Essays in Labor Economics

by

Amanda Dawn Pallais

B.A. Economics and Mathematics, University of Virginia (2006)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2011

Signature of Author ....
Department of Economics
May 15, 2011

Certified by ....
David Autor
Professor of Economics
Thesis Supervisor

Certified by ....
Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics
Thesis Supervisor

Accepted by ....
Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics
Chairman, Departmental Committee on Graduate Studies

Essays in Labor Economics

by

Amanda Dawn Pallais

Submitted to the Department of Economics
on May 15, 2011, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This dissertation consists of three chapters on topics in labor economics. In the first chapter, I present a model in which firms under-invest in hiring novice workers because they don't receive the full benefit of discovering novice talent. A firm must pay a cost to hire a novice worker. When it does, it obtains both labor services and information about the worker's productivity. This information has option value as a productive novice can be rehired. However, if competing firms also observe the novice's productivity, the option value of hiring accrues to the worker, not the employer. Firms will accordingly under-invest in discovering novice talent unless they can claim the benefit from doing so. I test this model's relevance in an online labor market by hiring 952 workers at random from an applicant pool of 3,767 for a 10-hour data entry job. In this market, worker performance is publicly observable. Consistent with the model's prediction, novice workers hired at random obtained significantly more employment and had higher earnings than the control group, following the initial hiring spell. A second treatment confirms that this causal effect is likely explained by information revelation rather than skills acquisition. Providing the market with more detailed information about the performance of a subset of the randomly-hired workers raised earnings of high-productivity workers and decreased earnings of low-productivity workers. Due to its scale, the experiment significantly increased the supply of workers recognized as high-ability in the market. This outward supply shift raised subsequent total employment and decreased average wages in occupations affected by the experiment (relative to non-treated occupations), implying that it also increased the sum of worker and employer surplus. Under plausible assumptions, this additional total surplus exceeds the social cost of the experiment.

In the second chapter, I estimate the sensitivity of students' college application decisions to a small change in the cost of sending standardized test scores to colleges. In 1997, the ACT increased the number of free score reports it provided to students from three to four, maintaining a \$6 marginal cost for each additional report. In response to this \$6 cost change, ACT-takers sent more score reports and applications, while SAT-takers did not. ACT-takers also widened the range of colleges to which they sent scores. I show that students' response to the cost change is inconsistent with optimal decision-making but instead suggests that students use rules of thumb to make college application decisions. Sending additional score reports could, based on my estimates, substantially increase low-income students' future earnings.

In the third chapter, I analyze the effects of the Tennessee Education Lottery Scholarships, a broad-based merit scholarship program that rewards students for their high school achievement with college financial aid. Since 1991, over a dozen states, comprising approximately a quarter of the nation's high school seniors, have implemented similar merit scholarship programs. Using individual-level data from the ACT exams, I find that the program did not achieve one of its stated goals, inducing more students to prefer to stay in Tennessee for college, but it did induce large increases in performance on the ACT. This suggests that policies that reward students for performance affect behavior and may be an effective way to improve high school achievement.

Thesis Supervisor: David Autor
Title: Professor of Economics

Thesis Supervisor: Esther Duflo
Title: Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics

Contents

Acknowledgements

I have been extremely fortunate while at MIT for the opportunity to work with my advisors David Autor, Esther Duflo, and Daron Acemoglu. Their deep insights and helpful suggestions have not only substantially improved this dissertation, but also my skills as an economist. Their unfailing willingness to review and challenge my work, assumptions, and conclusions as well as their intellectual and financial support has been left a deep imprint on this work and on me. I hope to pass on to others what they have given me.

I have also greatly benefitted from many discussions with professors and fellow graduate students. In particular, Jason Abaluck, Leila Agha, Josh Angrist, Sue Dynarski, Amy Finkelstein, Maria Fitzpatrick, Bob Gibbons, Jonathan Goldberg, Michael Greenstone, Jon Gruber, Jerry Hausman, Bill Johnson, Lisa Kahn, Daniel Keniston, Danielle Li, Alan Manning, Whitney Newey, Christopher Palmer, Iuliana Pascu, Jim Poterba, David Powell, Michael Powell, Jesse Rothstein, Simone Schaner, Joe Shapiro, Pian Shu, Chris Smith, Jean Tirole, Juuso Toikka, Sarah Turner, and Heidi Williams deserve special thanks. I would like to thank participants at the MIT labor and applied micro lunches, the MIT Labor/PF seminar, and the NBER Higher Education Working group meeting for their helpful comments.

