>st</sup> and 2<sup>nd</sup> 2-digit operands of the  $j^{\text{th}}$  sub-task, respectively,  $D_{ij}$  is a dummy indicator denoting the difficulty of the summation of the  $j^{\text{th}}$  sub-task,  $X_i^k$  is a vector of subject-specific attributes (presented in Table 1) and  $\varepsilon_{ij}$  is the disturbance term.

Table 2 includes the estimated coefficients for our experimental outcomes for each indicator of interest separately (columns 1-3) and for all indicators together (column 4). The latter is the preferred specification which has been also estimated for round 1 (column 5) and round 2 (column 6), separately. Since outcomes in round 2 may be affected by the piece rate scheme, we additionally include the corresponding bonus variable.

According to our results, subjects' effort increases with complexity and difficulty. Our preferred specification implies that, on average, when a subject has to add a pair of 2-digit numbers which is considered to be complex, it needs additionally 1.682 seconds (1<sup>st</sup> digit) and 1.511 seconds (2<sup>nd</sup> digit), compared to a non-complex pair. This additional time is 1.390 seconds for the case of a difficult pair.

Regarding the quality of results, we observe that complex and difficult subtasks are less likely to be answered correctly by the participants in our study. Based on a probit model specification, we show that, on average, when a subject has to add a complex pair of 2-digit numbers, has a probability of providing a correct answer which is 3.5 pp lower than the corresponding probability for a non-complex pair. The magnitude of this difference is the same for the 1<sup>st</sup> and 2<sup>nd</sup> digit of the 2-digit pair. In addition, when a subject has to add a difficult pair our results show a 2.5 pp lower probability of providing a correct answer (vs non-difficult pair).

Similarly, we show productivity losses for complex and difficult subtasks. More specifically, productivity is lower by 20% in a difficult subtask and by 25%-30% with complex ones. Lastly, our efficiency index exerts lower values when a subtask is considered to be complex and difficult. In particular, the time-adjusted correct answers are lower by 0.048 units (difficulty) and 0.060-0.074 units (complexity).

--Insert Table 2 about here--

## **Conclusions**

As in the real labor market, an online job may include several subtasks, with differences in difficulty and complexity (i.e. within the same task framework), resulting in different effort behavioral patterns by each worker. Thus, our experiment revealed a robust negative impact of task's difficulty and complexity on each and every workers' outcomes (i.e., correct answers, effort, effort-adjusted earnings and efficiency). Hence, subtask heterogeneity seems to play a significant role on the formation of the

online working behavior in crowdsourcing online labor markets. These findings have important implications for the design of tasks in online platforms and the analysis of worker-specific earnings differentials in online labor markets.

## References

Cubel, M., Nuevo-Chiquero, A., Sanchez-Pages, S., & Vidal-Fernandez, M. (2016). Do personality traits affect productivity? Evidence from the laboratory. *The Economic Journal*, 126(592), 654-681.

Dube, A., Jacobs, J., Naidu, S., & Suri, S. (2020). Monopsony in online labor markets. *American Economic Review: Insights*, 2(1), 33-46.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2), 171-178.

Horton, J. J. (2017). The effects of algorithmic labor market recommendations: Evidence from a field experiment. *Journal of Labor Economics*, 35(2), 345-385.

Horton, J.J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labour market. *Experimental Economics*, 14(3), 399-425.

Lemaire, P., & Arnaud, L. (2008). Young and older adults' strategies in complex arithmetic. *The American Journal of Psychology*, 1-16.

