Appendix For Contest or Contract: Assessing the Effectiveness of Contests in Online Labor Markets

Jason Chan[1], Jiahui Mo[2], Tingting Nian[3], Nila Zhang[4]

1. Carlson School of Management, University of Minnesota, Minneapolis, Minnesota 55455, jchancf@umn.edu
2. Poole College of Management, North Carolina State University, Raleigh, North Carolina 27695, jmo3@ncsu.edu
3. The Paul Merage School of Business, University of California, Irvine, Irvine, California 92697, tnian@uci.edu
4. School of Management, Fudan University, Shanghai 200433, China, nilazhang@fudan.edu.cn

Appendix A Summary Statistics and Correlation

In this part, we explain how we measure worker ability and employer characteristics, and provide the summary statistics and correlation tables.

***Worker Ability*.** For both types of jobs, we collect all workers' ratings (including both bidders and those who submitted solutions) for completed jobs. Worker ability is dynamic and may change over time. For each worker in the focal job, we use their average historical job ratings, prior to participating in the focal job, as a measure of ability.

***Employer Countries and Income Groups*.** For each job, we retrieve the employer's country of origin. Based on the World Bank's *Countries and Economies* classification, we categorize these countries into seven regions (i.e., North America, South Asia, Europe & Central Asia, East Asia & Pacific, Sub-Saharan Africa, Middle East & North Africa, and Latin America & the Caribbean) and four income groups (i.e., high income, upper middle income, lower middle income, and low income).

The statistics for each scheme are summarized in Table A1 and A2, and the correlation table is provided in A3.

Table A1. Descriptive Statistics for Contest-based Jobs

Variable	Obs	Mean	Std. Dev.	Min	Max
(1) Quality	8,669	4.558	.698	1	5
(2) Quality_sentiment	8,669	3.838	.388	0	5
(3) Num_hire_again	8,669	1.087	1.278	0	75
(4) Payment	8,669	49.83	44.031	6	750
(5) Duration	8,669	7.585	5.558	3	75

(6) DescripitionLengh	8,669	316.155	96.137	14	568
(7) Num_skills	8,669	2.804	1.152	2	5
(8) Job Complexity	8,669	4.671	3.131	0	20
(9) Graphic Design	8,669	.975	.155	0	1
(10) Website Design	8,669	.052	.223	0	1
(11) Topic1	8,669	.267	.296	.004	.976
(12) Topic2	8,669	.217	.271	.004	.977
(13) Topic3	8,669	.25	.29	.004	.972
(14) Topic4	8,669	.123	.207	.004	.973
(15) Topic5	8,669	.043	.116	.004	.972
(16) E_tenure	8,669	4.893	2.57	0	8.552
(17) Number of Workers	8,669	24.096	19.368	1	229
(18) Ratio_highquailty_worker	8,669	.63	.16	0	1
(19) Max_worker_quailty	8,669	4.983	.234	0	5
(20) Skew_worker_quailty	8,669	-.961	1.156	-7.082	5.292
(21) Var_worker_quailty	8,669	4.258	1.419	0	6.25
(22) Historical Task Scheme Ratio (HTSR)	8,669	.828	.658	0	17
(23) Employer Historical Scheme Ratio (EHSR)	8,669	.378	1.2	0	11

Table A2. Descriptive Statistics for Contract-based Jobs

Variable	Obs	Mean	Std. Dev.	Min	Max
(1) Quality	16,199	4.785	.732	0	5
(2) Quality_sentiment	16,199	3.936	.552	0	5
(3) Num_hire_again	16,199	3.096	7.915	1	146
(4) Payment	16,199	55.968	92.757	5	2500
(5) Duration	16,199	7.852	1.414	1	28
(6) DescripitionLengh	16,199	200.328	127.944	5	634
(7) Num_skills	16,199	3.532	1.756	2	6
(8) Job Complexity	16,199	3.057	3.033	0	28
(9) Graphic Design	16,199	.93	.255	0	1
(10) Website Design	16,199	.108	.31	0	1
(11) Topic1	16,199	.183	.265	.002	.972
(12) Topic2	16,199	.143	.232	.002	.977
(13) Topic3	16,199	.165	.252	.002	.971
(14) Topic4	16,199	.199	.28	.003	.973
(15) Topic5	16,199	.091	.187	.003	.986
(16) E_tenure	16,199	4.418	2.724	0	8.678
(17) Number of Workers	16,199	32.295	29.967	1	184
(18) Ratio_highquailty_worker	16,199	.862	.164	0	1
(19) Max_worker_quailty	16,199	4.963	.111	0	6.287
(20) Skew_worker_quailty	16,199	-1.778	1.49	-7.695	1.732
(21) Var_worker_quailty	16,199	1.741	1.532	0	6.156
(22) Historical Task Scheme Ratio (HTSR)	16,199	.507	.511	0	17
(23) Employer Historical Scheme Ratio (EHSR)	16,199	.099	.409	0	11

Table A3. Correlation Table

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
(1) Schema	1.00																							
(2) Quality	0.15	1.00																						
(3) Quality_sentiment	0.09	0.45	1.00																					
(4) Num_hire_again	0.15	0.08	0.07	1.00																				
(5) Payment	0.04	0.01	-0.03	0.02	1.00																			
(6) Duration	0.04	0.03	0.01	0.08	0.05	1.00																		
(7) DescripitionLengh	-0.42	-0.09	-0.13	-0.16	0.03	-0.02	1.00																	
(8) Num_skills	0.22	0.09	0.06	0.21	0.12	0.27	-0.13	1.00																
(9) Job Complexity	-0.24	-0.05	-0.06	-0.10	-0.01	-0.05	0.60	-0.08	1.00															
(10) Graphic Design	-0.09	-0.03	-0.01	-0.11	-0.08	-0.04	0.03	-0.09	0.03	1.00														
(11) Website Design	0.09	0.02	0.02	0.06	0.12	0.09	-0.07	0.39	-0.04	0.01	1.00													
(12) Topic1	-0.14	-0.04	-0.04	-0.08	0.04	-0.05	0.09	-0.22	-0.17	0.04	-0.08	1.00												
(13) Topic2	-0.14	-0.02	-0.02	-0.06	-0.06	-0.05	0.16	-0.09	0.36	0.05	-0.06	-0.21	1.00											
(14) Topic3	-0.15	-0.05	-0.04	-0.08	0.02	-0.01	0.07	-0.12	0.03	0.04	-0.03	-0.24	-0.21	1.00										
(15) Topic4	0.14	0.05	0.04	0.02	-0.06	0.00	-0.08	0.09	0.02	-0.02	0.00	-0.27	-0.20	-0.25	1.00									
(16) Topic5	0.14	0.02	0.02	0.06	-0.00	0.02	-0.07	0.07	0.03	-0.01	0.07	-0.15	-0.12	-0.14	-0.10	1.00								
(17) E_tenure	-0.08	0.02	0.06	0.13	-0.03	0.07	-0.03	0.20	0.00	-0.03	0.07	-0.13	-0.04	-0.04	0.13	0.06	1.00							
(18) Number of Workers	0.14	-0.01	-0.03	-0.12	0.05	-0.16	0.01	-0.46	0.00	0.06	-0.16	0.16	-0.01	0.05	-0.03	-0.02	-0.10	1.00						
(19) Ratio_highquailty_worker	0.56	0.11	0.09	0.17	0.10	0.11	-0.23	0.29	-0.15	-0.01	0.12	-0.09	-0.08	-0.13	0.07	0.09	-0.01	0.02	1.00					
(20) Max_worker_quailty	-0.06	-0.02	-0.02	-0.05	-0.01	-0.09	0.04	-0.21	0.04	0.04	-0.08	0.06	0.01	0.05	-0.02	-0.02	-0.03	0.18	0.01	1.00				
(21) Skew_worker_quailty	-0.27	0.00	0.01	0.11	0.02	0.17	0.00	0.44	-0.03	-0.04	0.14	-0.09	-0.00	-0.01	-0.03	-0.01	0.15	-0.50	-0.27	-0.18	1.00			
(22) Var_worker_quailty	-0.63	-0.15	-0.10	-0.21	-0.09	-0.18	0.28	-0.46	0.19	0.09	-0.16	0.14	0.10	0.16	-0.10	-0.10	-0.02	0.15	-0.72	0.21	0.18	1.00		
(22) Historical Task Scheme Ratio (HTSR)	-0.26	-0.08	-0.06	-0.15	-0.08	-0.12	0.17	-0.21	0.14	0.16	-0.13	0.07	0.05	0.07	0.01	-0.03	-0.00	0.25	-0.28	0.12	-0.18	0.38	1.00	
(23) Employer Historical Scheme Ratio (EHSR)	-0.17	-0.03	0.01	-0.02	-0.03	0.01	0.06	0.06	0.04	-0.00	0.03	-0.02	0.03	-0.00	0.02	-0.01	0.09	-0.08	-0.09	-0.01	0.08	0.08	0.06	1.00