I am grateful to Anand Hattiangi, Dmitry Diskin, and the oDesk Corporation for help understanding and accessing the oDesk data; to Jesse Rothstein, Princeton University, James Maxey, Julie Noble, and the ACT Corporation for allowing me access to the ACT database; Caroline Hoxby for help in accessing the American College Survey Database; and Robert Anderson, Erin O'Hara, David Wright and the Tennessee Higher Education Commission for providing summary statistics about the Tennessee Education Lottery Scholarship winners. Financial support from the National Science Foundation, the George and Obie Shultz Fund, David Autor, and Esther Duflo is gratefully acknowledged.

Finally, I would like to thank my parents, Don and Dee Ann Pallais, for their incredible support over the past five years. I cannot imagine any two people being any more supportive or encouraging over such a long period of time. This dissertation is dedicated to them.

Chapter 1

Inefficient Hiring in Entry-Level Labor Markets

1.1 Introduction

Young workers are more likely to be unemployed than workers who have had time to accumulate labor market experience. In August, 2010, for example, 14.7% of workers 20 to 24 years old were unemployed, compared with 9.5% of the general population.[1] Competition for entry-level jobs is intense. Newman (1996) found that the average success rate of applications to fast food restaurants, a common job for entry-level workers, in Harlem, NY was only 7%.

There are many potential explanations for the high unemployment rates of young workers, including that they are simply less skilled than more experienced workers. This paper proposes the hypothesis that firms hire inefficiently few inexperienced workers because they do not receive the full benefit of discovering workers' talent. It tests this hypothesis in a field experiment involving over 3,700 workers and finds strong support for it.

Employers are uncertain about the abilities of workers who lack experience. Hiring these workers generates information about their abilities (e.g., Farber and Gibbons, 1996; Altonji and Pierret, 2001), producing an option to hire the high-ability workers in the future. Generating this option is costly for firms. Managers must spend time explaining the jobs to workers and monitoring their progress. Moreover, firms incur an opportunity cost of lost time if jobs are not completed correctly or timely. If information about worker quality is partially public, high-ability workers receive

[1] This is from the Bureau of Labor Statistics's Table A-13, September 2010.

part of the option value as higher wages. In the extreme case of pure public learning, workers receive the entire option value. If workers cannot compensate firms for generating this option, for example because they cannot accept sub-minimum wages or credibly commit to low-wage long-term contracts, firms will hire inefficiently few inexperienced workers.[2]

Tervio (2009) proposed that this type of informational inefficiency could generate excessively high wages for CEO's and entertainers. This paper shows that it may affect entry-level labor markets. There is substantial uncertainty about the abilities of entry-level workers, particularly those with little education and few credentials. Firms cannot conceal whether they have fired, retained, or promoted a worker, an important signal of entry-level worker performance. Entry-level labor markets have developed several institutions that reduce this inefficiency. Internships and hiring subsidies for young workers reduce firms' costs of hiring inexperienced workers. Fixed-term contracts (in Europe) reduce firms' costs of hiring inexperienced workers by allowing firms to dismiss low-ability young workers more easily. In some European training contracts, unions and industry consortiums pay workers' initial salaries. These allow the parties who benefit from talent discovery to pay for it.

Discovering a worker's ability is similar to general skills training: both produce a form of human capital that raises workers' value to firms, but require up-front investments. A large literature examines firms' provision of general skills training. Similar to this paper's model, 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 theoretical work identifies circumstances in which firms receive part of the return on their training investments. 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., Autor, 2001; Loewenstein and Spletzer, 1998). However, neither the theoretical nor the empirical literature has shown that firms recoup the full value of their training investments and, thus, will provide the optimal level of training. This paper provides evidence that another form of human capital, information about worker ability, is underprovided by firms.

[2]These are not the only reasons why workers may not be able to compensate firms for creating the option. This ineffiency occurs at above-minimum wages if a wage rigidity, such as potential adverse selection of workers, prevents wages from falling.