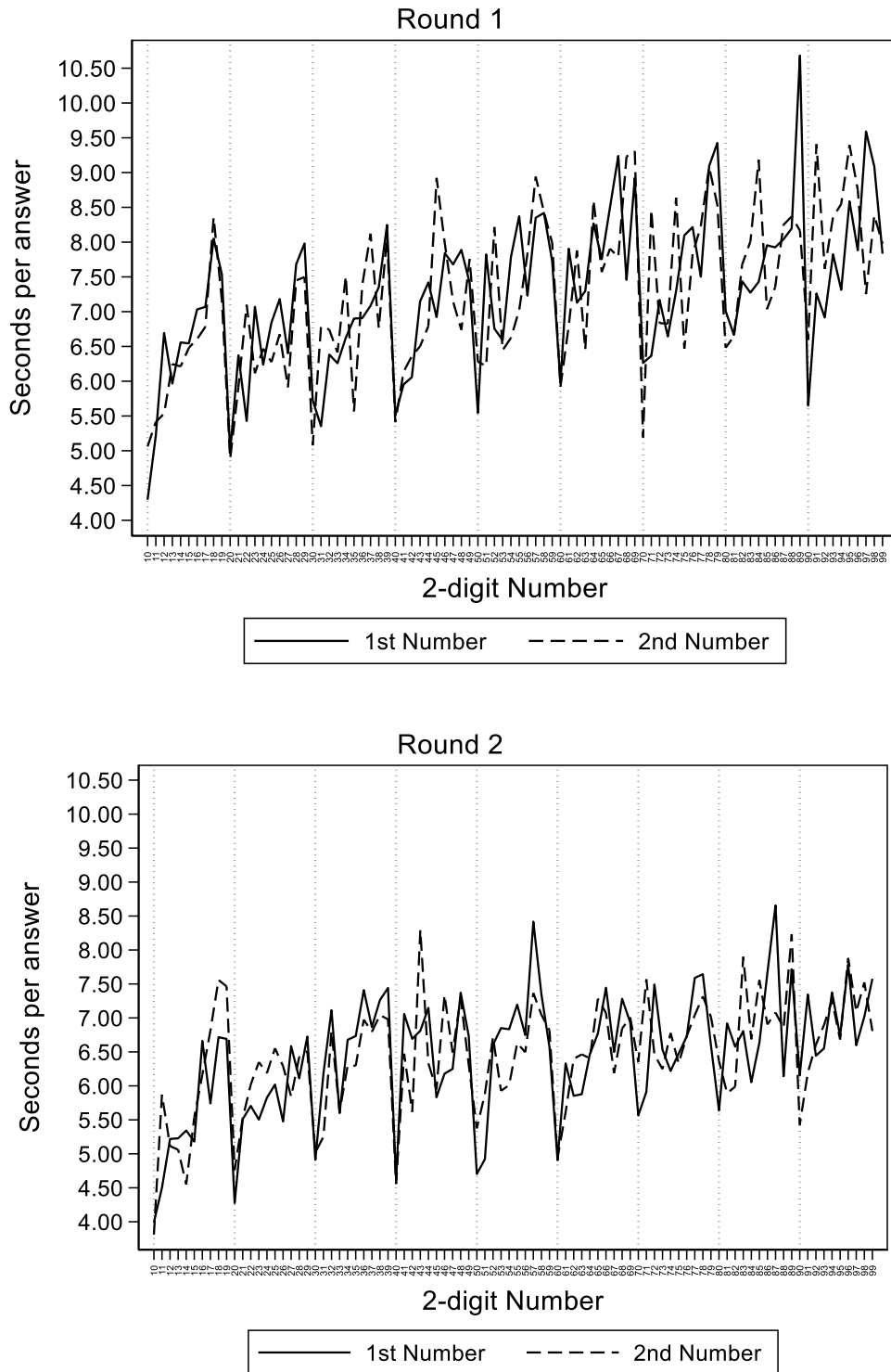
Lemaire, P., & Bruna, F. (2017). Age-Related Changes in Children's Strategies for Solving Two-Digit Addition Problems. *Journal of Numerical Cognition*, 3(3), 582-597.

Mourelatos, E., Giannakopoulos, N., & Tzagarakis, M. (2020). Personality traits and performance in online labour markets. *Behaviour & Information Technology*, 1-17.



