American Economic Review 2014, 104(11): 3565–3599
http://dx.doi.org/10.1257/aer.104.11.3565

Inefficient Hiring in Entry-Level Labor Markets†

By Amanda Pallais*

Hiring inexperienced workers generates information about their abilities. If this information is public, workers obtain its benefits. If workers cannot compensate firms for hiring them, firms will hire too few inexperienced workers. I determine the effects of hiring workers and revealing more information about their abilities through a field experiment in an online marketplace. I hired 952 randomly-selected workers, giving them either detailed or coarse public evaluations. Both hiring workers and providing more detailed evaluations substantially improved workers' subsequent employment outcomes. Under plausible assumptions, the experiment's market-level benefits exceeded its cost, suggesting that some experimental workers had been inefficiently unemployed. (*JEL* J23, J24, M51)

Young workers are more likely to be unemployed than older, more experienced workers.[1] A key question in designing policies to improve young workers' labor market outcomes is whether their poor outcomes result from human capital deficiencies or barriers to labor market entry. If it is the former, then these workers may need to engage in intensive education or training programs to succeed in the labor market. If it is the latter, then programs that simply give these workers a foot in the door may have long-lasting benefits.

This paper evaluates whether inexperienced workers would benefit, on average, from simply obtaining a job because it would give them a chance to demonstrate their abilities. Employers are uncertain about the abilities of inexperienced workers. Hiring these workers generates information about their abilities (e.g., Farber and Gibbons 1996; Altonji and Pierret 2001). This information is valuable because it allows firms to hire higher-ability workers in the future. But, firms might not have sufficient incentive to generate this information because, if it is partially

*Department of Economics, Harvard University, Littauer Center 234, 1805 Cambridge Street, Cambridge, MA 02138, and NBER (e-mail: apallais@fas.harvard.edu). I would especially like to thank my advisors David Autor, Esther Duflo, and Daron Acemoglu. I would also like to thank Jason Abaluck, Leila Agha, Josh Angrist, Bob Gibbons, Michael Greenstone, Jon Gruber, Lawrence Katz, Daniel Keniston, Danielle Li, Christopher Palmer, Iuliana Pascu, Jim Poterba, David Powell, Michael Powell, Simone Schaner, Joe Shapiro, Pian Shu, Jean Tirole, Juuso Toikka, Sarah Turner, Heidi Williams and seminar participants at Booth, Boston University, Columbia, Cornell, Harvard, Michigan State, MIT, Princeton, Stanford, UC-Irvine, UC-San Diego, University of Chicago, University of Virginia, Wharton, and Yale for their many helpful comments and suggestions. I am grateful to Anand Hattiangi, Dmitry Diskin, and the oDesk Corporation for help understanding and accessing the oDesk data. Financial support from the George and Obie Shultz Fund, David Autor, and Esther Duflo is gratefully acknowledged. This project received IRB approval. The author declares that she has no relevant or material financial interests that relate to the research described in this paper.

†Go to http://dx.doi.org/10.1257/aer.104.11.3565 to visit the article page for additional materials and author disclosure statement(s).

[1] For example, in 2013, the unemployment rate of US workers 20 to 24 years old was 12.8 percent, compared with only 6.3 percent for workers 25 to 54 years old. These statistics are from the Bureau of Labor Statistics (2014).

public, workers receive part of its value as higher earnings. Hiring workers is costly. Managers must spend time explaining the jobs to workers and monitoring workers' progress. Moreover, firms incur an opportunity cost of lost time if jobs are not completed correctly or timely. If workers cannot compensate firms for hiring them and producing the information, for example because a minimum wage or adverse selection (e.g., Weiss 1980) prevents wages from falling or because bonding is prohibitively difficult (e.g., Dickens et al. 1989), firms will hire inefficiently few entry-level workers.

Through a field experiment in an online marketplace, this paper assesses the impact of (i) giving jobs to relatively inexperienced workers and (ii) giving the market more information about workers' job performance on their future employment outcomes and the market as a whole. The online marketplace, oDesk, consists of workers all over the world who complete approximately 200,000 hours of work per week remotely.[2] Importantly, when an oDesk employer terminates a job, it is required to publicly disclose a rating of the worker on a one-to-five scale and can, if it chooses, provide a short comment.

In this experiment, I invited low-wage data-entry specialists to apply for 10-hour data-entry jobs. When workers applied, they proposed hourly wage rates for the job. The 3,767 workers who applied proposing wages of $3 per hour or less formed the experimental sample. Workers in the sample were randomized into three groups: two treatment groups (with 476 workers in each) and a control group (containing the remaining 2,815 workers). The size of the treatment groups (together approximately 25 percent of the sample) was determined by financial constraints. I did not hire control group workers. I hired workers in both the "coarse evaluation treatment" and "detailed evaluation treatment" groups. I provided workers in both treatment groups with a public one-to-five rating, calculated from their actual performance statistics and normed to match the distribution of ratings in the market. The ratings were calculated without reference to what treatment group the workers were in. The difference between the two treatment groups was the amount of information about workers' performance that was in the public comment I provided. The comments that workers in the coarse evaluation treatment received were designed to be as uninformative as the comments typically provided in the marketplace. However, in the detailed comment treatment, workers receiving a rating of four or higher received a detailed comment with objective information about their data entry speed, accuracy, following of directions, and timely task completion. Due to IRB restrictions, I was not allowed to provide detailed evaluations to workers with low ratings. Thus, for workers earning below four, the detailed evaluation treatment was identical to the coarse evaluation treatment. Because of the large fraction of workers earning high ratings on oDesk, only 17 percent of workers earned ratings below four.

Using the marketplace's administrative data, I then observed the experimental workers' subsequent oDesk employment outcomes. Workers benefitted from obtaining an experimental job. After the experiment, workers in the coarse evaluation treatment were more likely to be employed, requested higher wages, and had higher earnings than control group workers. In the two months after the experiment,

[2] All statistics about the marketplace describe oDesk in July 2010, immediately after the experiment. This statistic was generated by oDesk (2012).

inexperienced workers' earnings approximately tripled as a result of obtaining a job. Providing workers with more detailed evaluations also increased their earnings and the wages they requested; this is consistent with the idea that more information about worker quality makes workers more valuable to firms. Both the effect of receiving a job and the effect of receiving a detailed evaluation grew over time. But, as theory suggests, the benefits of detailed evaluations were not universal: detailed performance evaluations helped those who performed well and hurt those who performed poorly.

The interpretation of these results depends on whether the treatments affected outcomes by revealing information about worker ability or through another mechanism. I consider whether five alternative mechanisms could explain the experiment's results: (i) the treatment jobs provided human capital; (ii) the act of hiring workers led the market to positively update its belief about their abilities; (iii) the fact that workers received detailed evaluations led the market to positively update its belief about their abilities; (iv) obtaining an experimental job induced workers to apply to more oDesk jobs, but did not change employers' beliefs about workers' abilities; and (v) I gave workers more positive ratings than they deserved. None of these alternative explanations can explain all the experiment's results. For example, the first two explanations cannot explain the results of the detailed evaluation treatment. Moreover, it does not appear that the act of hiring workers in itself led the market to positively update its beliefs about workers' abilities: obtaining a job did not improve workers' employment outcomes when the market observed only that they had been hired (and not their evaluations), but outcomes improved immediately after the evaluations became public.

The fact that the treatments benefitted treatment group workers does not imply that they increased overall market welfare. That is, treatment group workers could have simply displaced other equivalent oDesk workers. I do not have experimental variation that allows me to estimate the effect of the experiment on the market as a whole. However, to shed light on the experiment's effect on the market, I compare how employment and wages changed after the experiment across oDesk's 74 job categories, based on the intensity with which the categories were affected by the experiment. This analysis relies on the assumption that without the experiment, employment and wages would have changed similarly in more- and less-affected categories. I find that, after the experiment, total employment increased in more-affected job categories relative to less-affected ones, while average wages decreased in the former relative to the latter. I use these results to benchmark the experiment's effects on total market surplus. Under plausible assumptions, the benefits to market participants of the increased employment induced by the experiment outweighed the experiment's social cost (the time workers spent working and I spent managing them). It suggests that inefficiently low hiring of novice workers led to diminished employment and output in this market.

This paper directly relates to three strands of the literature. First is the literature on firm provision of general skills training. Public information about workers' abilities is similar to general human capital. While public information does not increase workers' output conditional on their working, it increases the aggregate output of a group of workers by allowing firms to hire only the highest-ability workers. Thus, discovering a worker's ability is similar to general skills training: both produce future productivity benefits, but require up-front investments. Becker (1964) shows that, because workers receive the benefits of general skills training, it will be underprovided if firms cannot be compensated for providing it. More recent work shows that if firms have monopsony power in the labor market (Acemoglu and Pischke 1999), obtain private information about worker quality (Acemoglu and Pischke 1998), or can use training to screen workers (Autor 2001), they will provide some general skills training. There is some empirical evidence that firms provide general skills training that is not fully offset by lower wages (e.g., Loewenstein and Spletzer 1998; Autor 2001). However, neither the theoretical nor the empirical literature shows that firms recoup the full value of their training investments resulting in their providing the optimal level of training. This paper provides evidence that information about worker ability is underprovided by firms.

If not provided by firms, general skills training can be provided by schools; but output and information about workers' abilities are jointly produced. Worker attributes such as reliability, enthusiasm, and maturity are difficult to verify outside of an employment context. Thus, if firms do not generate this information, there may be few alternative mechanisms for its production.

The most closely related paper to this one is Terviö (2009), which proposes that the combination of hiring costs and publicly-observable performance could generate inefficiently high wages and low employment for CEOs and entertainers. This paper shows that a similar inefficiency may lead to inefficiently low employment in entry-level labor markets, making interventions that give workers a chance to demonstrate their abilities particularly effective. There is substantial uncertainty about the abilities of entry-level workers, particularly those with little education and few credentials. Firms often cannot conceal whether they have fired, retained, or promoted a worker, an important signal of entry-level worker performance. Workers' expected output is low, so minimum wages may be binding, making it difficult for workers to compensate firms for hiring them.

Entry-level labor markets have institutions and policies that, in theory, reduce this inefficiency: some reduce firms' cost of hiring inexperienced workers and some directly credential workers in return for compensation. For example, internships and hiring subsidies for young workers reduce firms' costs of hiring inexperienced workers. Fixed-term contracts (in Europe) reduce firms' hiring costs by allowing firms to dismiss low-ability young workers more easily. In many occupations, workers can pay to take tests demonstrating their competence at a given activity. Temporary help firms play a similar role. They screen workers for a variety of competencies (e.g., Microsoft Word skills) and, in return for endorsing the worker, receive part of the worker's compensation.[3] However, it is difficult for private firms to entirely remove the inefficiency. While tests can determine workers' skills and aptitude, they may not capture workers' dedication and enthusiasm. Moreover, while policies may reduce firms' costs of hiring workers, in most cases, workers are not legally allowed to pay firms for hiring them or agree to indentured servitude contracts.

[3] oDesk, the online marketplace that is the setting for the paper's empirical work, has similar institutions. Workers can take approximately 300 skills tests to demonstrate proficiency in subjects ranging from English to Microsoft Excel and C++. Moreover, Stanton and Thomas (2013) discuss agencies, which allow established workers to vouch for inexperienced workers in return for a percentage of their earnings.

Finally, this paper relates to the large literature evaluating whether programs that help young and disadvantaged workers enter the labor market can improve their long-term outcomes. The findings from this literature are mixed.[4] This paper has two primary advantages relative to this literature. First, while other programs typically combine many different elements, this experiment is able to isolate the effect of information about workers' abilities from on-the-job training, job placement services, or stigma from participating in a given program. Second, because the experiment was so large relative to the marketplace, this paper can address the concern that benefits for hired workers came entirely at the expense of other non-studied workers. My results suggest that the benefits to experimental workers outweighed any cost to other oDesk workers.

The rest of the paper is organized as follows. Section I describes the online marketplace, lays out the experimental design, and assesses the randomization. Section II presents the theoretical framework and generates testable predictions for the effects of the experimental treatments. Section III analyzes the worker-level effects of the experiment and discusses whether they could have been generated by alternative mechanisms. Section IV estimates the effect of the experiment on net market surplus. Section V concludes and discusses the application of these results to other settings as well as public policies that could potentially reduce the inefficiency.