If not provided by firms, general human capital can be provided by schools; but output and information about workers' abilities may be 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 are few alternative mechanisms for its production.

This paper first develops a stylized model that demonstrates this inefficiency and generates testable predictions. In the model, firms have to pay a non-wage cost to hire novice workers. Hiring these workers generates information about their productivities, generating an option to hire the high-ability workers in the subsequent period. However, the market observes this information so the option value accrues to high-ability workers as higher wages. Because workers cannot provide their labor services below a certain wage (e.g., because of adverse selection or an inability to take sub-minimum wages), firms underinvest in discovering novice talent. The model predicts that providing inexperienced workers with an opportunity to demonstrate their ability through employment would allow some of them to be recognized as high-ability. This would increase inexperienced workers' future employment, earnings, and reservation wages. Moreover, by increasing the supply of workers recognized as high ability, it would increase future market employment, decrease market wages, and increase total market surplus: the sum of worker surplus and firm profits. The model also predicts that providing the market with more information about the job performance of inexperienced workers would increase the employment, earnings, and reservation wages of workers who performed well and impair the analogous employment outcomes of workers who performed poorly. Overall, it would increase the option value of hiring inexperienced workers, and thereby, increase their future earnings.

The paper then tests these predictions through a field experiment in a large, online marketplace: oDesk. oDesk workers are located all over the world and complete approximately 200,000 hours of work per week remotely. Many oDesk workers work full-time in the marketplace, though many others apply for jobs but never get work. Employers are required to provide public evaluations of all of their workers. They must rate workers on a one-to-five scale and may, at their option, leave a short comment. Workers cannot remove the numerical rating.[3]

For the experiment, I posted 10-hour data entry jobs to the marketplace and invited low-wage data-entry specialists to apply. Out of an applicant pool of 3,767 workers, I hired 952 workers at random. Like all employers in the marketplace, after the job was complete, I rated the performance of each worker on a one-to-five scale. I gave most of these workers an uninformative comment,

[3]While they can remove the comment, only four percent of them do.

but I gave a randomly-selected subset a detailed comment with objective information about their data entry speed, accuracy, following of directions, and timely completion of the task. Using the marketplace's administrative data, I then observe all subsequent employment outcomes of the experimental workers on oDesk, including the jobs they obtain, the hours they work, and their total earnings. Three weeks later, I assess their reservation wages by inviting a randomly-selected subset of these workers to apply to a new job offered by a different employer with a randomly-selected wage.

The data support the model's predictions. In the two months after the experiment, workers randomly selected to receive treatment jobs were more likely to be employed, requested higher wages, and had higher earnings than the control group. Although the experiment did not target workers who were particularly likely to benefit from the treatment, it had very large effects. Excluding the experimental jobs, within two months of the treatment, the average treatment group worker had earned in excess of $20 more than the average control group worker. I spent less than $17 to hire each treatment group worker. Workers who received positive detailed comments earned substantially more after the experiment than workers who received uninformative comments. Workers who received negative detailed comments earned even less. Overall, workers who received detailed evaluations earned in excess of $23 more than those who received less informative comments.

The large size of the experiment relative to the marketplace and the wide variation in its impact on the marketplace's 74 job categories allow me to evaluate the experiment's effects on the marketplace as a whole. Using a difference-in-difference strategy, I find that, after the experiment, employment increased and wages decreased in job categories more heavily affected by the experiment. I use these estimates to estimate the experiment's effects on total market surplus and find that, under plausible assumptions, the benefits of the experiment to firms and workers outweighed its social cost. This result obtains despite the fact that this experiment was not targeted to increase surplus, output created during the experimental treatment is not counted in the total, and I only consider benefits within eight weeks of the experiment. This result suggests that inefficiently low hiring of novice workers in this market led to diminished output and employment.