## Figures

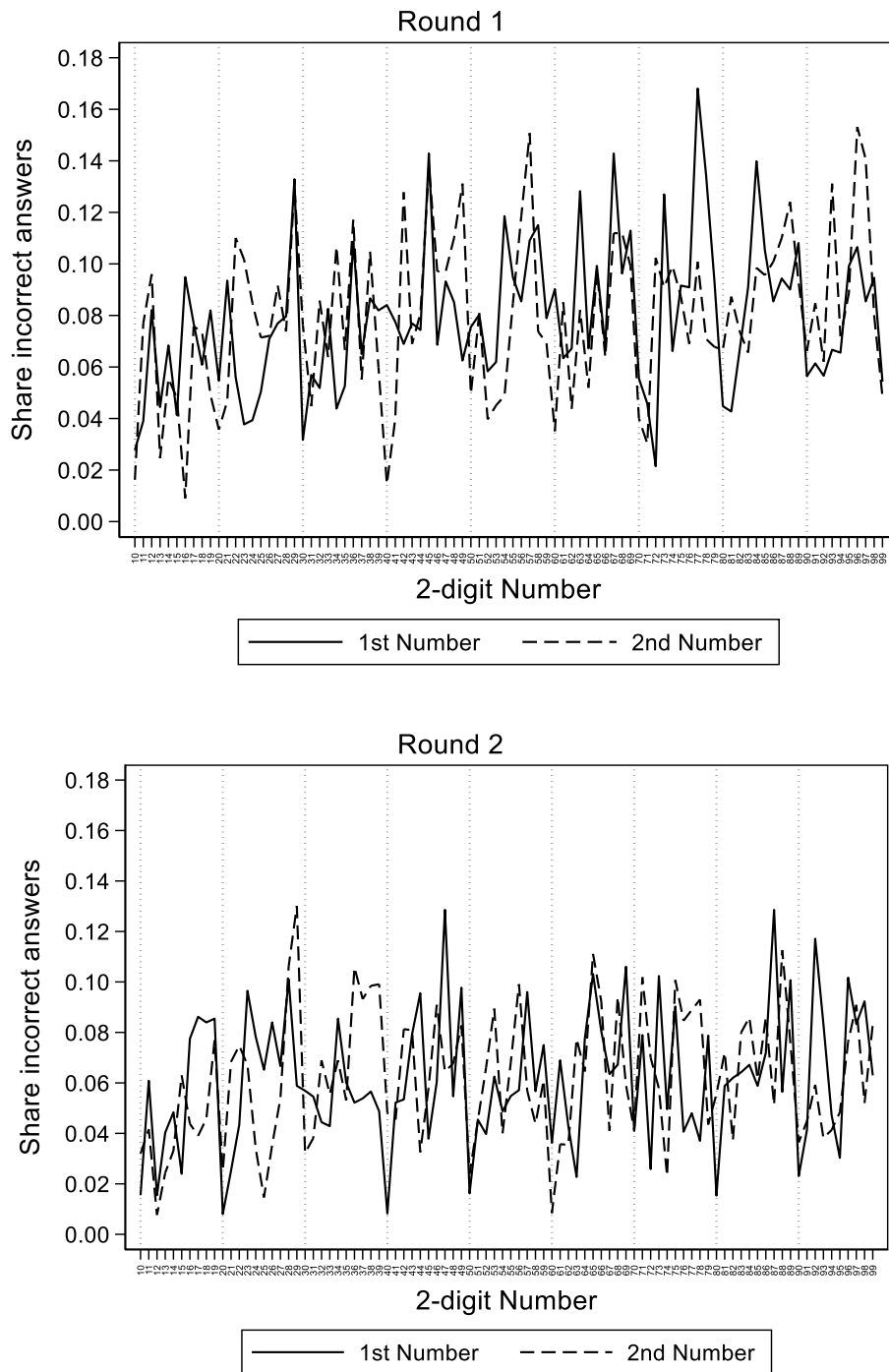
Figure 1. Average seconds per answer by round of task and pair of 2-digit random numbers



Source: Task on Amazon Mechanical Turk

Note: OLS estimates weighted by the number of responses per pair of 2-digit numbers. Vertical (dot) lines represent the 2-digit numbers: 10, 20, 30, 40, 50, 60, 70, 80, 90.

Figure 2. Average share of incorrect answers by round of task and 2-digit random numbers



Source: Task on Amazon Mechanical Turk

Note: OLS estimates weighted by the number of responses per 2-digit number. Vertical (dot) lines represent the 2-digit numbers: 10, 20, 30, 40, 50, 60, 70, 80, 90.

## Tables

Table 1. Summary statistics

Variable	All	Complex subtask 1 <sup>st</sup> digit	Complex subtask 2 <sup>nd</sup> digit	Difficult subtask
<b>Experimental outcomes</b>				
Effort (seconds per answer)	6.68	6.85	6.82	7.25
Quality (correct answers)	0.93	0.93	0.93	0.92
Productivity (earnings over effort)	0.68	0.66	0.67	0.64
Efficiency (quality over effort)	0.20	0.19	0.19	0.18
<b>Task Features</b>				
Complex task 1 <sup>st</sup> digit (0/1)	0.90	-	-	-
Complex task 2 <sup>nd</sup> digit (0/1)	0.90	-	-	-
Difficult task (0/1)	0.60	-	-	-
<b>Demographics</b>				
Age (years)	36.90	36.92	36.93	36.87
Female (0/1)	0.49	0.49	0.49	0.49
Whites (0/1)	0.84	0.84	0.84	0.84
Black (0/1)	0.06	0.06	0.06	0.06
Asian (0/1)	0.10	0.10	0.10	0.10
<b>Cognitive skills</b>				
Schooling (0/1)	0.19	0.19	0.19	0.19
IQ (0/1)	0.86	0.86	0.86	0.86
Computer competence (0/1)	0.82	0.82	0.82	0.82
<b>Personality traits</b>				
Openness	3.60	3.60	3.60	3.60
Conscientiousness	4.04	4.03	4.04	4.04
Extraversion	2.87	2.87	2.88	2.87
Agreeableness	3.83	3.83	3.83	3.82
Neuroticism	2.78	2.78	2.78	2.77
<b>Socio-economic status</b>				
Labor force participation (0/1)	0.83	0.83	0.83	0.83
Family income (0/1)	0.52	0.52	0.52	0.53
<b>Performance pay</b>				
Piece rate bonus	0.06	0.06	0.06	0.06