Appendix B: Text Mining Techniques to Analyze Job Descriptions

Employers usually provide narrative job descriptions that contain various dimensions of information and specific requirements. To better understand the content of these descriptions, we leverage multiple text mining techniques. We summarize these techniques and the operationalizations in Appendix B.

B.1 Topic Modeling and LDA

We employed a natural language processing technique, Latent Dirichlet Allocation (LDA), to identify thematic topics in the employer-specified job descriptions. As an unsupervised model, LDA does not require any prior classification of document, but explicitly extract latent topics from the unstructured text data. In general, a topic is a collection of words that frequently occur together to reflect a specific theme. Consisting of different words, a document (i.e., one job description) is a mixture of topics. Based on distributions of words, LDA identifies hidden topics and then derives the probability that documents belong to each topic. Our analyses proceed as follows. First, we preprocess all documents by removing stop words (e.g., "a", "the"). After the standard LDA preprocessing (excluding rare terms, such as, those appearing in fewer than 0.1% of job descriptions), 1,347 words were retained. Second, we apply LDA variational expectation maximization (VEM) algorithm to generate topics based on the distribution of words in all the descriptions. To identify the optimal number of topics when fitting the LDA model, we conduct CaoJuan2009 statistic model (as shown in Figure B1). The statistics suggested that the proper number of topics in our context should be between 6 and 8. We therefore fit the model with various numbers of topics, from 4 to 10 and inspected the distribution of words. When using 6 topics, the resulting clusters of words are most meaningful. When using fewer topics, each topic would include relatively coarse level of information. When using more topics, information and themes across each topic may have many overlaps and are not easy to distinguish. As such, we fitted the LDA model with six topics, and these topics are employers' information (Topic 1), detailed requirements (Topic 2), design type (Topic 3), exemplars (Topic 4), reference websites (Topic 5), and others (Topic 6). Frequently, words appearing in each topic are listed in Table B1. For each job description, we obtain the posterior topic probabilities inferred from words distribution in its corresponding descriptions. For example, the higher probability of being in one topic means that the employer discusses more about this

topic with more words in the descriptions. Therefore, the distribution among these topics reflects how an employer specified a job and we use such distribution among topics as the control variables.

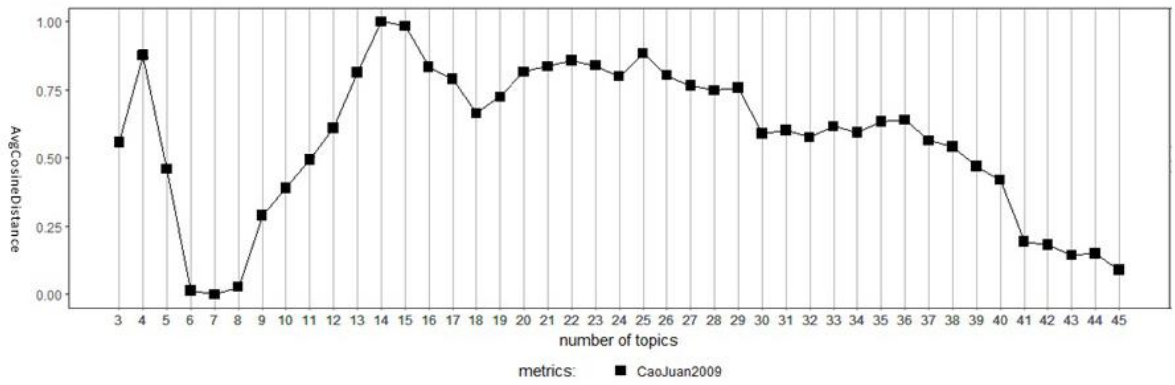


Figure B1. CaoJuan2009 Statistic Model Result

Table B1. Some Topic Keyword Examples

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
compani	color	websit	file	remov	busi
busi	black	brand	imag	login	card
product	white	app	attach	url	detail
servic	blue	icon	png	view	websit
profession	font	store	vector	websit	creat

B.2 Extracting Requirement-Related Keywords

Text-mining procedures were used to extract requirement-related keywords from job descriptions. Following standard methods (Büschken and Allenby 2016), each job description was divided into individual sentences. Using the Python package Natural Language Toolkit (NLTK), we extracted nouns and adjectives as candidate words that could represent job requirements. We then applied a stemming function to reduce words to their base forms (e.g., “logos” to “logo”), ensuring consistency and comparability across job descriptions. As a result, we obtained a comprehensive word list containing all the meaningful keywords, which can be used for future operations, such as selecting requirement-related and creativity-related keywords.

Not all words appeared frequently; some were specific to rare jobs. Including such infrequent terms could significantly increase computational complexity. Therefore, we retained only words that appeared in at least 0.1% of the job description sentences, resulting in 2,548 candidate words.

To ensure these words were truly related to job requirements, we hired three students to independently review them. The intercoder reliability, measured using Krippendorff's alpha (1980), exceeded 0.90, indicating strong agreement and validating the coding results. After manual coding, the authors reviewed all results and resolved any remaining ambiguities. In total, we identified 789 requirement-related keywords. Table B2 presents the 10 most frequent keywords.

Table B2. Most Frequently Appeared Requirement-Related Keywords

Keywords	Frequency
compani	3863
color	2435
websit	1766
brand	1094
idea	986
imag	914
product	868
file	698
black	662
white	616

B.3 Feature Engineering and Job Scheme Prediction

We implemented a comprehensive text analysis with python to identify key linguistic features that distinguish between contest-based and contract-based job schemes. This process involved several stages of data preprocessing, feature extraction, and classification.

We began with text preprocessing to standardize the input data and reduce noise. All job descriptions were converted to lowercase to ensure consistency and eliminate case sensitivity issues. Special characters and numbers were removed to retain only meaningful text, followed by tokenization to break sentences into individual word units. We eliminated common stopwords (e.g., "the," "and," "is") since these words typically carry little semantic weight in distinguishing between job types. Lemmatization was applied to normalize words to their root forms (e.g., "managing" to "manage"), which helps in consolidating word usage patterns. Additionally, we filtered out short words (less than three characters), as they tend to add little value and may introduce noise into the analysis.