The interpretation of these results depends on the mechanism that generates them. The paper considers whether three alternative mechanisms could generate these outcomes: (1) the experimental jobs provided workers with human capital, (2) hiring the workers in itself gave the market a positive signal about their abilities, and (3) job receipt induced workers to apply to more oDesk jobs, but did not change employers' beliefs about worker ability. These mechanisms cannot explain all of the experiment's results. None of them can explain the effects of giving workers a more detailed evaluation. Moreover, the treatment is unlikely to have substantially increased workers' human capital because, on average, it only lasted 7.6 hours and represented a small increase in workers' total offline work experience. Being hired in itself did not improve workers' outcomes. Job receipt only affected employment outcomes after the market observed the workers' evaluations, not when it observed only that they had been hired. The evaluations did change the market's assessment of these workers' value. After the experiment, treatment group workers were more likely to receive any given job they applied to than were control group workers.

The experimental results inform the literature on the effect of reputation on contracting. The theoretical literature shows that when legal or feasibility constraints prevent writing the optimal contract, agents' reputations affect the set of feasible outcomes (e.g., Kreps and Wilson, 1982; Tirole, 1996). Empirical research also shows reputations are important determinants of contracts (e.g., Banerjee and Duflo, 2000; McMillan and Woodruff, 1999; Resnick, *et al.* 2006). However, developing a reputation is a cost of market entry. An important question is to what extent this cost restricts economic activity. This is a particularly large concern in the trade of goods and labor with foreign countries where contracts are difficult to enforce and domestic firms have little familiarity with trading partners. In the oDesk marketplace, reputations are important because workers have limited liability constraints and cannot pledge their future output. The paper shows that developing a reputation is a significant barrier to entering this labor market, which reduces total market employment and output.[4]

The rest of the paper is organized as follows. Section 1.2 describes the online marketplace, Section 1.3 presents the theoretical framework, and Section 1.4 lays out the experimental design. Section 1.5 analyzes the worker-level effects of the experiment and analyzes whether they could be generated by alternative mechanisms. Section 1.6 presents the experiment's market-level effects and Section 1.7 concludes.

1.2 The Marketplace

oDesk is an online marketplace in which employers hire independent contractors to perform tasks remotely. The marketplace is large: as of July, 2010, oDesk workers completed approximately

[4]As the majority of oDesk jobs are offshored, these results speak directly to the feasibility of offshoring. Reputation concerns do not limit all trade, but they prevent offshoring from reaching the efficient level.

200,000 hours of work per week, the equivalent of 5,000 full-time employees. oDesk workers are located around the world. A plurality (37%) lives in the United States, while India (15%) and the Philippines (14%) are the next most common countries of residence. Employers are also located around the world, but approximately 80% are located in the United States.

Employers post job openings in a wide variety of fields, the majority being in web programming, website design, and data entry. Some jobs last only a few hours, while others constitute full-time employment. The average oDesk job lasts 69 hours.

When posting a job, employers choose one of two pay schemes: "hourly" and "fixed wage." Hourly jobs, the type of jobs created in this experiment, constitute 70% of jobs on oDesk. In this type of employment, workers earn a fixed hourly wage and oDesk tracks the number of hours worked. oDesk guarantees that the employer will pay for the hours worked regardless of the quality of the final product, though the employer can stop the job at any time or limit the number of hours worked. In fixed-wage jobs, workers and employers agree to a price for the entire project, but at the end of the project, employers have complete discretion over how much they pay. In these jobs, hours worked are not recorded.

Once a job is posted, workers can apply to the job directly or employers can invite them to apply. Under either method, the worker proposes 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 may hire as many or few applicants as they deem suitable.

Workers post public profiles, describing their skills and the types of jobs they are seeking. To verify these skills, workers can take over 200 multiple-choice skills tests on topics such as Microsoft Word or C++ proficiency and post the results to their profiles. Each worker posts her preferred hourly wage rate at the top of her profile. Workers can suggest a different wage to employers when applying for jobs, but employers see their posted wages as well. An example worker profile is displayed in Figure 1.[5]

As soon as a hired worker begins working in an hourly job, the job title, number of hours worked, and hourly wage rate are automatically posted to the worker's profile.[6] When a job is complete, employers and workers must rate each other from one to five on six dimensions: availability, communication, cooperation, deadlines, quality, and skills. These scores are averaged to form each party's overall rating for the job. Both composite ratings are automatically posted

[5] This worker was not included in the experiment because he did not join oDesk until after the experiment.