I. Experimental Context and Design

A. *The Marketplace*

oDesk is an online marketplace in which employers hire independent contractors to perform tasks remotely. The marketplace is large: immediately following the experiment in July, 2010, oDesk workers completed approximately 200,000 hours of work per week, the equivalent of 5,000 full-time employees. oDesk workers are located around the world. Right after the experiment, a plurality (37 percent) lived in the United States, while India (15 percent) and the Philippines (14 percent) were the next most common countries of residence.[5] In contrast, approximately 80 percent of employers were located in the United States. The most common types of jobs on oDesk were web programming, website design, and data entry. In general, oDesk jobs were shorter than traditional, offline jobs. But, there was a lot of variation in the length of oDesk jobs: some jobs lasted for only a few hours, while others constituted full-time employment. The average job lasted 69 hours. Repeat interactions occurred, but were not the norm. The average worker with any employment had 5.9 jobs with 4.7 unique employers since she joined the marketplace.

Employers posted job openings in 74 job categories. These postings described the job and any necessary worker characteristics. When employers posted, they chose whether to offer hourly or fixed wage jobs. Hourly jobs, the type created in this experiment, constituted 70 percent of jobs on oDesk. In these jobs, oDesk

[4] See, for example, Holister, Kemper, and Maynard (1984); Couch (1992); Bell and Orr (1994); Bloom et al. (1997); Bloom et al. (2009); Redcross et al. (2009); and Autor and Houseman (2010). Stanley, Katz, and Krueger (1998) and Bloom (2010) provide summaries of the literature.

[5] All statistics in this section aside from the total number of hours per week worked on oDesk are from my calculations using the oDesk database.

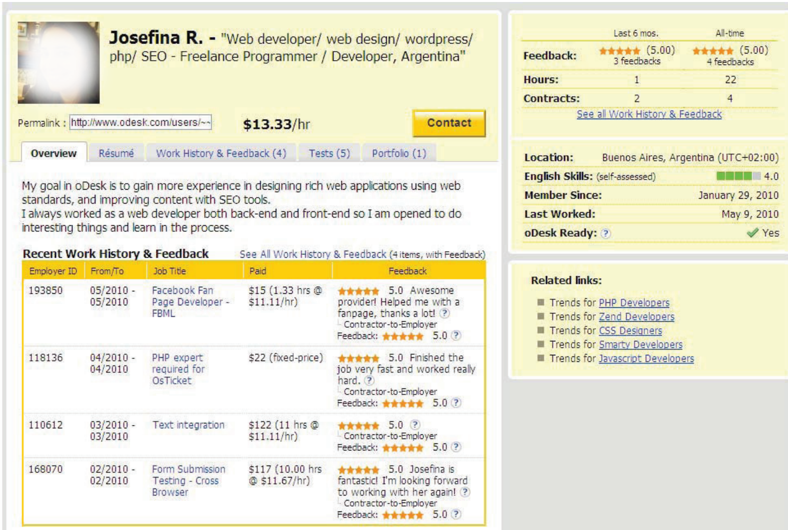

Figure 1. Example oDesk Profile

tracked the number of hours worked and each worker earned an established hourly wage. oDesk guaranteed that the employer would pay for the hours worked, though the employer could stop the job at any time. In a fixed wage job, the worker and employer agreed to a price for the entire project, hours worked were not recorded, and the employer had complete discretion over how much it paid.

Workers posted public profiles, describing their skills and the types of jobs they were seeking. An example is displayed in Figure 1. (This worker was not in the sample because her proposed hourly wage was too high.) Workers could apply directly to jobs; alternatively, employers could search for workers and invite them to apply. When employers searched for workers, they could search for workers with different levels of experience, choosing to contact only workers who had worked a certain number of hours or had a certain feedback score. However, employers rarely invited workers in this sample to apply: less than 9 percent of applications sent by treatment group workers after the experiment were initiated by employers. Under either application method, the worker proposed a price: an hourly wage (in an hourly job) or an amount for the entire project (in a fixed wage job). After reviewing their applicant pools, employers could hire as many or few applicants as they deemed suitable.

Each worker posted her preferred hourly wage rate at the top of her profile. When applying for a job, a worker could suggest a different wage to employers, but employers saw her posted wage as well. As soon as a worker began working in an hourly job, the job title, number of hours worked, and hourly wage were automatically posted to her profile. In fixed wage jobs, the job title and agreed job price were automatically posted. When an employer ended a job, it had to rate the worker from one to five on six dimensions: availability, communication, cooperation, deadlines, quality, and skills. These scores were averaged to form the worker's overall rating for the job. The worker rated the employer on the same six dimensions (before seeing her own rating); these scores were averaged to form the employer's overall rating. Because oDesk wanted the employee to be able to present her side of the story, both composite ratings were automatically posted to the worker's profile. A worker could not remove the ratings without refunding the remuneration received. Employers' ratings were typically very positive: before the experiment, 64 percent of low-wage data entry workers received a rating of exactly five, while 83 percent averaged at least four. Workers and employers could also choose to provide short comments about the employment experience, which were also automatically posted to the worker's profile. Comments were generally one or two positive sentences providing little objective information. Unlike the numerical ratings, workers could remove employer comments without penalty, but overall only 4 percent of oDesk workers did.

In addition to the employer feedback mechanism, oDesk developed a number of ways to let workers demonstrate their abilities. Because workers' listed skills and experience could be hard to verify, oDesk developed its own skills tests: 40 minute, 40 multiple choice question tests on subjects such as written English, Microsoft Word, and C++. Workers' scores and performance relative to other oDesk workers could then be directly posted to their profiles. Workers could also display their "qualifications," certifications from other online platforms, and post a portfolio of their prior work.

Additionally, workers could join agencies, groups of workers typically coordinated by an established worker, in return for a fraction of their earnings. The profile of each agency-affiliated worker contained the agency's average feedback score as well as the worker's own feedback score. Stanton and Thomas (2013) show that agencies were a way for workers to signal their quality. They find that agency-affiliated workers were much more likely to obtain a first job and earned higher wages in their first jobs than non-affiliated workers. However, once the market observed feedback on the workers' own abilities, workers no longer benefitted from being in an agency. While agencies were more common among high-wage workers (only 7 percent of my treatment group workers were in an agency), their presence suggests the difficulties oDesk workers had in developing reputations.

B. *Sample Selection*

I recruited subjects for this experiment by posting hourly data-entry jobs to the marketplace and inviting workers to apply.[6] The jobs were expected to take approximately ten hours and involved entering census records from a PDF file into a Microsoft Excel spreadsheet. I invited an application from every oDesk worker who had a public profile, listed her specialty as data entry, posted an hourly wage of $3 or less to her profile, and had applied for at least one job in the prior three months. Because hiring so many workers at one time would be both logistically difficult and a large shock to the market, I contacted workers in two waves, two weeks apart. Workers were randomly allocated to a wave. The 3,767 workers who applied to the jobs and requested a wage of $3 or less formed the experimental sample.

[6] I posted these jobs from the accounts of 23 different employers. Each employer posted ten separate (but identical) jobs, so that no one employer or job applicant pool would appear too large. Workers in the sampling frame were randomly assigned to an employer and a job.

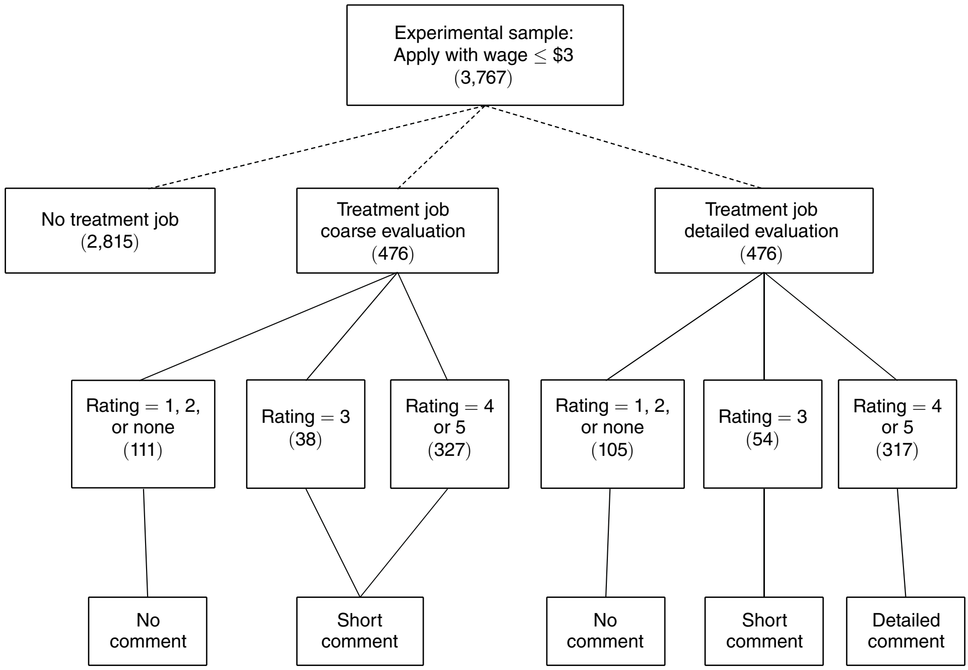

Figure 2. Experimental Design

Note: Dashed lines indicate random assignment.

Online Appendix Table 1 shows the sample selection. Slightly fewer than 10,000 workers fit the sample selection criteria, most of whom had never had an oDesk job. Thirty-nine percent of these workers applied to the jobs, all but 85 of whom requested a wage of $3 or less. Workers with prior oDesk experience were substantially more likely to apply than inexperienced workers (54 percent and 33 percent applied, respectively).

C. *Experimental Protocol*

Figure 2 displays the experimental design. Workers were first randomized into either the control group or one of two treatment groups: the detailed evaluation or the coarse evaluation treatment group. Randomization into any treatment group was stratified on prior oDesk experience, such that workers without oDesk experience had a higher chance of being in any treatment group (32 percent) than experienced workers (15 percent). Conditional on receiving any treatment, all workers had a 50 percent chance of receiving the detailed evaluation treatment. Inexperienced workers constituted approximately three quarters of each treatment group.

The coarse evaluation treatment was designed to be equivalent to being hired (and, thus, evaluated) by a typical employer in the marketplace. The detailed evaluation treatment was identical to the coarse evaluation treatment except that it provided the market with more information about some workers' job performance. Workers in both treatment groups were hired and given a maximum of ten hours over one week to enter the data. They were told that if, after spending ten hours on the task, they had not completed it, they should send the file back unfinished. I recorded objective measures of workers' performance: their data entry speed, their error rate, the date they returned the data file, and three measures of whether they had followed the data entry instructions. I rated all hired workers on a one-to-five scale using a weighted average of workers' scores on these performance measures. The distribution of scores from my job was designed to match the distribution of scores low-wage data entry workers received in the marketplace, adjusted for the fact that a worker in my sample was more likely to be inexperienced than a typical oDesk worker.[7] The scores were calculated in the same way for workers in both treatment groups. Approximately 18 percent of workers did not return the file or log any hours. Under oDesk's protocol, these workers were not rated. Thus, the treatments should be considered as an intent to hire.

The particular treatment group to which workers were assigned affected only the type of comment workers were eligible to receive. No workers in either treatment group received a comment if they earned a rating below three. The remaining workers in the coarse evaluation treatment received an uninformative comment. The remaining workers in the detailed evaluation treatment received a detailed comment if they scored at least a four and an uninformative comment if they scored between three and four. (The human subjects committee permitted detailed evaluations only for workers scoring at least four.)[8] Workers in the detailed evaluation treatment did not know that they would receive a detailed evaluation until it was posted.

The uninformative comment was chosen to be short and positive, like most of the comments in the marketplace. The detailed comment provided objective information on the worker's data entry speed and accuracy, whether the worker met the deadline, and whether she followed the job's instructions. Additionally, it repeated the uninformative comment, so the only difference between the two comment types was the objective information provided in the detailed evaluation.