For feature extraction, we adopted a dual approach to capture both lexical frequency and deep semantic meaning. The first method, TF-IDF vectorization, encodes how important a word is to a document in relation to the entire corpus. This method transformed each job description into numerical vectors based on the term frequency-inverse document frequency algorithm, including both unigrams

and bigrams to capture individual terms and frequent phrases. Complementing this, we used BERT semantic embeddings, which offer a context-aware representation of text. Unlike TF-IDF, which treats each word independently, BERT takes into account the surrounding words, enabling the model to grasp subtle nuances, synonyms, and variations in meaning. Each job description was transformed into a 768-dimensional dense vector, allowing us to capture deeper semantic structures that are otherwise lost in simpler models. By horizontally stacking TF-IDF and BERT embeddings, we constructed a comprehensive feature representation combining keyword frequency and semantic relationships.

We then trained a Random Forest classifier, a robust ensemble learning method that combines multiple decision trees. Random Forests are particularly well-suited for high-dimensional feature spaces like ours, and they reduce the risk of overfitting. We used 100 trees to strike a balance between performance and computational efficiency. Random Forests provide an intrinsic feature importance metric. We then assessed feature importance by evaluating each feature's contribution to reducing Gini impurity within the model, ranking both TF-IDF terms and BERT dimensions by their discriminative power.

Additionally, we conducted a word-level correlation analysis by calculating the Pearson correlation coefficient between each TF-IDF term and the job type label. Positive correlations indicate words more common in contest-based jobs, while negative values indicate association with contract-based jobs. The absolute values of these coefficients reflect the strength of each word's association with job type. For example, contest jobs may emphasize terms like "innovation," "challenge," or "prototype," reflecting a competitive and performance-oriented environment, while contract jobs may prioritize terms like "deadline," "client," or "reporting," suggesting structured and deliverable-focused expectations.

Finally, we produced a predictive word list ranked by correlation strength, highlighting which terms are most indicative of each job category.

B.4 Obtaining Creativity-Related Keywords

Based on the comprehensive word list generated by the first step in Section B.2, we identified the creativity related words by utilizing the "Creativity" dictionary generated by Neufeld and Gaucher (2017) in Linguistic Inquiry and Word Count (LIWC-22) tool (Boyd et al. 2022). The example words of the dictionary are listed below.

Table B3. Creativity-Related Keywords

advance	distinct	invent	prolific	unorthodoxactualiz
artistic	exceptional	inspir	radical	unprecedent
best-in-class	experimental	novel	resourceful	unusual
brainstorm	fresh	odd	revolution	unveil
breakthrough	futur	offbeat	surpris	upheav
creation	modern	open-mind	trailblaz	vicissitude
creativ	groundbreaking	origin	uncommon	vision
creativity	innovat	outofthebox	unconventional	wild
cutting-edge	imaginat	peculiar	unfamiliar	
different	initiat	pioneer	uniqu	

B.5 Examples of Requirement-Related Keywords Predictive of the Two Schemes

We display the top 20 predictive requirement-related keywords for each scheme in Table B4.

Table B4. Top 20 Predictive Requirement-Related Keywords for the Two Schemes

Top 20 Keywords for Contest Scheme	color 11.6%，modern 11.2%, feel 8.7%, prefer 8.6%, creative 8.3%, company 7.5%, idea 7.4%, blue 7.4%, black 7.3%, free 7.2%, mean 7.1%, service 7.1%, inspire 7%, contain 6.1%, word 6%, fun 6%, combine 5.8%, red 5.3%, clean 5.3%, classic 5.1%
Top 20 Keywords for Contract Scheme	url 10.2%, detail 6.5%, banner 4.4%, vector 4.3%, flyer 4.2%, pdf 4%, card 3.9%, template 3.9%, brochure 3.7%, interesting 3.4%, poster 3.3%, format 3.3%, cover 2.9%, photoshop 2.7%, jpeg 2.6%, graphic 2.4%, illustrator 1.9%, sketch 1.8%, rough 1.7%, basic 1.7%

Contest-based jobs are characterized by aesthetic and conceptual terminology such as “color” (11.6%), “modern” (11.2%), and “creative” (8.3%), suggesting that employers using contests seek innovative visual solutions with artistic freedom. These descriptions frequently include specific colors (“blue,” “black,” “red”), emotional cues (“feel,” “inspire,” “fun”), and abstract qualities that allow for interpretive flexibility.

Conversely, contract-based jobs emphasize technical specifications and deliverables through terms like “url” (10.2%), “detail” (6.5%), and file format requirements (“pdf,” “jpeg”), along with specific design outputs (“banner,” “flyer,” “brochure”). These terms indicate more structured tasks with predefined parameters, technical requirements, and industry-standard deliverables, reflecting a preference for precision and professional expertise over creative exploration.

References

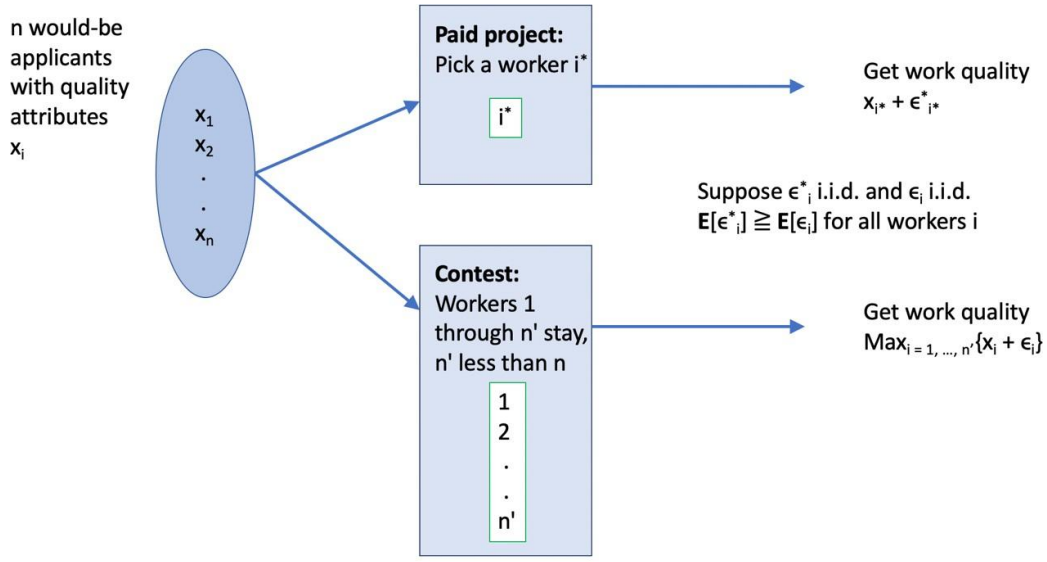
Boyd RL, Ashokkumar A, Seraj S, and Pennebaker JW (2022) *The Development and Psychometric Properties of LIWC-22* (University of Texas at Austin, Austin). 1–47.
Büschken J, Allenby GM (2016) Sentence-based text analysis for customer reviews. *Marketing Science*. 35(6): 953-975.
Krippendorff K (1980) *Validity in content analysis, chapter 1*. Moch mann E, ed. Computerstrategien F¨ur Die Kommunikationsanalyse (Campus, Frankfurt/New York). 69–112.
Neufeld KH, Gaucher D (2017) Creativity and innovation dictionary. *Dictionary for use with Linguistic Inquiry and Word Count software. Department of Psychology, University of Winnipeg*.

Appendix C Theoretical and Empirical Framings[1]

The diagram depicts a simplified setting. It considers a single design job with a predefined budget that has *n* would-be applicants. A contract-based job means selecting a single worker i^* on the basis of the observable quality x_{i^*}. After receiving private feedback and with the expectation of full pay, the worker produces the work quality $x_{i^*} + \varepsilon_{i^*}$, where ε_{i^*} is stochastic and unknowable at the time of hiring worker i^*.