[6] In fixed-wage jobs, only the job title and agreed job price are posted.

to the worker's profile.[7] A worker cannot remove the ratings without refunding the remuneration received. Employers' ratings are typically very positive: 64% of workers receive a rating of exactly five, while 83% average at least four.

Worker and Employers may also choose to provide short comments about the employment experience. These are also automatically posted to the worker's profile. Comments are generally one to two positive sentences providing little objective information. Unlike the numerical ratings, workers can remove employer comments without penalty, but only 4% of them do.

1.3 Model

This section provides a simple framework that formalizes the insight that inefficiently few inexperienced workers will be hired when firms do not receive the benefit from discovering talented novices. It also develops implications that are testable in the oDesk setting.

1.3.1 Model Set-Up

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

Workers vary in their ability (a_i), which is unknown to all market participants. Before the worker's novice period, all market participants observe a normally distributed signal of the worker's ability from the characteristics, x_i, listed on her resume. They use Bayesian updating to form their beliefs about worker ability:

$$a_i \sim N\left(\hat{a}_{i0}, \frac{1}{h_{x_i}}\right) \tag{1.1}$$

where $\hat{a}_{i0}$ denotes a worker's expected ability before her novice period. This expected ability is also distributed normally across workers.

Each firm (indexed by j) offers an identical task in which workers' output is $y_i = a_i$. It 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 and monitor her progress as well as any related overhead costs, such as for equipment or office space. It is continuously distributed across firms with cumulative distribution function F_c.

[7] According to oDesk's website, oDesk posts the worker's evaluation of the firm beside the firm's evaluation of the worker in order to allow "both sides of the story" to be presented.

A worker's net marginal product in the task is

$$a_i - c_j. \tag{1.2}$$

The output of employed workers is imperfectly observable. All market participants observe a signal of output, $\hat{y}_i$:

$$\hat{y}_i = a_i + \varepsilon_{iy} \text{ where } \varepsilon_{iy} \sim N\left(0, \frac{1}{h_y}\right). \tag{1.3}$$

Firms update their beliefs about the abilities of novice workers using this signal and Bayesian updating. The term $\hat{a}_i$ designates any worker's expected ability, while $\hat{a}_{i1}$ designates a worker's expected abilities before her veteran period. The term F_0 is defined as the exogenous cumulative distribution function of novice workers' expected abilities and F_1 as the endogenous cumulative distribution function of veteran workers' expected abilities.

The market clears each period. Wages, w_{ij}, are restricted to be non-negative. For simplicity, workers and firms are risk neutral and do not discount the future. If firms do not fill their jobs, they receive 0, while if workers do not receive a job, they earn their outside option: $w_0 = 0$.

1.3.2 Market Equilibrium

Proposition 1 *The Perfect Bayesian Equilibrium of this game involves a threshold, $\bar{c}$, such that all firms with fixed costs $c_j \leq \bar{c}$ will hire a worker, while no firm with $c_j > \bar{c}$ will. All workers with expected ability $\hat{a}_i \geq \bar{c}$ and only these workers will be employed. These workers will earn wages $w_{ij} = \hat{a}_i - \bar{c}$.*

Risk-neutral firm j earns expected profit

$$\hat{\pi}_{ij} = \hat{a}_i - c_j - w_{ij} \tag{1.4}$$

from hiring worker i. Because novice expected ability is normally distributed, there exists a mass of novice workers whose expected ability is below every firm's fixed cost. No firm will incur an expected loss by hiring them at non-negative wages. Thus, not all workers will be hired and not all firms can hire a worker.

Firms vary only by their fixed costs. Therefore, there exists a fixed cost threshold, $\bar{c}$, such that all firms with $c_j \leq \bar{c}$ will hire a worker.[8] All workers weakly prefer a market job to their

[8] Throughout the paper, there are measure zero sets of workers and firms that are indifferent between entering

outside option. To the market, they vary only by their expected abilities. Thus, there is an ability threshold, $\underline{a}$, such that all workers with $\hat{a}_i \geq \underline{a}$ will be hired.