The uninformative comment read as follows, where only the words in brackets varied by worker.

> *It was a pleasure working with* [*x*].

The detailed comment read:

> [*x*] *completed the project* [*y days before the deadline, by the deadline, z days after the deadline*] *and* [*followed our instructions perfectly, followed our instructions, followed most of our instructions, did not follow our instructions*]. [*x*] *was in the* [*top 10 percent, top third, middle third, bottom*

[7] In fact, the distributions of feedback scores received by experienced and inexperienced workers were not statistically distinct.

[8] MIT's human subjects committee was concerned that giving workers negative evaluations would harm workers. It allowed me to give low numerical ratings, which were essential to the experiment. However, it permitted me to provide detailed comments only to workers who did well overall on the task. There is no censoring of the comment for people receiving a rating of four or above, so the detailed comments do provide negative information about aspects of these workers' performance (e.g., they were in the bottom 10 percent of workers I hired in speed or accuracy).

> *third, bottom 10 percent*] *of providers in speed and the* [*top 10 percent, top third, middle third, bottom third, bottom 10 percent*] *in accuracy. It was a pleasure working with* [*x*].[9]

Because 83 percent of oDesk workers generally earned a rating of at least four and my ratings were a weighted average of workers' performance on the different criteria, many treatment group workers who received a rating of four or five were in the bottom third of speed, accuracy, or both. If employers, particularly those new to oDesk, did not realize that so many workers received high ratings, these detailed comments would have appeared very negative.

I did not hire workers in the control group. However, some of these workers were hired by outside employers while the treatment group worked on my job. This was rare for workers without prior experience: only 4 percent of inexperienced control group workers worked during this period. Unsurprisingly, a higher fraction of experienced control group workers, 27 percent, obtained jobs from other employers during this period. Because inexperienced workers comprised over three quarters of the treatment group, this suggests that fewer than 9 percent of treatment group workers would have been hired during this period in the absence of the experiment.

D. *Data Collection*

I directly collected data on workers' job performance. The remaining worker characteristic and outcome data used in this project are administrative data obtained from oDesk's server with oDesk's permission. oDesk's server automatically records information on workers' profiles, job applications, and employment. The primary worker-level outcomes are measures of workers' employment, earnings, and reservation wages.

I consider three measures of employment: whether a worker obtained any job after the experiment, the number of jobs obtained, and the number of hours worked (in hourly jobs). I also use the wages workers posted to their profiles as a measure of their reservation wages. All workers had to post a wage to their profiles, so this measure is free from selection concerns. I observe the wage workers posted before the experiment and the timing of all subsequent changes to this posted wage, so I can determine the wage posted at any point in time. In a fully competitive market, workers would post their reservation wages. While workers do accept wages below their posted wages, there is no reason to believe the treatment affected the relationship between workers' posted wages and their reservation wages. Finally, I calculate workers' earnings from all oDesk jobs.

Three weeks after the initial randomization, I invited 630 workers to apply to another data-entry job with a fixed wage rate of either \$0.75, \$1, or \$2 per hour.

[9] In order to test whether any effect of the detailed comment was a result of it simply being longer than the coarse comment or signaling that the worker was hired by a larger or more competent firm, I also randomized whether the comment mentioned that the hiring firm was large. I added the (true) sentence "Our organization has hired hundreds of providers on oDesk" to randomly-selected coarse comments and the sentence "This is based on our experience with hundreds of providers on oDesk," to randomly-selected detailed comments. These sentences had no effect on workers' subsequent employment outcomes.

These workers were randomly selected without reference to their prior experience or whether they had been placed into the control group or a treatment group. The invitation was sent from a new employer and workers were randomized into either the \$0.75, \$1, or \$2 job. I recorded which workers applied and offered a job to a randomly-selected 5 percent of applicants. I use data on whether workers applied to this job to calculate their opportunity cost of working in the welfare calculations.

E. *Randomization Assessment*

Tables 1 and 2 assess the randomization and present descriptive statistics about the sample. Table 1 shows that the majority (63 percent) of workers were from the Philippines, while relatively few (under 3 percent) were from the United States. On average, workers without prior experience had been on oDesk for just over 4 months, passed 2.7 oDesk skills tests, and sent about 23 applications. Workers with previous jobs had been on oDesk for about twice as long and sent over seven times as many applications. They had an average of seven previous jobs (the median worker had four).

Table 1 compares the pre-experiment characteristics of workers in a treatment group with those of workers in the control group. It compares workers separately by prior experience because the randomization stratified on this variable. The treatment and control samples look similar based on covariates. Out of the 24 comparisons examined, one is statistically different at the 5 percent level and one is significantly different at the 10 percent level. In neither case is there a significant difference between workers in the treatment and control groups when I pool the sample of workers with and without previous experience and control for workers' prior experience.

Table 2 compares workers randomized into the detailed and coarse evaluation treatment groups. Conditional on being in a treatment group, all workers had a 50 percent chance of being in either treatment group. The table also separately compares the covariates of workers who received ratings of four or five in the two treatment groups since these are the only workers for whom the two treatments differed. In both cases, the randomization produced similar samples.

II. Model

This section provides a simple framework that formalizes the insight that firms will hire inefficiently few inexperienced workers when they do not receive the benefit from discovering talented novices. It then defines two shocks to the market that are the model equivalents of the coarse and detailed evaluation treatments and generates predictions about the effects of these shocks.

A. *Model Setup*

The marketplace comprises a mass 1 of firms and potential workers. Workers (indexed by i) live for two periods (period 0, the "novice" period and period 1, the "veteran" period). Each period, one generation of workers with mass $\frac{1}{2}$ exits the market and a new generation enters. Firms (indexed by j) live for one period.

Table 1—Randomization Assessment: Treatment vs. Control Groups

	No previous job		With previous jobs		
	Treatment group	Control group	Treatment group	Control group	All workers
Posted wage	2.18	2.16	2.23	2.27	1.98
Days since joining oDesk	137	126	251	257	179
Number of applications sent	25*	22*	160	167	27
Proposed wage for treatment job	2.18	2.16	2.29	2.32	2.01
Number of tests passed	2.7	2.7	4.5	4.7	3.5
Number of qualifications	2.9	3.0	4.6	4.8	3.7
Percent with portfolio	7	6	26	25	14
Philippines (percent)	63	61	63	64	63
India (percent)	10	11	10	12	11
Bangladesh (percent)	10	10	15**	10**	10
Pakistan (percent)	6.3	7.0	5.1	4.6	5.9
United States (percent)	2.9	2.6	2.3	2.5	2.6
Number of previous jobs			7.3	6.9	6.9
Average feedback score			4.4	4.4	4.4
Observations	736	1,562	216	1,253	3,767

Notes: Each cell presents the mean value of the indicated characteristic for the indicated group of workers immediately before the experiment. "Qualifications" are certifications from entities other than oDesk that are posted to the worker's profile. A "portfolio" is where a worker posts examples of her prior work.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 2—Randomization Assessment: Detailed vs. Coarse Evaluation Treatment Groups

	All treatment group workers		Workers with ratings of 4 and 5	
	Detailed treatment	Coarse treatment	Detailed treatment	Coarse treatment
Posted wage	2.17	2.22	2.21	2.25
Days since joining oDesk	161	164	163	164
Number of applications sent	58	53	55	53
Proposed wage for experiment job	2.19	2.22	2.22	2.24
Number of tests passed	3.1	3.2	3.2	3.4
Number of qualifications	3.2	3.4	3.4	3.6
Percent with portfolio	12	11	13	12
Philippines (percent)	63	62	67	67
India (percent)	10	10	9	8
Bangladesh (percent)	12	11	11	9
Pakistan (percent)	6.1	5.9	4.7	4.9
United States (percent)	2.1	3.4	1.6	2.8
Fraction with previous job (percent)	23	23	25	25
Workers with previous jobs only				
Number of previous jobs	7.3	6.9	6.2	6.7
Average feedback score	4.6***	4.2***	4.6*	4.3*
Observations	476	476	317	327

Notes: Each cell presents the mean value of the indicated characteristic for the indicated group of workers immediately before the experiment. "Qualifications" are certifications from entities other than oDesk that are posted to the worker's profile. A "portfolio" is where a worker posts examples of her prior work.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Workers vary in their ability (a_i), which is normally distributed in the population.[10] Each firm offers one identical task in which workers' output is $y_i = a_i$. Before a worker's novice period, the market observes her expected ability: $\hat{a}_{i0} = a_i + \varepsilon_{ia}$ where $\varepsilon_{ia} \sim N(0, \sigma_{ai}^2)$ and is independent of a_i. If the worker does not work in her novice period, the market does not update its beliefs about her ability and her expected ability before her veteran period ($\hat{a}_{i1}$) equals $\hat{a}_{i0}$. If a worker is employed in her novice period, the market observes a signal of her output, $\hat{y}_{iM}$:

$$\hat{y}_{iM} = a_i + \varepsilon_{iM} \quad \text{where} \quad \varepsilon_{iM} \sim N(0, \sigma_M^2). \tag{1}$$

Firms use this signal and Bayesian updating to update their beliefs. For simplicity, workers have the same information about their abilities as the market.

Each firm must pay a firm-specific fixed cost, c_j, to hire a worker. This cost includes the time to explain the job to the worker as well as any related overhead costs, such as for equipment or office space. It is continuously distributed across firms on $[0, \infty)$, generating a downward-sloping labor demand curve. Worker i's net marginal product at firm j is $a_i - c_j$.

Each period, firms make wage offers to workers, who can accept at most one offer. Each agent is either in an employment relationship with wages w_{ij} or takes her outside option. Wages are restricted to be non-negative. Both firms' and workers' outside options are zero. Workers and firms are risk neutral and discount the future at rate $\beta < 1$. If agents are indifferent between an employment relationship and their outside option, they enter the employment relationship. The timing of events within each period is as follows:

(i) A new generation of firms and novice workers enters the market.

(ii) Firms observe each worker's novice-period expected ability ($\hat{a}_{i0}$) and novice-period output signal ($\hat{y}_{iM}$), if it exists. They calculate each worker's expected ability.

(iii) Firms make wage offers to workers.

(iv) Workers accept or reject their wage offers.

(v) Hired workers work, producing output y_i and receiving their wages.

[10] The key aspect of this assumption is that it ensures that some workers have expected abilities below firms' hiring costs, and thus, they will not be hired. This assumption seems reasonable on oDesk: 5 percent of the workers I hired charged time for the job, but never turned in the entered data and 30 percent of those who did turn in the data entered over a third of the cells incorrectly. Hiring these workers for an actual job would provide low benefits compared to the time cost of administering the job and the opportunity cost of waiting for the worker to complete the task. In a more general context, it seems reasonable that some workers would have expected marginal products below the cost of hiring them (including any minimum wage). Low-skilled workers who steal or do not show up for work impose large costs on firms compared to the benefits they produce if they perform well.

(vi) Veteran workers and all firms exit the market. Novice workers become veterans.

This model is stylized, but the assumptions are empirically-motivated. The assumptions that firms live only one period and that wages are non-negative rule out bonding and long-term contracts. In the oDesk setting, these assumptions seem reasonable. More than 50 percent of oDesk firms offered only one job ever. oDesk did not allow contracts with negative wages and it would have been difficult for third-world oDesk workers to transfer money to US firms outside of the marketplace. More generally, Dickens et al. (1989) suggests bonding is difficult in the labor market. Workers may be liquidity constrained and bonding may negatively affect workers' attitudes and morale and harm firms' public images. Bonding may generate moral hazard problems where firms have incentives to report that workers performed poorly. Moreover, courts will not enforce contract provisions that call for workers to pay large penalties for poor performance and most indentured servitude contracts are unenforceable.

There was a minimum wage of 0 on oDesk. However, this assumption is not necessary. Any wage rigidity that prevents wages from falling will lead to a similar inefficiency where firms hire too few novice workers. For example, Weiss (1980) presents a model where adverse selection prevents wages from falling. While some unemployed workers would be willing to work for lower wages, because outside options are correlated with unobserved ability, high-ability workers would select out of the job if wages fell. Thus, firms will not decrease wages.