If the same design job is turned into a contest, some of the n applicants are unlikely to participate, because of the required effort for uncertain pay. (Formally, the contest is an all-pay auction mechanism.) Thus, the job is left with n' participants, but they all contribute, or "bid", work, and the employer receives the maximum in work quality, $max_{i=1,\ldots,n'}\{x_i + \varepsilon_i\}$. Presumably, each participant puts in less effort, $E[\varepsilon^*] \geq E[\varepsilon_i]$, or the applicants with the best attributes x_i only accept paid projects, since otherwise contests are clearly superior.

[1] We sincerely appreciate that R3 developed the theories motivating our empirical models.



The empirical question regarding whether contests are superior given the budget is: Is $max_{i=1,\ldots,n}\{x_i\}+E[\varepsilon^*]>E[max_{i=1,\ldots,n'}\{x_i+\varepsilon_i\}]$, or vice versa?

Appendix D Matching Performance

We present performance statistics for all the matching procedures used in the main manuscript in Appendix D.

D.1 PSM in the Main Analysis and the Robustness Checks

In the main analyses in Section 4.2.1, we conducted PSM using a caliper of 0.05 and without replacement, and the corresponding balance statistics for this procedure are presented in Table D1. We also conducted a robustness check using a stricter caliper of 0.01 in Section 5.2.4, with the corresponding mean comparison presented in Table D2. Figure D1 shows the distribution of propensity scores before and after matching, based on one to one matching with calipers of 0.05 and 0.01.

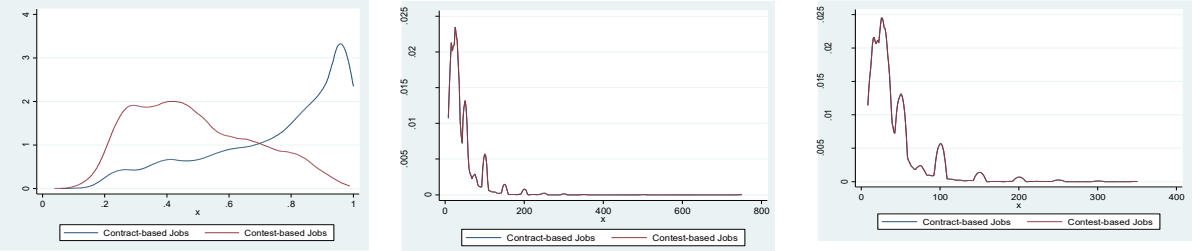
Table D1. Sample Mean Comparison Before and After Matching (Calipers of 0.05, Matching in Section 4.2.1)

Variable	Matched	Treated	Control	%bias	%reduce bias	t	p>t	V(T)/V(C)
Payment	U	49.83	55.97	-8.5		-5.82	0	0.23*
	M	42.66	42.63	0	99.5	0.04	0.964	1

Task Duration	U	7.59	7.852	-6.6		-5.77	0	15.44*
	M	7.49	7.563	-1.7	73.5	-0.9	0.369	20.06*
Description Length	U	316.16	200.3	102.4		73.87	0	0.56*
	M	294.66	280.2	12.8	87.5	6.66	0	0.78*
Number Skills	U	2.80	3.53	-49		-34.79	0	0.43*
	M	2.90	3.12	-14.6	70.3	-7.91	0	0.57*
Job Complexity	U	0.98	0.93	21.3		15.02	0	.
	M	0.97	0.97	2.1	90.4	1.28	0.199	.
Graphic Design	U	0.05	0.11	-20.5		-14.72	0	.
	M	0.06	0.07	-4.4	78.6	-2.45	0.014	.
Website Design	U	4.89	4.42	17.9		13.35	0	0.89*
	M	4.79	4.63	6	66.6	3.07	0.002	0.91*
Ln(E_tenure)	U	4.67	3.06	52.4		39.56	0	1.07*
	M	4.47	4.27	6.3	88.1	3.04	0.002	0.87*
Topic1	U	0.22	0.14	28.9		22.27	0	1.36*
	M	0.21	0.20	4.5	84.4	2.18	0.029	1.01
Topic2	U	0.25	0.17	31.4		24.13	0	1.32*
	M	0.23	0.21	9.6	69.6	4.8	0	1.04
Topic3	U	0.12	0.20	-30.6		-22.02	0	0.55*
	M	0.14	0.16	-8.3	72.8	-4.45	0	0.83*
Topic4	U	0.27	0.18	29.9		22.82	0	1.24*
	M	0.25	0.22	7.7	74.2	3.87	0	0.99
Topic5	U	0.04	0.09	-31.1		-21.97	0	0.39*
	M	0.05	0.07	-7.6	75.5	-4.41	0	0.90*

Table D2. Sample Mean Comparison Before and After Matching (Calipers of 0.01, Matching in Section 5.2.4)

Variable	Matched	Treated	Control	%bias	%reduce bias	t	p>t	V(T)/V(C)
Payment	U	49.74	56.60	-9.3		-6.41	0	0.21*
	M	40.28	40.25	0	99.6	0.04	0.97	1
Task Duration	U	7.576	7.89	-7.8		-7.03	0	15.09*
	M	7.303	7.54	-6	23.7	-2.95	0.003	19.14*
Description Length	U	315.08	191.19	107.8		78.43	0	0.56*
	M	287.63	282.03	4.9	95.5	2.38	0.017	0.81*
Number Skills	U	2.80	3.57	-51.4		-36.71	0	0.42*
	M	2.93	3.09	-10.3	79.9	-5.21	0	0.59*
Job Complexity	U	4.65	2.90	56.9		43.56	0	1.08*
	M	4.40	4.33	2.2	96.1	1	0.318	0.88*
Graphic Design	U	0.98	0.93	21		14.88	0	.
	M	0.97	0.97	1.6	92.5	0.91	0.362	.
Website Design	U	0.05	0.11	-21.6		-15.6	0	.
	M	0.06	0.07	-1.9	91.4	-0.97	0.334	.
Ln(E_tenure)	U	4.89	4.43	17.3		13.07	0	0.90*
	M	4.73	4.68	2.2	87.1	1.05	0.294	0.93*
Topic1	U	0.22	0.14	28.9		22.27	0	1.36*
	M	0.21	0.20	2.7	90.7	1.19	0.235	1
Topic2	U	0.25	0.16	31.4		24.13	0	1.32*
	M	0.22	0.21	6.8	78.4	3.15	0.002	1.01
Topic3	U	0.12	0.20	-30.6		-22.02	0	0.55*
	M	0.15	0.16	-5.6	81.7	-2.74	0.006	0.87*
Topic4	U	0.27	0.18	29.9		22.82	0	1.24*
	M	0.25	0.22	6.6	77.8	3.08	0.002	0.98
Topic5	U	0.04	0.09	-31.1		-21.97	0	0.39*
	M	0.05	0.07	-7.3	76.6	-3.86	0	0.91*



Before Matching | Caliper=0.05 No Replacement | Caliper=0.01 No Replacement

Figure D1. P-Score Distribution of Contract-based and Contest-based Jobs Before and After Matching (Matching in Section 4.2.1 and Section 5.2.4)

D.2 Two-Step Matching with Keywords in Section 4.2.2

In the two-step matching procedure described in Section 4.2.2, we first conducted an additional propensity score matching (PSM) analysis using the same set of covariates, specifying one-to-ten nearest-neighbor matching with a caliper of 0.05 and allowing for replacement. The corresponding balance statistics are reported in Table D3. Following the PSM procedure, we calculated the paired keyword-based description distance and, among the ten matched contract jobs, retained the job with the highest description similarity. Figure D2 presents the distribution of propensity scores before matching, after the first-step PSM matching, and after the second-step keyword matching.

We also report new balance statistics for the textual covariates, showing that job descriptions in terms of requirements are likewise better balanced after the two-step matching. To evaluate the textual description, we compare keywords (e.g., the total number of keywords, as well as for keywords related to requirements, creativity, and topics) across the two groups, with the results reported in Table D4. We also compute keyword-based similarity between the treatment and control groups. Before matching, the similarity score between the two groups is 0.0577, whereas after matching it increases to 0.2343. This increase indicates that, following matching, job descriptions in the two groups are substantially more similar in their textual content.