Firms must all be indifferent to hiring any employed worker at her market wage, so all employed workers must be offered a wage

$$w_{ij} = \hat{a}_i - \bar{w} \tag{1.5}$$

for some constant $\bar{w}$. The marginal firm must earn zero profit from hiring this worker so $\bar{w} = \bar{c}$ and wages equal

$$w_{ij} = \hat{a}_i - \bar{c}. \tag{1.6}$$

The threshold worker must earn a zero wage; otherwise firms could earn a higher profit by offering slightly lower-ability workers lower wages. This implies $\underline{a} = \bar{c}$. These thresholds are determined by equating the number of hired workers with the number of filled jobs:

$$\underbrace{\frac{1}{2}(1 - F_0(\bar{c})) + \frac{1}{2}(1 - F_1(\bar{c}))}_{\text{Supply of Labor}} = \underbrace{F_c(\bar{c})}_{\text{Demand for Labor}} \tag{1.7}$$

where F_1 is the cumulative distribution function of veteran workers when all novice workers with $\hat{a}_{i0} \geq \bar{c}$ are hired. The left-hand side of this equation is the mass of workers hired: the mass of novice and veteran workers with $\hat{a}_i \geq \bar{c}$. The right-hand side is the mass of firms hiring: the mass of firms with $c_j \leq \bar{c}$.

1.3.3 The Social Planner's Solution

Proposition 2 *The desired equilibrium of a social planner who has the same information as the market and maximizes market surplus is as follows. There exists a threshold, $c^* > \bar{c}$, 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 $(\hat{a}_{i0} - c^*) + \Pr(\hat{a}_{i1} \geq c^*) \times E[\hat{a}_{i1} - c^* | \hat{a}_{i1} \geq c^*] \geq 0$. More novice workers are employed than in the market equilibrium.*

The social planner will have only the most efficient firms hire workers. Thus, there is a threshold, c^*, such that every firm with $c_j \leq c^*$ will hire a worker and no firm with $c_j > c^*$ will. Veteran workers are hired only when hiring them weakly increases expected worker surplus in the current

the market and taking their outside options (e.g., firms with $c_j = \bar{c}$). Throughout the paper, I use weak inequalities indicating that these agents enter the market. However, no other features of the equilibria would change if these measure zero sets of workers and firms did not enter.

period. That is, when

$$\hat{a}_{i1} \geq c^*. \tag{1.8}$$

In the social planner's solution, novice workers will be hired if the expected benefit from hiring them, calculated over both periods of life, is non-negative. All novices expected to generate positive surplus in the current period ($\hat{a}_{i0} \geq c^*$) will be hired. Workers expected to generate a loss in their novice period ($\hat{a}_{i0} < c^*$), but whose veteran-period option value is greater than this expected loss will also be hired. Novice workers will be hired if

$$\underbrace{(\hat{a}_{i0} - c^*)}_{\text{Novice-Period Benefit}} \quad + \underbrace{\Pr(\hat{a}_{i1} \geq c^*) \times E[\hat{a}_{i1} - c^* | \hat{a}_{i1} \geq c^*]}_{\text{Veteran-Period Option Value}} \geq 0. \tag{1.9}$$

The first term in this equation is the expected current-period benefit of hiring a novice worker. The second term is the option value of being hired for workers with $\hat{a}_{i0} < c^*$. (This equation is automatically satisfied for workers with $\hat{a}_{i0} \geq c^*$.) If not employed in their novice periods, these workers will not be employed in their veteran periods. However, if employed in their novice periods, these workers will be hired with probability $\Pr(\hat{a}_{i1} \geq c^*)$ in their veteran periods and generate expected surplus $E[\hat{a}_{i1} - c^* | \hat{a}_{i1} \geq c^*]$ if hired. In contrast, in the market equilibrium, because firms do not value workers' veteran-period option value, novice workers are hired only if their novice-period benefit is non-negative.