The assumption that firms live only one period precludes firms from having monopsony power over veteran workers as they do in models of asymmetric information (e.g., Waldman 1984 and Greenwald 1986). While there is evidence of asymmetric information (e.g., Gibbons and Katz 1991 and Kahn 2013), there is also evidence that public learning about workers' abilities is important for inexperienced workers and early career earnings dynamics (e.g., Farber and Gibbons 1996; Altonji and Pierret 2001; and Schönberg 2007). It is certainly possible that the oDesk employers that remain in the market have private information about their previous hires. To the extent that private information or mobility costs allow firms to hire previously-employed workers at wages below the workers' expected marginal products, this would increase the benefit of hiring novice workers and reduce the inefficiency. Asymmetric information would also decrease the benefits of being hired for an experimental job because the market would negatively update its beliefs about workers' abilities from the fact that I did not rehire them.

B. *Market Equilibrium and Social Planner's Solution*

PROPOSITION 1: *The perfect Bayesian equilibrium of this game exists and has unique actions along the equilibrium path. There is a threshold, $\overline{c}$, such that all firms with fixed costs $c_j \leq \overline{c}$ will hire a worker, while no firm with $c_j > \overline{c}$ will. All workers with expected ability $\hat{a}_i \geq \overline{c}$ and only these workers will be employed. These workers will earn wages $w_{ij} = \hat{a}_i - \overline{c}$.*

PROOF:

See online Appendix A.

Because there is a mass of workers with expected abilities below every firm's fixed cost, not all workers will be employed. In particular, low-ability novices expected to generate losses when working at the marginal firm will not be hired. However, they would like to work and would compensate firms for hiring them because being hired in their novice period would increase their expected veteran-period earnings. If a low-ability novice does not work, her expected ability will not change and she will not be hired in her veteran period. If she works, with some probability, she will perform well enough to earn strictly positive veteran-period wages. However, novices cannot compensate firms for hiring them: they cannot accept negative wages, post bonds, or commit to accept low wages in the future. This generates the inefficiency. It also implies that this equilibrium would not change if workers knew their own abilities. High-ability workers would be willing to compensate firms more for hiring them than would lower-ability workers, but they would not be able to do so.

In specifying the social planner's solution, it is helpful to define $\hat{a}_{i1H}$: worker *i's* veteran-period expected ability if she is hired in her novice period.

PROPOSITION 2: *The solution of a utilitarian social planner who has the same information as the market and maximizes expected market surplus (the sum of expected worker and firm surplus) is as follows. There exists a threshold, $c^* > \overline{c}$, such that every firm with $c_j \leq c^*$ and only these firms hire workers. All veteran workers with expected ability $\hat{a}_{i1} \geq c^*$ are employed as are all novice workers with $(\hat{a}_{i0} - c^*) + \beta \Pr(\hat{a}_{i1H} \geq c^*) \times E[\hat{a}_{i1H} - c^* | \hat{a}_{i1H} \geq c^*] \geq 0$. A larger mass of novice workers is employed in the social planner's solution than in the market equilibrium c.*

PROOF:

See online Appendix A.

The social planner's solution is the equilibrium that would be enacted if novices could accept negative wages.[11] This solution employs some novices who would be unemployed in the market equilibrium because they generate an expected loss in novice-period work. However, the expected veteran-period benefit these workers receive from novice-period work exceeds this expected loss. In general, hiring novices provides benefits in the subsequent period because it produces information about their abilities that allows them to be more efficiently allocated to either market work or unemployment. While firms have to pay the hiring cost to produce the information, workers obtain its benefits (they earn their expected marginal products which are higher because workers have been more efficiently allocated to sectors).

C. *Model Predictions*

Motivated by my experiment, I consider the comparative statics of two shocks to the market equilibrium. The first models the coarse evaluation treatment. In this

[11] The only constraint preventing the social planner's solution from being enacted in the market is the fact that novices cannot accept negative wages. If this constraint were removed, the resulting market equilibrium would be the social planner's solution.

shock, an outside employer (Employer C) enters the market for one period. This employer has the same information as the market at the beginning of the period, but hires novices without regard to their expected abilities. At the end of the period, it relays an output signal for each hired worker to the market. This output signal has the same distribution as the output signal generated in market jobs. (It represents the one-to-five rating left by employers.) The second shock models the detailed evaluation treatment. As in the prior shock, an outside employer (Employer D) enters the market for one period with the same information as the market, hires novice workers without regard to their expected abilities, and relays an output signal for each hired worker to the market. However, Employer D relays a more precise output signal than Employer C. It relays $\hat{y}_{iD}$ where $\hat{y}_{iD} = a_i + \varepsilon_{iD}$, $\varepsilon_{iD} \sim N(0, \sigma_D^2)$ and is independent of all other variables and $\sigma_D^2 < \sigma_M^2$. (Here, $\hat{y}_{iD}$ represents the one-to-five rating plus the detailed comment.) In both shocks, if a worker is not hired by the outside employer, she remains in the market and is either hired by another firm or takes her outside option.

An important assumption is that neither the fact that a worker was hired by an outside employer nor the precision of the output signal conveys information about the worker's ability. The market updates its beliefs based only on the output signal itself. This assumption seems plausible. In the experiment, the market should not have updated its beliefs based on either of these factors since they were randomly determined (conditional on observables). Moreover, since all oDesk employers saw the same worker profile and there was no face-to-face interaction, there was less scope for an employer to have private information before hiring on oDesk. The detailed evaluations were formulaic, often negative, and commonplace, characteristic of a particular employer, not the sign of a particularly talented worker. Nevertheless, I revisit this assumption when testing for alternative explanations in Section IIIB.

PROPOSITION 3: *Relative to being in the market equilibrium, being hired by Employer C during the novice period weakly increases a worker's expected veteran-period probability of employment, earnings, and reservation wages. It strictly increases these outcomes for workers with* $\hat{a}_{i0} < \overline{c}$. *Relative to being hired by Employer C, being hired by Employer D during the novice period strictly increases a worker's expected veteran-period earnings and reservation wages, regardless of her novice-period expected ability,* $\hat{a}_{i0}$. *It increases her probability of veteran-period employment when* $\hat{a}_{i0} < \overline{c}$ *and decreases her probability of veteran-period employment when* $\hat{a}_{i0} > \overline{c}$.

PROOF:

See online Appendix A.

The key intuition is that being hired by Employer C affects the expected veteran-period outcomes only of workers who would not have been hired in the market equilibrium (workers with $\hat{a}_{i0} < \overline{c}$). If they remain in the market equilibrium and are not hired in their novice periods, their expected abilities will not change and they will not be hired in their veteran periods. Their earnings and reservation wages will equal zero. But, if they are hired by Employer C, there is some probability that they will receive a sufficiently positive output signal to be employed with positive earnings and reservation wages in their veteran periods.

On the other hand, being hired by Employer D (relative to being hired by Employer C) affects workers of all expected abilities. The more precise performance signal causes the market to update its beliefs to a greater extent based on the output signal. This leads to lower probabilities of employment for workers who would have been employed without the output signal (those with $\hat{a}_{i0} > \overline{c}$) and to higher probabilities of employment for workers who would have been unemployed without the signal (those with $\hat{a}_{i0} < \overline{c}$). However, more updating increases expected earnings and reservation wages for workers of all expected abilities.[12] Consider a worker with expected ability $\hat{a}_{i0} > \overline{c}$. If the market did not update its beliefs about her ability, she would be employed with certainty in her veteran period. Her earnings would equal her expected marginal product at the marginal firm, $\hat{a}_{i0} - \overline{c}$. This is the average of positive marginal products for states of the world when her true ability is above the hiring threshold and negative marginal products for states of the world when her true ability is below the threshold. If, instead, the market learned her true ability in her novice period, she would be unemployed in states of the world where her actual marginal product was negative. However, her expected earnings would no longer be driven down by the fact that she could generate a negative marginal product. She would still be rewarded for her positive marginal product in states of the world where her ability was above the hiring threshold, but states of the world where she was unemployed would contribute a zero marginal product. Thus, her expected earnings would be higher when the market knew her true ability even though her expected employment rate was lower.

In this model, a more precise output signal allows workers to be more effectively sorted into either market work or unemployment. This is the mechanism through which the signal's precision affects expected earnings. Without unemployment, the signal's precision would only affect the variance of earnings, not mean earnings.

Sections 4, 5, and 6 of online Appendix A present and prove three additional propositions. The first, Proposition 5, says that the effect of being hired by Employer C, relative to remaining in the market equilibrium, is increasing in the initial uncertainty over the workers' ability (σ_{ai}^2). This is intuitive. When it is more uncertain about a worker's ability, the market puts more weight on the output signal, so being hired by Employer C has a larger effect on the worker's expected veteran-period outcomes. However, counterintuitively, the proposition also says that the effect of being hired by Employer D relative to being hired by Employer C is not necessarily increasing in σ_{ai}^2. Employer D's output signal does cause the market to update its beliefs about workers with uncertain abilities more, but so does Employer C's output signal. Thus, relative to being hired by Employer C, being hired by Employer D does not necessarily have larger effects for workers with more uncertain abilities.

Proposition 6 says that while, ex ante, being hired by Employer D (relative to Employer C) increases all workers' expected earnings, ex post, being hired by Employer D can decrease workers' actual earnings. Workers who receive sufficiently

[12] Note that when I refer here to expected earnings, this refers to workers' expected earnings unconditional on whether they are employed.

poor output signals from Employer D have lower earnings and reservation wages than they would have if they had been hired by Employer C. Similarly, workers who receive sufficiently positive output signals from Employer D have higher earnings and reservation wages.

The stylized model presented above does not include the worker's choice to exit the market because including this choice adds little insight. However, because I observe workers exiting the market, I extend the model in Section 6 of online Appendix A to give workers non-zero outside options and a choice to exit the market before each period. Proposition 7 says that being hired by Employer C in the novice period (relative to remaining in the market equilibrium) weakly increases workers' probabilities of remaining in the market in their veteran periods. This is because being hired by Employer C weakly improves all workers' subsequent market employment outcomes. On the other hand, being hired by Employer D (relative to being hired by Employer C) can increase or decrease the probability that a worker remains in the market. The market updates its beliefs more after a worker is hired by Employer D. This leads to lower probabilities of remaining in the market for workers who would have done so without the performance signal and higher probabilities of remaining in the market for those who would not have.

The final proposition considers the effect of Employer C on market employment, wages, and surplus.

PROPOSITION 4: *If Employer C hires a non-zero fraction of novice workers in the subsequent period, the hiring threshold, $\overline{c}$, will increase to c', total market employment will increase, wages will decrease conditional on expected ability from $\hat{a}_i - \overline{c}$ to $\hat{a}_i - c'$, and market surplus will increase.*

PROOF:

See online Appendix A.

Employer C's hiring increases the mass of veterans with expected ability $\hat{a}_{i1} \geq \overline{c}$ in the subsequent period. This produces an excess supply of labor at the old equilibrium wage, which causes wages to fall (conditional on workers' expected ability), while employment increases. Market surplus increases both because employment increases (more employment relationships generate surplus) and because workers hired by Employer C replace lower expected ability workers (the same mass of jobs produces more surplus).

Because being hired by Employer D decreases the expected veteran-period employment probabilities of workers with $\hat{a}_{i0} > \overline{c}$, Employer D's hiring a non-zero fraction of novices could decrease the mass of veterans with $\hat{a}_{i1} \geq \overline{c}$. This would decrease employment and increase wages, conditional on expected ability. This seems unlikely to happen in the empirical context as far less than 50 percent of control group workers were hired during the treatment period, suggesting that most workers had $\hat{a}_{i0} < \overline{c}$. Thus, I expect that the detailed and coarse evaluation treatments should have the same qualitative effect on market wages, employment, and surplus. However, even if Employer D decreased market employment, it would still increase market surplus by allowing employers to hire higher-ability workers.