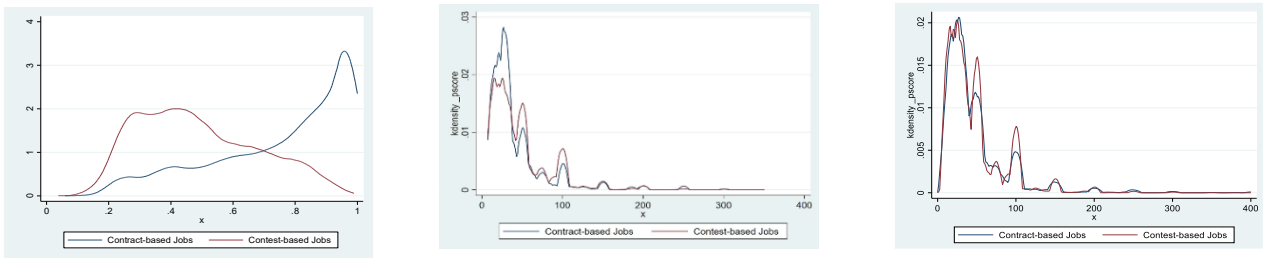
Table D3. Sample Mean Comparison Before and After the PSM Matching (Section 4.2.2)

Variable	Matched	Treated	Control	%bias	%reduce bias	t	p>t	V(T)/V(C)
Payment	U	49.74	56.61	-9.3		-6.41	0	0.21*
	M	45.62	45.58	0.1	99.4	0.07	0.944	1
Task Duration	U	7.58	7.89	-7.8		-7.03	0	15.09*
	M	7.53	7.51	0.5	93.1	0.34	0.731	21.44*

Description Length	U	315.08	191.19	107.8		78.43	0	0.56*
	M	312.90	302.92	8.7	91.9	5.99	0	0.74*
Number Skills	U	2.80	3.57	-51.4		-36.71	0	0.42*
	M	2.83	3.03	-13.3	74.1	-9.2	0	0.57*
Job Complexity	U	4.65	2.90	56.9		43.56	0	1.08*
	M	4.66	4.50	5.1	91.1	3.02	0.002	0.85*
Graphic Design	U	0.98	0.93	21		14.88	0	.
	M	0.98	0.97	2.6	87.7	2.09	0.037	.
Website Design	U	0.05	0.11	-21.6		-15.6	0	.
	M	0.05	0.06	-3.4	84.2	-2.51	0.012	.
Ln(E_tenure)	U	4.89	4.43	17.3		13.07	0	0.90*
	M	4.89	4.79	4	76.8	2.56	0.01	0.91*
Topic1	U	0.22	0.14	28.9		22.27	0	1.36*
	M	0.22	0.20	5.2	82.2	3.01	0.003	0.99
Topic2	U	0.25	0.16	31.4		24.13	0	1.32*
	M	0.25	0.22	8.9	71.6	5.34	0	1.02
Topic3	U	0.12	0.20	-30.6		-22.02	0	0.55*
	M	0.13	0.14	-6.2	79.6	-4.41	0	0.83*
Topic4	U	0.27	0.18	29.9		22.82	0	1.24*
	M	0.26	0.25	5.9	80.4	3.52	0	0.97
Topic5	U	0.04	0.09	-31.1		-21.97	0	0.39*
	M	0.04	0.06	-8.1	74.1	-6.36	0	0.81*

Table D4. Sample Mean Comparison for keywords Before and After Matching with Keywords

	Matched	Treated	Control	%bias	%reduce bias
number of all keywords	U	10.4728	5.6238	104.9	
	M	10.8447	8.9655	34.6	67.1
number of requirement keywords	U	5.2388	3.0109	73.4	
	M	5.4648	4.7557	17.6	76
number of creative keywords	U	0.1634	0.0628	27.7	
	M	0.1615	0.1157	13.4	51.5
number of topic keywords	U	9.5406	4.9773	105.9	
	M	9.8893	8.1169	35.8	66.2



Before Matching After the 1st step, Caliper=0.05,1:10 With Replacement After the 2nd step keyword matching

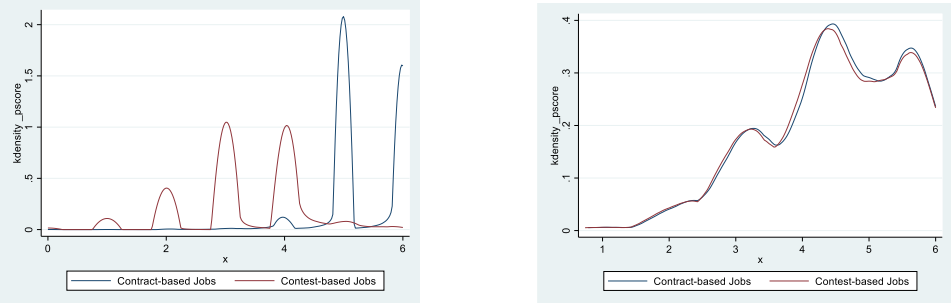
Figure D2. P-Score Distribution of Contract-based and Contest-based Jobs After Two-Step Matching (Matching in Section 4.2.2)

D.3 Matching in Section 5.2.2

We conducted another matching procedure in Section 5.2.2, considering workers' expected winnings and the ability distribution among workers in a task. The performance statistics and the distribution of propensity scores are presented in Table D4 and Figure D3.

Table D4. Sample Mean Comparison Before and After Matching (Matching in Section 5.2.2)

Variable	Matched	Treated	Control	%bias	%reduce bias	t	p>t	V(T)/V(C)
Ln(Expected Prize)	U	3.60	0.99	171.9		116.86	0	0.16*
	M	3.21	3.02	12.9	92.5	2.37	0.018	0.19*
Task Duration	U	7.59	7.85	-6.6		-5.77	0	15.44*
	M	8.09	8.11	-0.6	90.7	-0.1	0.923	14.60*
Description Length	U	316.16	200.33	102.4		73.87	0	0.56*
	M	275.67	259.74	14.1	86.3	2.29	0.022	0.73*
Number Skills	U	2.80	3.5319	-49		-34.79	0	0.43*
	M	3.58	3.83	-17.1	65.1	-2.88	0.004	0.56*
Job complexity	U	0.98	0.93	21.3		15.02	0	.
	M	0.95	0.93	10.3	51.8	1.64	0.102	.
Graphic Design	U	0.05	0.11	-20.5		-14.72	0	.
	M	0.14	0.14	-2.3	88.8	-0.32	0.748	.
Website Design	U	4.89	4.42	17.9		13.35	0	0.89*
	M	4.73	4.70	1.3	92.9	0.23	0.817	1
Ln(E_tenure)	U	4.67	3.06	52.4		39.56	0	1.07*
	M	4.25	4.02	7.3	86	1.19	0.236	0.85*
Topic1	U	0.22	0.14	28.9		22.27	0	1.36*
	M	0.19	0.15	16.7	42.4	2.97	0.003	1.29*
Topic2	U	0.25	0.16	31.4		24.13	0	1.32*
	M	0.21	0.21	-2	93.6	-0.36	0.721	0.92
Topic3	U	0.12	0.20	-30.6		-22.02	0	0.55*
	M	0.17	0.16	2.1	93.2	0.37	0.711	1.06
Topic4	U	0.27	0.18	29.9		22.82	0	1.24*
	M	0.20	0.21	-1.4	95.3	-0.26	0.793	0.99
Topic5	U	0.04	0.09	-31.1		-21.97	0	0.39*
	M	0.07	0.07	-0.7	97.9	-0.11	0.911	1.12
Number of Workers	U	24.10	32.30	-32.5		-23.03	0	0.42*
	M	14.52	13.55	3.8	88.2	1.15	0.251	0.67*
Ratio of High-quality Workers	U	0.630	0.86	-143		-107.07	0	0.95*
	M	0.76	0.70	38.4	73.1	4.4	0	0.39*
Max Worker Quality	U	4.98	4.96	10.5		8.81	0	4.40*
	M	4.98	4.92	32.6	-209	3.61	0	0.30*
Skewness of Worker Quality	U	-0.96	-1.78	61.3		44.43	0	0.60*
	M	-1.62	-0.93	-51.7	15.7	-8.83	0	2.06*
Variance of Worker Quality	U	4.26	1.74	170.4		126.61	0	0.86*
	M	2.56	2.56	0	100	0	0.996	0.82*