Source: Dataset with results drawn from Amazon Mechanical Turk. Author's calculation. The number of subjects is 504 and the total number of responses in both rounds is 21136 (10379 in round 1 and 10757 in round 2).

Notes: Schooling refers to subjects attained some college education. IQ refers to subjects without mistakes in the IQ tests. Computer competence refers to subjects with at least 75% success rate in the computer competence tests. Family income refers to subjects with reported family income equal or higher than 3,000 US dollars per month.

Table 2. The impact of task complexity and difficulty on task outcomes

	Both rounds				Fixed wage	Piece rate
	[1]	[2]	[3]	[4]	[5]	[6]
Effort (OLS)						
Complex task 1 <sup>st</sup> digit	1.755*** (0.127)	-	-	1.682*** (0.124)	1.740*** (0.213)	1.630*** (0.124)
Complex task 2 <sup>nd</sup> digit	-	1.552*** (0.114)	-	1.511*** (0.112)	1.640*** (0.172)	1.380*** (0.134)
Difficult task	-	-	1.445*** (0.085)	1.390*** (0.084)	1.721*** (0.131)	1.071*** (0.105)
R <sup>2</sup>	0.036	0.034	0.042	0.055	0.055	0.055
Quality (Probit, average marginal effects)						
Complex task 1 <sup>st</sup> digit	-0.036*** (0.005)	-	-	-0.035*** (0.005)	-0.027*** (0.008)	-0.041*** (0.006)
Complex task 2 <sup>nd</sup> digit	-	-0.035*** (0.005)	-	-0.035*** (0.005)	-0.037*** (0.007)	-0.032*** (0.007)
Difficult task	-	-	-0.026*** (0.004)	-0.025*** (0.004)	-0.026*** (0.005)	-0.025*** (0.005)
Pseudo R <sup>2</sup>	0.020	0.020	0.021	0.029	0.027	0.034
Productivity (OLS)						
Complex task 1 <sup>st</sup> digit	-0.324*** (0.018)	-	-	-0.300*** (0.018)	-0.299*** (0.023)	-0.297*** (0.022)
Complex task 2 <sup>nd</sup> digit	-	-0.252*** (0.019)	-	-0.247*** (0.018)	-0.250*** (0.022)	-0.235*** (0.023)
Difficult task	-	-	-0.213*** (0.011)	-0.204*** (0.012)	-0.239*** (0.014)	-0.195*** (0.015)
R <sup>2</sup>	0.256	0.242	0.248	0.263	0.101	0.445
Efficiency (OLS)						
Complex task 1 <sup>st</sup> digit	-0.077*** (0.003)	-	-	-0.074*** (0.003)	-0.070*** (0.004)	-0.078*** (0.005)
Complex task 2 <sup>nd</sup> digit	-	-0.062*** (0.003)	-	-0.060*** (0.003)	-0.060*** (0.004)	-0.061*** (0.005)
Difficult task	-	-	-0.050*** (0.002)	-0.048*** (0.002)	-0.049*** (0.003)	-0.048*** (0.003)
R <sup>2</sup>	0.106	0.093	0.111	0.165	0.161	0.163

Source: Authors' Calculations. Data drawn from Amazon Mechanical Turk. The number of subjects is 504 and the total number of responses in both rounds is 21136.

Notes: Productivity is expressed in logs during estimations. All specifications include controls for age, gender, race, schooling, IQ, computer competence, personality traits, labor force participation and family income. A dummy indicator for the task round is included in columns 1-4. Piece rate bonuses are included in column 6. In parentheses, heteroscedasticity corrected standard errors with clustering at subject's level.

Statistical significance: \*\*\* p<0.01, \*\* p<0.05, \* p<0.10