Table 3—The Effects of the Treatments on Employment Outcomes During the Two Months after the Experiment

	Workers with no previous jobs			Workers with previous jobs		
	Detailed treatment	Coarse treatment	Control	Detailed treatment	Coarse treatment	Control
Total jobs	0.883	0.731	0.284	2.889	2.037	1.958
p-value: equal to control	(0.000)	(0.000)		(0.005)	(0.809)	
p-value: equal to coarse treatment	(0.307)			(0.079)		
Any job	0.302	0.296	0.117	0.694	0.528	0.545
p-value: equal to control	(0.000)	(0.000)		(0.003)	(0.729)	
p-value: equal to coarse treatment	(0.872)			(0.012)		
Hours worked	11.32	13.49	5.36	74.46	47.52	47.80
p-value: equal to control	(0.002)	(0.000)		(0.006)	(0.977)	
p-value: equal to coarse treatment	(0.482)			(0.074)		
Posted wage	2.31	2.25	2.03	2.68	2.32	2.38
p-value: equal to control	(0.000)	(0.000)		(0.010)	(0.565)	
p-value: equal to coarse treatment	(0.573)			(0.043)		
Earnings	29.72	27.14	10.06	186.84	101.19	120.60
p-value: equal to control	(0.000)	(0.000)		(0.008)	(0.423)	
p-value: equal to coarse treatment	(0.750)			(0.018)		
Observations	368	368	1,562	108	108	1,253

Notes: Each statistic not in parentheses is the mean of the indicated employment outcome for workers in the indicated experimental group. Employment outcomes are calculated for the two months after the experiment; all experimental jobs and earnings are excluded. Each statistic in parentheses is a p-value from a test that the means for the groups indicated by the row and column are equal.

III. Worker-Level Effects

A. *Treatment Effects*

I first assess the effects of the coarse evaluation treatment on workers' subsequent employment outcomes. Proposition 3 predicts that obtaining a job with a coarse evaluation will increase workers' employment rates, earnings, and reservation wages relative to being in the control group. Proposition 5 predicts that, conditional on workers' expected ability, the coarse evaluation treatment will have larger effects on workers about whom the market is more uncertain.

Table 3 compares the employment outcomes of the three experimental groups in the two months following the experiment. Workers are categorized by the treatment they were assigned to at the beginning of the experiment, even though 18 percent of workers in the treatment groups did not accept treatment jobs and workers earning low ratings in the detailed evaluation treatment did not receive detailed comments. Posted wages are measured at the end of the two-month period and the experimental jobs themselves are excluded from any outcomes. The results are presented separately for workers with and without prior oDesk experience as the randomization stratified on this variable. The market should be more uncertain about the abilities of inexperienced workers than experienced workers; inexperienced workers may also have lower expected abilities.

The coarse evaluation treatment's effects reflect the model's predictions: it had positive effects on employment outcomes for workers without prior oDesk experience, but no effect for experienced workers (whose expected abilities are less uncertain). Inexperienced control group workers performed poorly in the labor market: only 12 percent obtained any job in the next two months for an average earnings (unconditional on working) of approximately $10. The coarse evaluation treatment significantly improved all five employment outcome measures for inexperienced workers. It almost tripled the fraction of these workers with any employment from 12 percent to 30 percent and the average worker's earnings from $10 to $27. It also increased the wage these workers posted on their profiles by approximately 10 percent.

Experienced control group workers performed much better than inexperienced control group workers: over half (55 percent) worked on oDesk in the two months after the experiment for an average earnings of $121. However, the coarse evaluation treatment did not significantly improve any of the five employment outcomes for experienced workers.

Table 3 also allows me to assess the effects of the detailed evaluation treatment. Proposition 3 predicts that workers in the detailed evaluation treatment will have higher earnings and reservation wages than those in the coarse evaluation treatment, while the treatment's effect on employment is ambiguous. The table shows that, relative to the coarse evaluation treatment, the detailed evaluation treatment increased experienced workers' average earnings from $101 to $187 and their average posted wages by approximately 15 percent. The earnings gains did not come at the expense of employment; the detailed evaluation increased the fraction of workers with any subsequent employment from 53 percent to 69 percent.

However, the detailed evaluation treatment did not improve average employment outcomes of inexperienced workers relative to the coarse evaluation treatment. While workers in the detailed evaluation treatment had better outcomes on four out of the five employment measures than workers in the coarse evaluation treatment, the differences are neither large nor significant. There are two potential explanations for the somewhat surprising result that the detailed evaluation treatment had larger effects for experienced than inexperienced workers. First, Proposition 5 says that, conditional on expected ability, the effect of the detailed evaluation relative to the coarse evaluation is not necessarily increasing in uncertainty about worker ability. Second, the detailed evaluation should have the most impact for workers with expected abilities near the hiring threshold because there is the most uncertainty about whether those workers should be hired. Experienced workers may have expected abilities that are closer to the hiring threshold. Another potential explanation is that a higher fraction of experienced than inexperienced workers in the detailed evaluation treatment actually received a detailed evaluation (74 percent versus 66 percent), but Table 5 shows that even among workers who were eligible for detailed evaluations, the detailed evaluation treatment had larger effects for experienced workers.

I next consider the robustness of the treatments' effects to the addition of control variables. Panel A of Table 4 displays the results of regressing each employment outcome on indicators for being in the detailed evaluation treatment and for being in the coarse evaluation treatment both interacted with dummies for having prior

TABLE 4—REGRESSION ESTIMATES OF THE EFFECTS OF THE TREATMENTS WITH CONTROLS DURING THE TWO MONTHS AFTER THE EXPERIMENT

	Total jobs (1)	Any job (2)	Hours worked (3)	Posted wage (4)	Earnings (5)
Panel A. Treatments separately					
Detailed treatment × no previous job	0.573***	0.180***	7.14***	0.26***	20.14***
	(0.125)	(0.025)	(2.28)	(0.10)	(6.38)
Detailed treatment × previous job	0.773**	0.132***	24.75**	0.29**	60.20**
	(0.344)	(0.045)	(11.52)	(0.14)	(28.43)
Coarse treatment × no previous job	0.399***	0.171***	8.92***	0.20***	17.05***
	(0.091)	(0.024)	(2.43)	(0.05)	(5.60)
Coarse treatment × previous job	0.151	−0.005	1.80	−0.00	−15.74
	(0.300)	(0.047)	(9.20)	(0.09)	(20.45)
Previous job	0.327***	0.221***	19.25***	0.22***	40.24***
	(0.105)	(0.020)	(3.47)	(0.04)	(8.07)
Control group mean: no previous job	0.284	0.117	5.36	2.03	10.06
Control group mean: previous job	1.958	0.545	47.80	2.38	120.60
Observations	3,767	3,767	3,767	3,767	3,767
Panel B. Treatments combined					
Treatment job × no previous job	0.486***	0.176***	8.03***	0.23***	18.59***
	(0.081)	(0.018)	(1.78)	(0.06)	(4.35)
Treatment job × previous job	0.463*	0.064*	13.29*	0.15*	22.28
	(0.237)	(0.034)	(7.68)	(0.09)	(18.43)
Previous job	0.326***	0.221***	19.19***	0.21***	40.04***
	(0.105)	(0.020)	(3.46)	(0.04)	(8.07)
Control group mean: no previous job	0.284	0.117	5.36	2.03	10.06
Control group mean: previous job	1.958	0.545	47.80	2.38	120.60
Observations	3,767	3,767	3,767	3,767	3,767

Notes: Each column in panel A displays the results of regressing the dependent variable indicated by the column on indicators for being in the detailed evaluation treatment and being in the coarse evaluation treatment, both interacted with whether or not the worker had prior oDesk experience. Each column in panel B displays the results of regressing the same dependent variable on an indicator for being in any treatment group interacted with whether or not the worker had prior oDesk experience. All regressions include a dummy for having prior oDesk experience as well as the covariates listed in footnote 13. Employment outcomes are calculated for the two months after the experiment; all experimental jobs and earnings are excluded. Huber-White standard errors are in parentheses.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

oDesk experience and not having prior oDesk experience. I control for having prior oDesk experience as well as a number of covariates.[13] This table tells a similar

[13] I control for the tests the worker took (dummies for passing one test, two or three tests, and four or more tests, an indicator for having taken the most popular skills test among the sample, and an indicator for passing it), the number of qualifications the worker had (dummies for listing two to four qualifications and five or more

Table 5—The Effects of Receiving a Detailed Evaluation for Workers Earning Ratings of 4 or 5 During the Two Months After the Experiment

	Total jobs (1)	Any job (2)	Hours worked (3)	Posted wage (4)	Earnings (5)
Panel A. Main effect of detailed evaluation					
Detailed treatment × no previous job	0.063	0.004	−3.32	0.20	6.10
	(0.195)	(0.044)	(4.22)	(0.16)	(10.48)
Detailed treatment × previous job	0.642	0.154**	25.27	0.37*	77.67*
	(0.595)	(0.075)	(18.44)	(0.22)	(43.63)
Coarse evaluation mean: no previous job	1.012	0.366	17.94	2.31	33.27
Coarse evaluation mean: previous job	2.358	0.568	54.29	2.43	118.41
Observations	644	644	644	644	644
Panel B. Differential effect of detailed evaluation					
Met deadline × detailed treatment × no previous job	0.060	0.005	−3.39	0.20	6.05
	(0.197)	(0.044)	(4.26)	(0.16)	(10.58)
Missed deadline × detailed treatment × no previous job	−0.333	−0.333	−7.11	−1.08***	−15.79
	(0.274)	(0.274)	(5.84)	(0.42)	(12.97)
Met deadline × detailed treatment × previous job	0.623	0.156**	26.73	0.38	79.17*
	(0.616)	(0.076)	(19.10)	(0.23)	(45.16)
Missed deadline × detailed treatment × previous job	−0.250	0.000	−42.62**	0.07	−49.20**
	(0.548)	(0.436)	(21.13)	(0.65)	(22.93)
Coarse evaluation mean: no previous job	1.012	0.366	17.94	2.31	33.27
Coarse evaluation mean: previous job	2.358	0.568	54.29	2.43	118.41
Observations	644	644	644	644	644

Notes: Each column in panel A displays the results of regressing the dependent variable indicated by the column on an indicator for being in the detailed evaluation treatment interacted with whether or not the worker had prior oDesk experience. The regressions also include a dummy for having prior oDesk experience. Each column in panel B displays the results of regressing the same dependent variable on the four variables listed in the left-most column. Two-way interactions of meeting the deadline with having or not having prior oDesk experience are also included as is an indicator for having prior oDesk experience. Only workers who obtained a rating of at least four in an experimental job are included. Employment outcomes are calculated for the two months after the experiment; all experimental jobs and earnings are excluded. Huber-White standard errors are in parentheses. The "Coarse evaluation mean" rows are limited to workers in the coarse evaluation treatment who earned a rating of at least four.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

story to the previous one. The coarse evaluation treatment only improved outcomes for inexperienced workers. However, its effects on inexperienced workers across each of the five outcomes were large and significant. For example, on average, it increased inexperienced workers' earnings by $17.05, which exceeds the average amount I spent to hire them ($15.93).

Relative to the control group, the detailed evaluation treatment had large effects for both experienced and inexperienced workers. On all but one outcome (the

qualifications), the number of oDesk applications the worker sent before the experiment (dummies for having sent 3 to 5, 6 to 15, 16 to 50, 51 to 100, and over 100 applications), the wage proposed for the experimental job (dummies for offering $1 to $1.99, $2 to $2.99, and exactly $3), the number of jobs the worker had before the experiment, and an indicator for being in the second experimental wave.

probability of obtaining any job), the measured effects of the detailed treatment are larger for experienced workers, though due to the large standard errors, the estimated effects for inexperienced workers and experienced workers are not significantly different. In percentage terms, however the effects are much larger for inexperienced workers. For example, the detailed evaluation treatment increased the total number of jobs and earnings of inexperienced workers by over 200 percent relative to around 40 percent to 50 percent for experienced workers.

Panel B shows the effect of receiving any experimental job, regardless of whether it came with the possibility of a detailed evaluation. It displays the results of regressing each employment outcome on an indicator for receiving any treatment job interacted with both having some and having no prior oDesk experience. It also controls for having prior oDesk experience and the same covariates as the regressions in the previous panel. It shows that, on average, receiving a treatment job improved the employment outcomes for both experienced and inexperienced workers and that both experienced and inexperienced workers earned more as a result of the treatment job than I paid them for completing it. Online Appendix Table 2 shows the effect of the treatments on the pooled sample of experienced and inexperienced workers.