Before Matching Calliper=0.05 No Replacement

Figure D3. P-Score Distribution of Contract-based and Contest-based Jobs Before and After Matching (in Section 5.2.2)

Appendix E IV Estimation

While propensity score matching may help make the sample more comparable, it primarily accounts for observed factors and may not be able to control for unobservable factors. To address potential endogeneity issues arising from unobserved factors, we supplement the main analyses in Section 4.2 with an instrumental variable (IV) approach.

We propose the Historical Task Scheme Ratio (*HTSR*) and Employer Historical Scheme Ratio (*EHSR*) as potentially suitable instruments for our study. *HTSR* refers to the propensity of similar tasks to be posted as a contest or a contract-based scheme. To calculate *HTSR* for the focal task, we first identify similar tasks based on the required skills. Each employer specifies the skills needed to complete the task by using platform-provided skill tags. Two jobs are considered similar if they have identical skill requirement tags (*i.e.*, the exact same set of tags). Among these similar tasks, we determine how many tasks were posted as contests and how many were completed under the contract-scheme. *HTSR* is then calculated as the ratio of contests within the similar tasks to contracts within the similar tasks.

EHSR builds on the concept that employers tend to have a preferred job scheme. Here, we define *EHSR* as the ratio of an employer's past tasks that were listed as a contest to those that were listed under a contract scheme. To construct this measure, we obtain all tasks posted by each employer prior to the focal task and identify the number of contest-based jobs and contract-based jobs. *EHSR* is calculated as the ratio of the two figures.

For *HTSR* and *EHSR* to be valid instruments, they must satisfy two conditions: 1) they should be correlated with the potentially endogenous independent variable (i.e., job scheme), and 2) they should

not be directly related to the outcome (i.e., work quality). The fulfilment of these two conditions ensures that the effect of the endogeneous variable (job scheme) on the outcomes is influenced solely through the instrumental variables, which generates random perturbations to the endogeneous variables.

We believe that both conditions are fulfilled. First, *HTSR* and *EHSR* are likely correlated with the choice of job scheme that employers use. *HTSR* could influence scheme choice for a focal task due to the tendency of employers to follow established patterns or norms based on similar tasks. Employers may look to the choices made for similar tasks as a benchmark. The rationale is that if a significant proportion of similar tasks have opted for a contest scheme, it suggests that contests may be more effective or suitable for that type of task. The choice made by the majority of employers serve as a signal of optimized decision, guiding employers to choose the scheme that has been commonly and successfully used by others for similar tasks. Thus, *HTSR* serves as a heuristic or informational shortcut that influences employers' decision-making processes by reflecting the collective wisdom and experiences of their peers in similar situations.

The relevance of *EHSR* as an instrument follows a similar rationale. *EHSR* reflects an employer's historical experience with the two schemes. By aligning with common historical practices, it reduces the perceived risk and uncertainty associated with choosing a scheme. Therefore, *EHSR* serves as a context-specific heuristic that guides employers toward the scheme previously used, thereby influencing their choice for the focal task.

In addition to the relevance condition, both *HTSR* and *EHSR* meet the exclusion restriction condition, meaning that they do not have a direct effect on the quality of the focal task. Both instrumental variables provide a heuristic based on the behaviors of other jobs and historical jobs, but do not inherently carry information about the specific characteristics of the focal task itself. As a result, they should not directly affect the quality of the current task. In summary, both *HTSR* and *EHSR* influence the choice of job scheme based on patterns observed in other and historical jobs, yet remain uncorrelated with the actual quality of the focal task. This ensures that their effect on the task quality is exclusively through their impact on the scheme choice, satisfying the exclusion restriction condition.

We perform a set of tests to assess both the relevance and exclusion restriction to ensure the appropriateness of instrument variable in our subsequent sections. A range of statistics tests further

supports that the instrument variables are valid and strong. The Kleibergen-Paap rk Wald F statistics are 405.46 and 158.09, and the Cragg-Donald Wald F statistics are 387.23 and 250.63, both exceeding the thresholds for weak instrument tests in 2SLS (Stock and Yogo 2005). Additionally, the Hansen J test (Hansen 1982) for overidentifying restrictions shows $\chi 2 = 0.41$ ($p = 0.52$) and $\chi 2 = 0.17$ ($p = 0.68$), indicating that the exclusion restriction is likely satisfied. The Anderson canonical correlation LR statistic ($\chi 2 = 724.02$ for Model (1) and $\chi 2 = 480.22$ for Model (2)) indicates that our estimation did not suffer from underidentification, supporting the relevance of our IVs. Furthermore, the instrument variables pass the exclusion restriction test (KS statistic = 0.054, p = 0.964) based on D'Haultfoeuille, Hoderlein, & Sasaki (2021).

We use a two-stage least squares regression (2SLS) to estimate the matched sample derived from Section 4.2.1 (Angrist and Krueger 1994) and the findings are presented in Column (1) and (2) in Table E2. Although these two instrumental variables (IVs) do not carry direct information about the focal jobs, some unobserved factors related to seekers may still indirectly influence job quality through the IVs. For example, employers who prefer a contract scheme are more inclined to work closely with some workers, and this close collaboration may improve the perceived final job quality. To ensure our IV estimations did not suffer these potential unobserved issues, we conducted IV estimations based on three different subsamples.

(1) a sample of jobs in which the winner and the employers do not have prior experience working with each other (used in the analysis in Column (3) and (4) of Table 4). This could help address the concern that previous relationships or familiarity might artificially inflate work outcome.

(2) a sample of seekers who have experience with both compensation schemes (used in the analysis in Column (7) and (8) of Table 2). Seekers being familiar with the norms and expectations of both schemes could reduce the risk that outcomes are biased by scheme-specific familiarity or mechanisms.

(3) a newly matched sample based on job descriptions containing specific keywords (used in the analysis in Column (5) and (6) of Table 2). These subsamples help us account for potential unobserved factors across different employers. When employers have used a specific scheme multiple times and have worked with the same workers repeatedly, they may specify job descriptions using particular keywords (i.e., some words with which they are already familiar). Such specifications could influence

job performance. Therefore, in the third subsample, we ensure that job descriptions are closely matched in terms of keyword similarity to rule out potential differences in job descriptions.

The first-stage results are presented in Table E1 in Appendix E and the main findings are in Table E2. Across all estimations in E2, we have positive coefficients. The significantly positive coefficients estimated by 2SLS, suggest that, on average, the quality scores of contract-based jobs are higher than those of contest-based jobs across both model specifications. These results further substantiate the robustness of our main model's findings in Section 4.2.