Panel A of Table 5 further probes the impact of the detailed evaluations by estimating their effects on workers who received a score of four or five in my jobs (those who were eligible to receive detailed evaluations). (The treatment a worker was assigned to did not affect her rating and the treatments were identical until the evaluation was made public.) The table displays the results of regressing each employment outcome on the indicator for being in the detailed evaluation treatment group interacted with dummies for having some and having no prior oDesk experience. It controls for whether the worker had prior oDesk experience. It shows that receiving a detailed evaluation relative to a coarse evaluation is estimated to have increased the earnings and posted wages of inexperienced workers by approximately 20 percent and 10 percent, respectively, though neither estimate is significant. Detailed evaluations are estimated to have increased the earnings and posted wages of experienced workers by about 65 percent and 15 percent, respectively, both significant at the 10 percent level. Online Appendix Table 3 shows that these results are robust to adding the control variables, while online Appendix Table 4 shows that detailed evaluations significantly increased the earnings and posted wages of the pooled sample of experienced and inexperienced workers.[14]

The model predicts that both treatments should have the largest effects for workers whose expected abilities are close to the hiring threshold.[15] Intuitively, workers far below the threshold have such low expected abilities that even completing one job successfully cannot raise their expected abilities above the threshold, while workers far above the threshold will be hired and have high earnings regardless of the treatment they are in. Figure 3 presents suggestive evidence that the treatments

[14] Online Appendix Table 3 shows that when the control variables are added, the effect of the detailed evaluations on experienced workers' earnings is significant at the 5 percent level. Moreover, the effects of detailed evaluations on experienced workers' posted wages and the posted wages and earnings of the pooled sample are all significant at the 5 percent level after one month.

[15] The coarse evaluation treatment should have the biggest effects on all outcomes for workers whose expected abilities are just below the hiring threshold. The detailed evaluation treatment should have the biggest positive effects on employment for workers just below the threshold, while it should have large effects on earnings and posted wages for workers just above or just below the threshold.

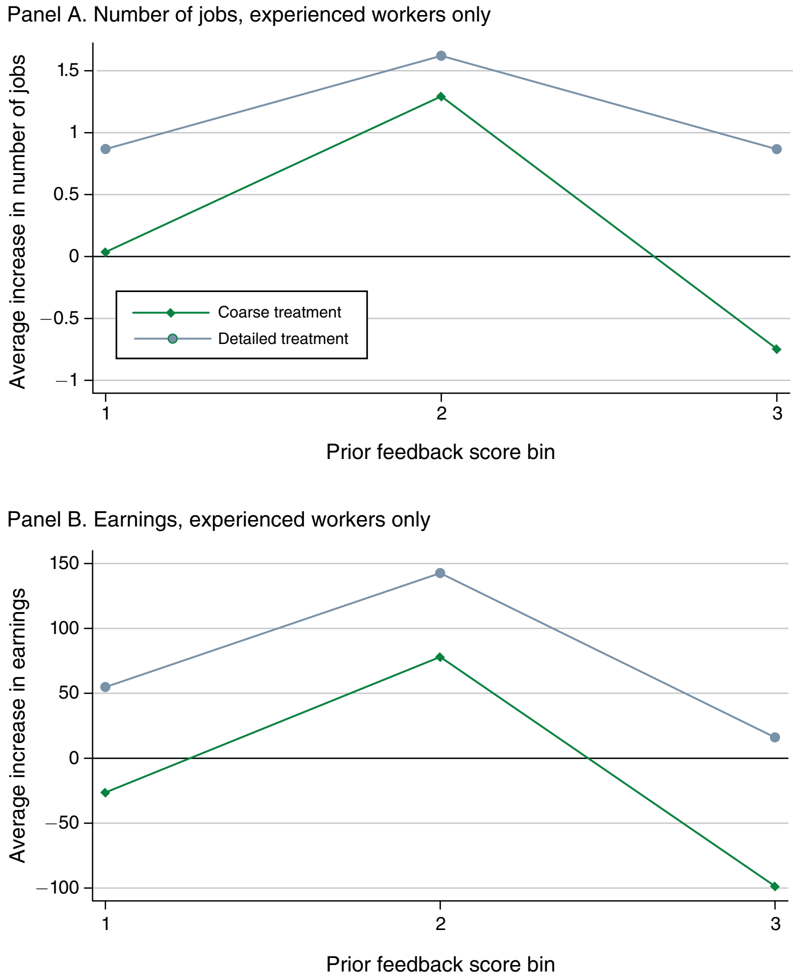

Figure 3. Differential Effect of the Treatments by Prior Feedback Score

Note: Prior feedback score bins 1, 2, and 3 contain experienced workers whose feedback scores put them in the bottom, middle, and top thirds of the feedback score distribution, respectively, before the experiment.

may indeed have had the largest impacts for workers around the hiring threshold. It breaks the sample of experienced workers into thirds based on their average feedback score from employers before the experiment. About 55 percent of experienced workers in the control group were hired in the two months after the experiment, suggesting that workers in the middle bin have expected abilities near the threshold. For each third of workers, the figure plots the average effect of the detailed and coarse evaluation treatments on total jobs (panel A) and earnings (panel B) relative to the control group. Of course, a worker's previous feedback score is only a proxy for her expected ability before the experiment. Moreover, these estimates are noisy and few are significantly different from each other. With these caveats in mind, the figure shows that workers around the threshold (those in the middle bin) benefit the most from both the detailed and coarse evaluation treatments. Workers farther away

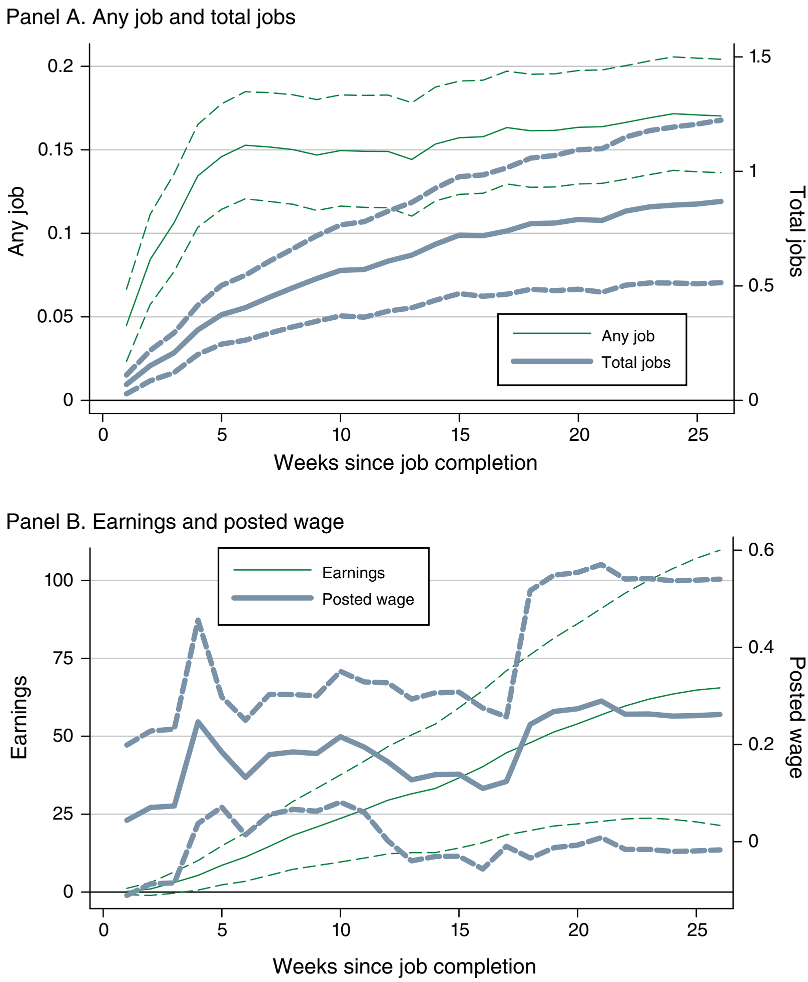

FIGURE 4. EFFECT OF TREATMENT JOB BY WEEK

Note: Dashed lines indicate 95 percent confidence intervals.

from the threshold do not benefit from the coarse evaluation treatment and benefit less from the detailed evaluation treatment.[16]

[16] This analysis may suggest that one reason why inexperienced workers did not benefit as much from the detailed evaluation treatment as experienced workers is that inexperienced workers were farther from the hiring threshold. When I regress prior feedback scores on workers' covariates and use these covariates to predict the feedback scores of inexperienced workers, 76 percent of inexperienced workers are predicted to fall in the bottom decile of experienced workers' feedback scores. I calculate the effect of the treatments for experienced workers in each decile of the feedback score distribution. Then, I use these estimates to predict the effect the treatments would have had on inexperienced workers' earnings and employment if the treatments had had the same effect on inexperienced as experienced workers, conditional on their (predicted or actual) feedback scores. While this is imprecise and far from perfect, the resulting estimates suggest that, if conditional on feedback score, the detailed evaluation treatment had had the same effect on inexperienced workers as it did on experienced workers, the effect of the detailed evaluation treatment (relative to no treatment) on inexperienced workers' earnings would have been negative and its effect on their employment would have been only about a third of the size of the effect for experienced workers.

The effects of the treatments persisted and increased over time. I calculate each worker's cumulative employment and earnings outcomes for each week from one to 26 weeks after the treatment jobs ended. I also record each worker's posted wages at the end of each week. Then, I regress these employment outcomes on an indicator for receiving any treatment job, controlling for prior oDesk experience. Panels A and B of Figure 4 plot the coefficients on the treatment indicator by week. The effect of obtaining a treatment job on earnings more than tripled from \$19.72 after two months to \$65.56 after six months. The effect of the treatment job on hours worked (which is not plotted) almost tripled, increasing from 9.7 after two months to 25.3 after six months (the p-value is only 0.11 after six months). It is striking that the treatment effects are so large given that the treatment was not restricted to those who would most benefit from it: some treatment group workers already had so much prior experience and such high expected abilities that they were unlikely to benefit, while others had such low expected abilities that even successfully completing one job would not increase their expected abilities above the hiring threshold. (Online Appendix Figure 1 disaggregates these results by the rating workers received in my job. It compares the outcomes of the control group with those of treatment group workers earning each different rating in my job.)

The effects of receiving a detailed evaluation also appears to persist over time, but they are much noisier. For example, when I regress cumulative earnings on an indicator for receiving a detailed evaluation in the sample of workers earning fours and fives in my jobs, the effect is positive and significant at the 5 percent level through week seven and significant at the 10 percent level through week nine. After six months, the measured effect on earnings is \$60.17, 2.5 times the effect after two months, but it has a t-statistic of only 1.1.

Finally, I test Proposition 6, which says that detailed evaluations do not always help workers. It says that detailed evaluations that are more negative than both a worker's initial expected ability and the coarse evaluation she would have received impair subsequent employment outcomes relative to coarse evaluations. Detailed evaluations that are more positive than both a worker's initial expected ability and the coarse evaluation she would have received improve outcomes. Because I know all treatment group workers' performance, I can determine the effect of a particular detailed evaluation by comparing the outcomes of workers who received that evaluation with the outcomes of workers in the coarse evaluation treatment who would have received that evaluation had they been in the detailed evaluation treatment.

I consider the effect of revealing whether a worker met the deadline on workers earning fours and fives in my jobs. Since the vast majority of workers earning fours and fives met the deadline, not meeting the deadline was a very negative signal. It is likely more negative than the coarse rating of four or five and most workers' initial expected abilities. Thus, revealing it should have impaired workers' subsequent employment outcomes. Meeting the deadline was a positive signal, but may not have been more positive than a rating of five or many workers' initial expected abilities. Panel B of Table 5 shows the results of regressing each employment outcome on the three-way interaction of (i) an indicator for meeting or missing the deadline with (ii) an indicator for being in the detailed treatment and (iii) an indicator for having or not having prior oDesk experience. The appropriate two-way interactions and fixed effects are included as controls.

Revealing whether an experienced worker met or missed the deadline had large effects on her subsequent earnings: revealing she missed the deadline decreased her earnings in the subsequent two months by $49 on average, while revealing that she met the deadline increased her earnings by $79. The estimates indicate that revealing whether the worker met the deadline had large effects on the other employment outcomes for experienced workers as well, but these are less precisely measured. The effects are typically insignificant for inexperienced workers, though nine out of the ten estimates have the expected sign.

Whether the worker met the deadline is the only piece of information that appeared to matter to employers. Online Appendix Table 5 shows that revealing that workers did not follow all the instructions, were in the bottom third of workers in speed, or were in the bottom third of workers in accuracy did not impair their employment outcomes. This could be because these are less negative signals than missing the deadline. Alternatively, it could be because whether the worker met the deadline was the first piece of information revealed in the comment. Because of the comment's length, the parts about speed and accuracy were not immediately visible on most workers' profiles; one had to click on the continuation to see them. This was unintentional.

B. *Mechanisms*

The previous section showed that the treatments' effects were consistent with the model's predictions. Here, I assess whether the experimental results could also be explained by alternative mechanisms. If the model's mechanism is correct, then workers with higher ratings should have better subsequent employment outcomes, all else equal. This is true in the data: after two months, workers who received ratings of five had earned $34 more than the control group, while workers who received ratings of one and two had earned $23 less.[17] However, since these ratings were not randomly assigned, it could simply be that highly-rated workers would have earned more even without the experiment.

The first alternative is that completing a treatment job provided workers with human capital. Many existing programs that employ disadvantaged workers explicitly provide on-the-job training and, even in those that do not, workers could easily gain human capital by working for several months. It is much less likely that workers accumulated human capital in these jobs. Workers worked a maximum of 10 hours; the average hire worked for only 7.6 hours. Given workers' offline experience, this was a very small increment to their total work experience. I did not provide training or guidance as it was impractical given the number of workers hired at one time. Moreover, workers in both treatment groups completed the same task, so human capital accumulation cannot explain the effects of the detailed evaluations.

The second alternative is that the act of hiring a worker caused the market to positively update its belief about the worker's ability. Hiring a worker would cause the market to positively update its belief if different employers received different signals of worker quality. This is less likely on oDesk than in a traditional labor market because, on oDesk, all employers saw exactly the same resume and there were

[17] Online Appendix Figure 1 shows how these differences grew over time.

no face-to-face interactions. However, employers could interview workers and they might have valued the same information differently.

An empirical test of this explanation utilizes the fact that the market observed that treatment group workers were hired as soon as they began working, but could not see their ratings until 9 to 11 days later. In the week the workers were completing my job (during which the market only observed that they had been hired, not their evaluations), treatment group workers were no more likely to obtain jobs from other employers than control group workers. In contrast, in the week immediately after the rating became public, treatment group workers obtained significantly more jobs and were more likely to be employed than control group workers. This is not simply because treatment group workers were too busy to apply to jobs while they were completing my task: they actually applied to more jobs than control group workers during this period. Moreover, it does not appear that oDesk employers typically penalized job applicants for being currently employed on other jobs. Among oDesk workers with at least twenty previous jobs (whose reputation should not have substantially changed with another evaluation), a given worker's job application was slightly more likely to be successful if she applied while working on another job.

This alternative cannot explain the effects of the detailed evaluations since the market observed that workers in both treatment groups were hired. However, a similar alternative is that the market positively updated its beliefs about a worker's ability based on the fact that she received a detailed evaluation. I think this is unlikely. The detailed comments were often negative. Twenty-seven percent of workers were revealed to be in the bottom third or bottom 10 percent of speed, while 28 percent were revealed to be in the bottom third or bottom 10 percent of accuracy. (These are less than 33 percent because workers earning ratings below four were not given detailed comments.) Less than 19 percent of workers were described as being in the top third or top 10 percent of both speed and accuracy. Perhaps more importantly, these comments were formulaic, contained no subjective information, and were common in the marketplace. In the two months after the experiment, only 12 percent of the applications workers with detailed evaluations sent went to firms that did not have another applicant who had received a detailed evaluation (with the exact same formula). Thus, these comments likely appeared to be the hallmark of a particular employer, not a particularly good worker.

A fourth alternative is that receiving a treatment job induced workers to apply to more oDesk jobs for reasons unrelated to the evaluations. For example, treatment group workers may have realized oDesk jobs were more desirable than they had thought or their initial hiring may have led them to believe it was easy to obtain oDesk employment. However, Proposition 7 also predicts that the coarse evaluation treatment should have induced workers to remain in the market by improving their oDesk employment opportunities. It is difficult to distinguish these two explanations because they have the same prediction. This prediction is borne out in the data: treatment group workers did apply to more jobs than control group workers after the experiment. However, the alternative explanation cannot explain the entire effect of the treatments because the treatments significantly increased the probability that a worker obtained a given job she applied to. Online Appendix B describes these results in much more detail. The explanation similarly cannot explain why receiving a treatment job increased workers' posted wages or the effects of the detailed evaluation, which did not alter workers' application patterns.

A final alternative is that I gave workers more positive ratings and comments than they deserved, despite the facts that my ratings matched the distribution of one-to-five ratings in the market (controlling for the relative inexperience of my workers) and that the detailed comment contained some subjective information. If this explanation were correct, the treatments should have had a diminished effect over time as the market learned more about the workers' true capabilities. The data do not show this pattern: the effects of both treatments on weekly earnings appear to remain constant over time.

IV. Welfare Analysis

In this section, I combine data with assumptions to estimate a lower bound on the experiment's effect on oDesk market surplus in the six months after the experiment. Proposition 4's proof suggests that the experiment increased welfare through two channels: (i) by increasing employment and (ii) by allowing firms that would have hired workers in the absence of the experiment to hire workers with higher expected abilities. In this section, I estimate the effect of the extra employment on market surplus and compare this increase in surplus to the costs of the experiment: the direct costs of the experimental jobs and workers' and firms' opportunity costs of the extra employment. I consider this a lower bound on oDesk market surplus because I cannot estimate the benefit firms received from hiring workers with higher expected abilities. However, it is important to note that my estimate of the extra employment generated by the experiment is not experimentally identified. Moreover, the estimate of market surplus only includes surplus obtained by oDesk workers and firms. For example, when workers increased their oDesk employment, they may have forgone offline jobs. This calculation includes their opportunity costs of not taking the offline jobs, but not the lost profit of the firms that would have hired them or the increased earnings of the workers who took the offline jobs in their absence.

I first estimate the effect of the experiment on market employment. While I do not have experimental variation I can use to estimate this, I can compare the change in employment after the experiment in those oDesk sectors more and less affected by the experiment. For each of oDesk's 74 job categories, I calculate a measure of the experiment's effect on the number of experienced workers in the category: "percent change experience." First, I estimate the number of workers I hired in each category: 952 (the total number of workers I hired) multiplied by the share of treatment group applications sent to jobs in that category in the month before the experiment. Then, I divide this by the number of experienced workers working in that category before the experiment.[18]

Panels A and B of Table 6 show the results of regressing two measures of log employment (log jobs created and log hours worked) in a job category-week on the interaction of percent change experience and an indicator for a week after the experiment. I control for week fixed effects, job category fixed effects, and job category-specific linear time trends. Each regression includes 26 weeks of data before

[18] This fraction averaged 8.5 percent for the entire marketplace, ranging from 55 percent in data entry and 79 percent in e-mail response handling to less than 1 percent in 25 job categories, primarily ones that required specific computer skills such as web programming or game development.

Table 6—Market-Level Effects of the Experiment

Panel A. log jobs					
Percent change experience × after	0.220**	0.246**	0.273**	0.272**	0.352*
	(0.108)	(0.105)	(0.091)	(0.092)	(0.201)
Mean of dependent variable	3.11	3.20	3.25	3.25	3.21
Panel B. log hours					
Percent change experience × after	0.142*	0.153**	0.099	0.102*	0.107
	(0.073)	(0.072)	(0.061)	(0.061)	(0.196)
Mean of dependent variable	6.45	6.55	6.63	6.63	6.59
Panel C. log wages					
Percent change experience × after	−0.418***	−0.437***	−0.265***	−0.315***	−0.034
	(0.112)	(0.097)	(0.095)	(0.096)	(0.149)
Mean of dependent variable	1.86	1.86	1.85	1.85	1.94
Weeks after experiment	8	17	26	26	26
Category-specific time trends	Yes	Yes	Yes	Yes	Yes
Category-specific quadratic	No	No	No	Yes	Yes
Data-entry included	Yes	Yes	Yes	Yes	No

Notes: Panels A, B, and C present the results of regressing the natural logarithm of jobs created in a job category-week (panel A), the natural logarithm of hours worked in a job category-week (panel B), and the natural logarithm of a job's hourly wages (panel C), on "Percent change experience" interacted with a dummy for being after the experiment. All regressions include week fixed effects, job category fixed effects, and job category linear time trends. The last two columns additionally include job category quadratic time trends. These regressions contain outcomes from 26 weeks before the experiment to 8, 17, or 26 weeks afterwards, excluding the weeks of the experiment. In panels A and B, observations are job category-weeks and are weighted by the number of jobs created in the category in a pre-period. In panel C, observations are individual job-weeks and are unweighted. All standard errors are clustered by job category.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

the experiment, omitting the weeks of the experiment and all jobs I offered. The key assumption is that, conditional on job category-specific trends, employment would have changed similarly after the experiment in job categories more and less affected by the experiment. In all three time windows considered, the measured effect of the experiment on new jobs was positive and significant. The elasticity ranged from 0.22 after two months to 0.27 after six months. The effect on hours worked was also positive across all three time periods, with an elasticity ranging from 0.10 to 0.14.

To determine the level increase in hours worked per week from these estimates, I multiply the most conservative hours elasticity from Table 6 (0.099) by the average number of hours worked per week on oDesk and the average value of percent change experience. The result is an increase of 950 hours per week in hours worked. (This excludes any increase in fixed wage employment.) This increase of 950 hours worked per week is smaller than the increase in the number of hours worked by the treatment group (1010), but not much smaller, suggesting that treatment group workers crowded out relatively little other employment.[19] To determine the increase in worker earnings

[19] While the model in Section III predicts that experimental workers should displace non-experimental workers, other models would predict that an increase in the number of oDesk workers recognized to be of adequate ability

due to this increased employment, I multiply this hours increase by the average wage in these new jobs. Online Appendix C details how I calculate this average wage ($2.11, slightly less than treatment group workers' average wage after the experiment: $2.20) and provides additional details on this section. Using these estimates, the increase in worker earnings from the employment expansion was

$$(2) \qquad \underbrace{\left(26 \text{ weeks} \times 950 \frac{\text{hours}}{\text{week}}\right)}_{\text{Increase in employment}} \times \underbrace{\$2.11}_{\text{Wage level}} \approx \$52{,}000.$$

Firms also benefitted from the increased employment. In particular, the theory predicts that firms with fixed costs between $\overline{c}$ and c' hired oDesk workers after the experiment with an average profit of $E[\overline{c}\,' - c_j | \overline{c} < c_j \leq c']$. If c_j is uniformly distributed between $\overline{c}$ and c', this is $\frac{1}{2}(c' - \overline{c})$. I can approximate $c' - \overline{c}$ using the decrease in market-level wages induced by the experiment.[20]

Panel C of Table 6 estimates the experiment's effect on market-level wages by regressing the hourly wage for a given job in a given week on the same independent variables as in panels A and B. The results suggest that the experiment substantially decreased wages. I determine the experiment's effect on the wage level ($0.22) analogously to the level change in hours using the same column's estimate of the wage change (-0.265). If c_j is uniformly distributed, the increase in firm profit from the additional employment was

$$(3) \qquad \underbrace{\left(26\,\text{weeks} \times 950 \frac{\text{hours}}{\text{week}}\right)}_{\text{Increase in employment}} \times \underbrace{\frac{1}{2} \times \$0.22}_{\text{Average profit/job}} \approx \$2{,}800.$$

Table 7 estimates the overall effect of the experiment on market surplus in the six months after the experiment. The middle column presents the most realistic estimates, while the right- and left-hand columns make more and less conservative assumptions, respectively. Panel A presents the benefits of the experiment quantified above. Panel B estimates workers' and firms' opportunity costs of the additional employment after the experiment, while panel C estimates the cost of the experimental jobs themselves.