Table E1 Results for First Stage of 2SLS in Different Samples

	PSM		No prior work Experience		Seekers with both scheme		PSM-keywords	
First Stage	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
HTSR	-.1685***	-.1112***	-.1192***	-.0799***	-.1575***	-.1239***	-.1617***	-.1179***
	(.0082)	(.0066)	(.0092)	(.0073)	(.0152)	(.0134)	(.0128)	(.0103)
EHSR	-.0809***	-.0503***	-.0826***	-.05***	-.0686***	-.0522***	-.0655***	-.0397***
	(.0047)	(.0037)	(.0056)	(.0043)	(.0055)	(.0047)	(.0063)	(.005)
Job Attributes								
Payment	-.0003**	-.0022***	-.0004***	-.0026***	-.0003	-.0023***	-.0005***	-.0029***
	(.0001)	(.0001)	(.0001)	(.0001)	(.0003)	(.0003)	(.0002)	(.0001)
Duration	-.0015	-.0015	-.0023*	-.002**	.0077***	.0081***	-.0029**	-.0024**
	(.0011)	(.0009)	(.0012)	(.0009)	(.0024)	(.0021)	(.0014)	(.0011)
DescriptionLength	-.0002***	-.0002***	-.0001***	-.0001***	-.0001	.0000	-.0004***	-.0003***
	(0)	(0)	(.0001)	(0)	(.0001)	(.0001)	(.0001)	(.0001)
Num_skills	.0326***	.046***	.0031	.0302***	.0848***	.0859***	.0321***	.0453***
	(.0037)	(.0031)	(.0047)	(.0038)	(.0075)	(.0069)	(.0056)	(.0046)
Job Complexity	.0006	.0009	.001	.0007	-.0005	-.0013	-.0014	.0002
	(.0018)	(.0014)	(.002)	(.0015)	(.0039)	(.0033)	(.0023)	(.0018)
Graphic Design	.0606**	-.0326	.0363	-.06***	-.0663	-.0817*	.0497	.0006
	(.0266)	(.0207)	(.0296)	(.023)	(.0491)	(.0418)	(.0399)	(.0314)
Website Design	-.0367*	-.015	-.0244	.0001	-.0951**	-.0534	-.0127	-.0112
	(.0199)	(.0154)	(.0238)	(.0185)	(.0401)	(.0341)	(.0292)	(.0229)
Topic1	-.0885***	-.125***	-.0663**	-.1215***	-.1241**	-.1642***	-.1534***	-.1603***
	(.0255)	(.0198)	(.0285)	(.0221)	(.0598)	(.0511)	(.0398)	(.0313)
Topic2	-.0719***	-.0707***	-.0656**	-.0786***	-.0104	-.0496	-.1384***	-.0969***
	(.0272)	(.0211)	(.0304)	(.0236)	(.0591)	(.0503)	(.0419)	(.033)
Topic3	-.1058***	-.0933***	-.075***	-.084***	-.0594	-.0271	-.1696***	-.1323***
	(.0258)	(.02)	(.029)	(.0224)	(.0587)	(.05)	(.0409)	(.0321)
Topic4	.0505*	.0152	.0704**	.0213	.1404**	.0784	.0731	.0308
	(.0278)	(.0216)	(.0316)	(.0244)	(.0606)	(.0515)	(.0455)	(.0358)
Topic5	.0509	.0332	.0667	.0392	.03	.0184	.1925***	.1245**
	(.0388)	(.0301)	(.0435)	(.0337)	(.0847)	(.072)	(.064)	(.0503)
Employer Attributes								
Ln(E_tenure)	-.0088***	-.0076***	-.0137***	-.0094***	-.0023	-.0038	-.0165***	-.0134***
	(.0018)	(.0014)	(.0019)	(.0015)	(.0051)	(.0043)	(.0024)	(.0019)
Competition Structure								
Number of Workers		.0058***		.0061***		.0056***		.0062***
		(.0002)		(.0002)		(.0005)		(.0002)
Ratio_highquality_workers		1.2789***		1.2765***		.8789***		1.1844***
		(.0198)		(.0219)		(.045)		(.0285)
_cons	.6548***	-.3202***	.679***	-.2842***	.3402***	-.3389***	.72***	-.255***
	(.0426)	(.0353)	(.0474)	(.0391)	(.0943)	(.0857)	(.0656)	(.0548)
Observations	10642	10642	9214	9214	1539	1539	5396	5396
R-squared	0.12	0.47	.0895	.4542	.3009	.4971	.1218	.4574

Standard errors are in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table E2 Estimated Coefficients for 2SLS in Different Samples

	PSM		No Prior Work Experience		Seekers with Two Schemes		PSM-Keywords	
Second Stage	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Typehut	.4791***	.6198***	.4863***	.6684***	.6406***	.7943***	.4359***	.5783***

	(.056)	(.0895)	(.079)	(.126)	(.1045)	(.1424)	(.0962)	(.1444)
Job Attributes								
Payment	.0004**	.0013***	.0004*	.0016***	-.0001	.002***	.0008***	.0019***
	(.0002)	(.0003)	(.0002)	(.0004)	(.0006)	(.0007)	(.0003)	(.0005)
Duration	.0004	.0005	.0014	.0016	-.0096**	-.0115**	.0017	.0019
	(.0018)	(.0018)	(.0019)	(.0019)	(.0047)	(.0047)	(.0022)	(.0022)
DescriptionLength	-.0002**	-.0002**	-.0002**	-.0002**	.0001	.0001	-.0002	-.0002
	(.0001)	(.0001)	(.0001)	(.0001)	(.0002)	(.0002)	(.0001)	(.0001)
Num_skills	.0081	-.0049	.0073	-.0086	.0093	-.0145	-.0003	-.0085
	(.0059)	(.0069)	(.0073)	(.0077)	(.0164)	(.0183)	(.0089)	(.01)
Job Complexity	.0018	.0016	.0024	.0024	.0018	.0028	-.0014	-.0018
	(.0028)	(.0028)	(.0031)	(.0031)	(.0073)	(.0073)	(.0038)	(.0038)
Graphic Design	-.0204	.0124	-.015	.0301	.0831	.1185	-.0011	.0123
	(.0411)	(.0417)	(.0464)	(.048)	(.0926)	(.0937)	(.0637)	(.0641)
Website Design	-.024	-.0286	-.0353	-.045	-.1038	-.1252*	.0401	.0401
	(.0309)	(.0308)	(.0374)	(.0374)	(.0758)	(.0753)	(.0467)	(.0466)
Topic1	-.0216	.0094	-.0319	.0108	.0528	.1268	-.058	-.0341
	(.0403)	(.0418)	(.0456)	(.0482)	(.1136)	(.1162)	(.0666)	(.069)
Topic2	-.006	.0043	-.0172	.0016	.0629	.1118	.0569	.0598
	(.0427)	(.043)	(.0487)	(.0496)	(.1106)	(.1113)	(.0696)	(.0696)
Topic3	-.0644	-.0526	-.0782*	-.0583	-.029	-.0314	-.1013	-.0928
	(.0409)	(.0413)	(.0464)	(.0473)	(.1105)	(.1105)	(.0688)	(.0694)
Topic4	-.001	.0076	-.0065	.0063	.0429	.0788	-.0354	-.0283
	(.0433)	(.0432)	(.0499)	(.0498)	(.1131)	(.1131)	(.0734)	(.0732)
Topic5	-.1347**	-.1348**	-.1836***	-.183***	-.0006	.0216	-.2087**	-.2067**
	(.0604)	(.0602)	(.069)	(.0688)	(.1583)	(.1586)	(.1047)	(.1044)
Employer Attributes								
Ln(E_tenure)	.0012	.002	.0014	.0018	.0099	.012	.0026	.0036
	(.0029)	(.0029)	(.0032)	(.0033)	(.0095)	(.0096)	(.0043)	(.0044)
Competition Structure								
Number of Workers		-.0028***		-.0035***		-.0063***		-.0023**
		(.0006)		(.0008)		(.0012)		(.001)
Ratio_highquality_workers		-.5195***		-.6242***		-.6566***		-.5074***
		(.1298)		(.1738)		(.1812)		(.1921)
_cons	4.4343***	4.75***	4.4288***	4.7869***	4.2446***	4.7268***	4.5055***	4.8097***
	(.0764)	(.0778)	(.0938)	(.0881)	(.1804)	(.1986)	(.1283)	(.1193)
Observations	10642	10642	9214	9214	1539	1539	5396	5396
R-squared	0.02	0.02	.011	.015	.0592	.0614	.0148	.0191
Adj R^2	0.01	0.02	.0077	.0115	.0399	.0408	.0091	.013
Aic	23559.04	23496.79	20994.01	20960.494	3520.5677	3521.0328	11935.071	11915.566
Bic	23791.76	23744.06	21222.121	21202.862	3691.4121	3702.555	12146.06	12139.742