I assess workers' opportunity costs of working by using their willingness to apply to the jobs I posted three weeks after the experiment. I assume that workers applied to these jobs if and only if their opportunity costs were below the jobs' wages.[21] Because 49 percent of the treatment group applied to the $1 wage job, I use this as

might actually improve other workers' employment outcomes, for example by inducing either new employers to join the marketplace or existing employers to remain in the marketplace.

[20] This is only an approximation, however, because the change in market wages captures both the change in the marginal firm's hiring cost and the change in hired workers' expected abilities. If hired workers' expected abilities increased after the experiment, the market-level wage change is a lower bound on the change in the fixed cost of the marginal firm.

[21] If there was a time cost of applying, this overstates workers' opportunity costs, particularly for workers with low probabilities of getting the jobs.

Table 7—Estimated Effect of the Experiment on Market Welfare

	High benefit, low cost	Medium benefit, medium cost	Low benefit, high cost
Panel A. Increased market surplus			
Increased worker surplus (excluding the cost of effort)	\$52,036	\$52,036	\$52,036
Increased firm profit (from increased employment)	\$2,757	\$2,757	\$2,757
Panel B. Opportunity cost of increased work			
Workers' opportunity cost (alternative hourly wage)	\$12,351 (\$0.50/hour)	\$24,701 (\$1 per hour)	\$36,884 (\$1.49 per hour)
Firms' opportunity cost (fraction of increased profits)	\$689 (25%)	\$1,378 (50%)	\$2,757 (100%)
Panel C. Cost of experimental jobs			
Workers' opportunity cost (alternative hourly wage)	\$3,622 (\$0.50/hour)	\$7,245 (\$1 per hour)	\$10,818 (\$1.49 per hour)
Fixed cost of employing (\$10 per hour spent)	\$476 3 min/worker	\$793 5 min/worker	\$1,587 10 min/worker
Panel D. Overall welfare change			
Total market surplus: cost	\$37,655	\$20,675	\$2,747

the "best-guess" estimate of workers' average opportunity cost. Online Appendix C details how I calculate a more conservative estimate (\$1.49), by assuming that if workers did not apply to this second job, they were not willing to accept any wage below their initial posted wage. I have less information on firms' opportunity costs. However, as I estimate that the increase in firm profits resulting from the experiment was only about 5 percent of the increase in worker earnings, even assuming that firms did not benefit from the increased employment does not affect the welfare conclusions.

Panel C assesses the opportunity cost of the experimental jobs themselves. The opportunity cost of workers' time is calculated as in the previous panel. In a traditional job, these costs would have been offset by the value of the output produced. However, in this experiment, the output had no value. Because workers were not expected to produce usable output, my time cost of employing them was relatively low (five minutes per worker).

Even under the conservative assumptions in the right-most column, the estimates suggest that hiring 952 randomly-selected, relatively inexperienced workers for a meaningless task increased market surplus by over \$2,700 in the subsequent six months. As the benefits of this experiment for treatment group workers and the overall market appear to have increased steadily over time (see Figure 4 and Table 6), extending the time frame over which the benefits were calculated would likely lead to larger estimates of the experiment's effect on market surplus. These calculations suggest that, before the experiment, novice employment on oDesk was inefficiently low.

V. Conclusion

There is a debate in the literature over whether simply helping young and disadvantaged workers enter the labor market can improve their long-term employment outcomes or whether intensive skills training is required. This paper proposes that merely giving workers a first job benefits them by providing the market with information about their abilities, which in turn, makes them more valuable to firms. However, to the extent this information is public, its benefits accrue to workers, so firms may hire too few inexperienced workers. In particular, firms will hire too few inexperienced workers if hiring is costly, they do not receive the benefits of the information produced, and workers cannot fully compensate them for being hired.

This paper uses a field experiment in an online marketplace to test whether giving workers a chance to demonstrate their abilities improves their labor market outcomes. In the experiment, workers were randomly selected to receive a job with a coarse evaluation, a job with a detailed evaluation, or no job. Simply giving workers a job substantially increased their subsequent employment rates, earnings, and reservation wages. Giving the market more detailed information about their job performance also increased their average earnings and reservation wages. These results are consistent with the paper's proposed mechanism: information about their abilities made workers more valuable to employers, but are inconsistent with several alternative mechanisms. Despite the fact that the experiment was not designed with this purpose, under plausible assumptions, it increased market surplus by more than its social cost, suggesting that, before the experiment, oDesk firms hired inefficiently few inexperienced workers.

These results come from a particular marketplace and an important question is whether and how they would generalize to other contexts. The oDesk setting is probably most similar to traditional low-wage labor markets. It is characterized by high unemployment rates and its data-entry workers earn wages just above the allowable minimum. However, oDesk employers may be more uncertain about the abilities of job applicants than offline employers of low-wage workers. They typically have less hiring experience, may be unfamiliar with credentials from foreign schools or employers, and have limited ability to verify these credentials. This suggests that the benefits of performance evaluations may be particularly large in the oDesk marketplace. On the other hand, oDesk has particularly low hiring costs. There is no positive minimum wage on oDesk and firms do not provide employment benefits or supplies. oDesk workers can be hired with the click of a mouse and fired instantaneously with no penalty. The tasks they complete are typically well-defined, easy to explain, and require no training. This suggests that the labor market inefficiency and the benefits of the treatments might be larger in a traditional labor market.

Assuming these results would generalize to traditional low-wage labor markets, there are several public policy responses that might reduce the inefficiency. First, a government could partially or fully subsidize firms for hiring young workers. Second, the government could itself hire young workers. For this to be maximally effective, the government would have to provide the market with honest measures of worker performance and ensure the jobs would not be stigmatized. Avoiding stigmatization might entail hiring some workers already recognized by the market to be high-ability (as I did in this experiment). This would increase the program's costs but also the output of hired workers. Finally, the government or a subsidized firm could provide employment tests, simulated work experiences designed to reliably measure workers' capabilities and diligence. These tests could be designed (as was the task in this experiment), to require little managerial time, but provide a useful signal of workers' performance to potential hiring firms.

As most oDesk jobs are offshored from US employers to foreign workers, the paper's results may shed light on whether developing a reputation is a significant barrier to offshoring, and on a grander scale, trade between foreign and domestic firms. Unlike in other forms of offshoring and international trade, the only significant barrier to transacting on oDesk is the difficulty of building a reputation. Firms offshoring offline may face significant costs of identifying available labor. Similarly, foreign and domestic firms wanting to trade must invest in identifying and communicating with each other as well as, potentially, in new plants and capital. In contrast, oDesk workers and firms can join the marketplace and search for each other costlessly and quickly. This experiment shows that the cost of building a reputation alone is sufficient to reduce the volume of trade, but, when reputations are established, trade volume increases. The extent to which the results of this experiment can be applied to more general trade contexts is an important question for future research.

REFERENCES

Acemoglu, Daron, and Jörn-Steffen Pischke. 1998. "Why Do Firms Train? Theory and Evidence." *Quarterly Journal of Economics* 113 (1): 79–118.

Acemoglu, Daron, and Jörn-Steffen Pischke. 1999. "Beyond Becker: Training in Imperfect Labour Markets." *Economic Journal* 109 (453): 112–42.

Altonji, Joseph G., and Charles R. Pierret. 2001. *Quarterly Journal of Economics* 116 (1): 313–50.

Autor, David H. 2001. "Why Do Temporary Help Firms Provide Free General Skills Training?" *Quarterly Journal of Economics* 116 (4): 1409–48.

Autor, David H., and Susan N. Houseman. 2010. "Do Temporary-Help Jobs Improve Labor Market Outcomes for Low-Skilled Workers? Evidence from 'Work First'." *American Economic Journal: Applied Economics* 2 (3): 96–128.

Becker, Gary S. 1964. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education.* Chicago: University of Chicago Press.

Bell, Stephen H., and Larry L. Orr. 1994. "Is Subsidized Employment Cost Effective for Welfare Recipients? Experimental Evidence from Seven State Demonstrations." *Journal of Human Resources* 29 (1): 42–61.

Bloom, Dan. 2010. *Transitional Jobs: Background, Program Models, and Evaluation Evidence.* New York: MDRC.

Bloom, Dan, Sarah Rich, Cindy Redcross, Erin Jacobs, Jennifer Yahner, and Nancy Pindus. 2009. *Alternative Welfare-to-Work Strategies for the Hard-to-Employ: Testing Transitional Jobs and Pre-Employment Services in Philadelphia.* New York: MDRC.

Bloom, Howard S., Larry L. Orr, Stephen H. Bell, George Cave, Fred Doolittle, Winston Lin, and Johannes M. Bos. 1997. "The Benefits and Costs of JTPA Title II-A Programs: Key Findings from the National Job Training Partnership Act Study." *Journal of Human Resources* 32 (3): 549–76.

Bureau of Labor Statistics. 2014. Household Data Annual Averages. "Table 3: Employment Status of the Civilian Noninstitutional Population by Age, Sex, and Race." US Department of Labor. http://www.bls.gov/cps/cpsa2013.pdf (accessed September 15, 2014).

Couch, Kenneth A. 1992. "New Evidence on the Long-Term Effects of Employment Training Programs." *Journal of Labor Economics* 10 (4): 380–88.

Dickens, William T., Lawrence F. Katz, Kevin Lang, and Lawrence H. Summers. 1989. "Employee Crime and the Monitoring Puzzle." *Journal of Labor Economics* 7 (3): 331–47.

Farber, Henry S., and Robert Gibbons. 1996. "Learning and Wage Dynamics." *Quarterly Journal of Economics* 111 (4): 1007–47.

Gibbons, Robert, and Lawrence F. Katz. 1991. "Layoffs and Lemons." *Journal of Labor Economics* 9 (4): 351–80.

Greenwald, Bruce C. 1986. "Adverse Selection in the Labour Market." *Review of Economic Studies* 53 (3): 325–47.

Hollister, Robinson G., Jr., Peter Kemper, and Rebecca A. Maynard. 1984. *The National Supported Work Demonstration.* Madison: University of Wisconsin Press.

Kahn, Lisa B. 2013. "Asymmetric Information between Employers." *American Economic Journal: Applied Economics* 5 (4): 165–205.

Loewenstein, Mark A. and James R. Spletzer. 1998. "Dividing the Costs and Returns to General Training." *Journal of Labor Economics* 16 (1): 142–71.

oDesk Corporation. 2012. *Online Material*. https://www.odesk.com/oconomy/activity (accessed February 13, 2012).

Pallais, Amanda. 2014. "Inefficient Hiring in Entry-Level Labor Markets: Dataset." *American Economic Review.* http://dx.doi.org/10.1257/aer.104.11.3565.

Redcross, Cindy, Dan Bloom, Gilda Azurdia, Janine Zweig, and Nancy Pindus. 2009. *Transitional Jobs for Ex-Prisoners: Implementation, Two-Year Impacts and Cost of the Center for Employment Opportunities (CEO) Prisoner Reentry Program.* New York: MDRC.

Schönberg, Uta. 2007. "Testing for Asymmetric Employer Learning." *Journal of Labor Economics* 25 (4): 651–91.

Stanley, Marcus, Lawrence F. Katz, and Alan Krueger. 1998. *Developing Skills: What We Know About the Impact of American Employment and Training Programs on Employment Earnings, and Educational Outcomes.* Report for G8 Economic Summit.

Stanton, Christopher, and Catherine Thomas. 2013. "Landing the First Job: The Value of Intermediaries in Online Hiring." Unpublished.

Terviö, Marko. 2009. "Superstars and Mediocrities: Market Failure in the Discovery of Talent." *Review of Economic Studies* 76 (2): 829–50.

Waldman, Michael. 1984. "Job Assignments, Signalling, and Efficiency." *RAND Journal of Economics* 15 (2): 255–67.

Weiss, Andrew Murray. 1980. "Job Queues and Layoffs in Labor Markets with Flexible Wages." *Journal of Political Economy* 88 (3): 526–38.