Standard errors are in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Appendix F Image Analysis

F1. Procedure

Due to the high computational cost, it was not feasible for us to analyze all the images. Therefore, we adopted the following steps to sample images for the analysis. First, we classified contests into 10 groups based on the payment distribution of all contests. Then, from each group, we randomly sampled 100 contests and downloaded the submissions from those sampled contests.

To detect similarity among submitted designs, we need an image similarity calculation method. We developed a Python program with Turi Create module, a deep learning toolset for creating Core ML models, to calculate the similarity distance among images. Turi Create uses Convolutional Neural Nets (CNNs), one of the most popular methods in the computer vision domain (Krizhevsky et al. 2012). Turi Create has a pre-trained CNN classifier on a large dataset with 1.2 million images. The image classifier transfers each image data into multiple layers of feature vectors and creates a nearest neighbors model

with those feature vectors. With the nearest neighbors model, the distance function calculates the distance among vectors. By using this approach, we calculated the similarity distance (where a smaller distance indicates greater similarity) between all pairs of submissions within a given contest. Then, we calculated the average similarity distance among all design submissions for each contest.

To illustrate examples of image similarity distance, we present four design submissions from the same contest in Figure F1. The first image serves as the reference submission. The second image represents the most similar submission, with a computed similarity distance of 6.12. The third one is the design submission with the median distance score, which is 12.72. The last one the most dissimilar design submission, whose distance score is 24.24.

	1	2	3	4
1	0			
2	6.12	0		
3	12.72	12.61	0	
4	24.24	23.96	21.29	0

Figure F1. Similarity Distance Scores Among Pairwise Images

F2. Statistics on Key Measures in Section 6.3.1 and Section 6.3.2.

We report summary statistics for image similarity distance, creativity-related keyword measures in job descriptions, and information volume across the different subsamples used in the copy behavior analysis in Section 6.3.1 and Section 6.3.2. These statistics are presented in Table F1.

Table F1. Statistics on Key Measures in Section 6.3.1 and Section 6.3.2.

Variables	Samples	Obs.	Mean	Std. Dev.	Min	Max
Avg similarity distance (Design image)	All Contests	989	37.416	5.186	17.691	58.178
	Contests with high competition	743	37.288	4.947	18.006	53.146
	Contest with low competition	194	37.920	5.728	17.691	55.867
	Solution Level	19,878	37.058	7.744	7.755	70.282
Number of creativity-related keywords	All Contests	989	1.268	2.153	0	12
	Contests with high competition	743	1.341	2.222	0	12
	Contest with low competition	194	1.098	2.037	0	9
	Solution Level	19,878	1.301	2.175	0	12
Number of feedback	All Contests	989	9.198	23.103	0	363
	Contests with high competition	743	10.035	25.970	0	363
	Contest with low competition	194	6.768	10.339	0	64

Reference

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *In Advances in neural information processing systems*. 1097-1105.

Appendix G. Solution Diversity and Job Quality

Having established that a larger number of workers reduces solution diversity, we next examine how this diversity affects overall submission quality in contest-based design jobs. To measure diversity, we calculate the pairwise distances between all submissions within each contest and use the average distance as our metric for overall design diversity. For submission quality, we employ the average rating across all submissions as our measure of contest-level performance. The analysis model is specified in Equation (G1).

$$Average\ Submission\ Quality_j = \beta_0 + \beta_1 Avg\ Solution\ Distances_j + \beta_2 JobAttributes_j + \beta_3 EmployerAttributes_j + \beta_4 CompetitionStructure_j + Month_j + DOW_j + \varepsilon_j \quad \text{(G1)}.$$

This analysis reveals a positive relationship between solution diversity and submission quality, as shown in Column (1) of Table G1. The results demonstrate that contests with more diverse submissions tend to produce higher-quality work overall. This finding suggests that while increasing the number of participants may reduce diversity, maintaining diverse approaches is crucial for achieving superior outcomes in creative contests.

Table G1. The Effect of Number of Workers on Design Distance

	(1)
	Average Submission Quality (Rating)
Average solution distances	.0103* (.0057)
_cons	.4459 (.9553)
Job Attributes	√
Employer Attributes	√
Competition Structure	√
Month and Day Effects	√
r2 (r2_a)	0.14 (0.10)
aic	2606.3294
bic	2831.3442
N	984

** p<0.1; ** p<0.05; *** p<0.01.*
Note: Job Attributes: Payment, Duration, DescriptionLength, Num_skills, Job complexity, Design Types, Description Topics, Creativity
Employer Attributes: Employer Experience, Region, Income;
Competition Structure: Number of workers, Ratio_highquality_workers, Max, Skewness, and Variance of workers' ability

Appendix H. Heterogenous Effects of Compensation Scheme on Job Quality

Thus far, we have compared job quality between contract-based and contest-based jobs and examined the mechanisms underlying the relatively lower quality observed in contest scheme. While limited information and the prevalence of copying behavior emerge as primary drivers of reduced performance in contests, it remains unclear under what specific conditions contest-based jobs might yield improved outcomes. We investigate heterogeneous treatment effects by examining whether the effectiveness of contest-based jobs varies with the level of monetary incentives.

Specifically, we assess whether prize amount moderates the effect of job schemes on job quality. To do so, we replicate the analysis from Model (2) using the matched sample, including an interaction term between job scheme and prize amount. The results are presented in Columns (1) and (2) of Table H1.

The findings reveal that higher monetary amount attenuates the positive effect of contract-based schemes on job quality. When the prize amount is relatively low (i.e., less than $245), contract-based jobs outperform contests in terms of quality. However, as prize amounts increase beyond this threshold, the relative advantage of contracts diminishes and eventually reverses. In high-incentive settings (i.e., prize amounts exceeding $245), contest-based jobs outperform contract-based jobs in generating higher-quality outputs.

This reversal likely arises because greater financial incentives in contests motivate workers to exert higher effort, thereby compensating for the limitations associated with restricted information. Furthermore, substantial rewards may deter low-effort imitation behavior, encouraging participants to invest in the creation of original and high-quality submissions.

Table H1. The Findings on The Heterogeneous Effects of Schemes

DV: Quality	**(1)**	**(2)**
Scheme (contract = 1)	.2171*** (.0144)	.2041*** (.0175)
Payment	.0008*** (.0002)	.0008*** (.0002)
Scheme × Payment	-.0009*** (.0002)	-.0009*** (.0002)
_cons	4.5328***(.0383)	4.5056***(.0405)
Job Attributes	√	√
Employer Attributes	√	√
Competition Structure		√
r2 (r2_a)	0.03 (0.03)	0.03 (0.03)
N	24868	24868

** p<0.1; ** p<0.05; *** p<0.01.*
Note: Job Attributes: Duration, DescriptionLength, Num_skills, Job complexity, Design Types, Description Topics
Employer Attributes: Region, Income, Tenure;
Competition Structure: Ratio_highquality_workers, Max, skewness, and variance of workers' ability