The social planner's cutoff, c^*, is determined by the market clearing condition. Define N_{c^*} as the fraction of novices for whom Equation (1.9) holds and F_1^* as the cumulative distribution of veteran ability if novices for whom Equation (1.9) is satisfied are hired. Then, c^* is chosen so that

$$\underbrace{\frac{1}{2} N_{c^*} + \frac{1}{2}(1 - F_1^*(c^*))}_{\text{Supply of Labor}} = \underbrace{F_c(c^*)}_{\text{Demand for Labor}}. \tag{1.10}$$

For any threshold, c, more novices and veterans are hired in the social planner's solution than in the market equilibrium, so there would be an excess supply of labor in the social planner's solution if $c^* = \bar{c}$. The social planner would want to employ novices with $\hat{a}_{i0} < \bar{c}$, which would lead to a larger mass of veterans with $\hat{a}_{i1} \geq \bar{c}$. To equilibrate supply and demand, more firms must hire workers in the social planner's solution than the market equilibrium: $c^* > \bar{c}$. With more firms hiring workers and a higher veteran hiring threshold than in the market equilibrium, the social planner's solution must involve hiring more novice workers than the market equilibrium. Otherwise, there would be excess demand for labor.

The social planner's solution would be enacted as the competitive equilibrium if workers could accept negative wages in their novice period. In this case, novices employed in the social planner's solution, but not by the constrained market, would receive wages of

$$w_{ij} = \hat{a}_{i0} - c^* < 0. \tag{1.11}$$

1.3.4 Adverse Selection

In the above model, wages for the marginal worker must equal zero or (equivalently) a binding minimum wage, for the inefficiency to be present. This is a function of the simplified set-up, not a true requirement for the model. This inefficiency will be present at higher wages if a wage rigidity prevents wages from dropping below a certain level. I consider a simple extension in which potential adverse selection of workers keeps wages artificially high.

While adverse selection is a more general problem, I illustrate its effects with a simple extension of the model with three groups of workers: A, B, and C, which comprise fractions γ_A, γ_B, and γ_C of the population, respectively. A worker's ability, a_i, and outside option, w_0, depend on to which of three groups a worker belongs. Group A and B workers have the same (non-zero) distribution of ability and outside options $w_0 = w_A$ and w_B, respectively that satisfy

$$w_A > w_B > 0. \tag{1.12}$$

Group C workers all have ability $a_i = 0$ and an outside option $w_0 = 0$. These groups represent the fact that workers with higher productivity in a given market will, in general, have better outside options. However, outside options vary among workers with the same market productivity. For simplicity, in this section, I assume that employers face equal uncertainty about each novice worker: $h_{x_i} \equiv h_x$ for all i. The rest of the model is unchanged.

Before workers' novice period, firms observe characteristics x_i from each worker's resume. Firms do not know workers' group affiliations and these characteristics provide no information about affiliations. Workers and firms have the same information except that workers know their group affiliations because they know their outside options. Firms learn workers group affiliations when workers are employed.

Define $\hat{a}_i$ as firms' expectation of a worker's ability if she were known to be in group A or

B. Define $\hat{\eta}_i$ as firms' expectation of the ability of workers who will accept a given wage offer, conditional on $\hat{a}_i$ and their beliefs about workers' strategies. For veteran group A or B workers who worked in their novice period, $\hat{a}_{i1} = \hat{\eta}_{i1}$.

I make two assumptions about the model's parameters. The first guarantees that group $A's$ outside option is sufficiently high that novice group A workers accept their outside options over some positive market wages. If group A workers never accept their outside option over positive wage offers, then there is no adverse selection problem. The second guarantees that labor demand is sufficiently low that at the ability levels at which group A workers select out of the market, firms will not hire only group B and C workers. Without this assumption, adverse selection would not lead to reduced market employment. Before specifying these assumptions mathematically, I define two functions, $c(a)$ and $G(a)$.

The function $c(a)$ gives the fixed cost of the marginal hiring firm if all novices with $\hat{a}_{i0} \geq a$, all veteran group A workers with $\hat{a}_{i1} - c \geq w_A$, and all veteran group B workers with $\hat{a}_{i1} - c \geq w_B$ are hired. It is defined implicitly by

$$\underbrace{\frac{1}{2}[1 - F_0(a)]}_{\text{Novice Workers Hired}} + \underbrace{\frac{1}{2}\gamma_A[1 - F_{1a}(w_A + c(a))]}_{\text{Veteran Group A Workers Hired}} + \underbrace{\frac{1}{2}\gamma_B[1 - F_{1a}(w_B + c(a))]}_{\text{Veteran Group } B \text{ Workers Hired}} = \underbrace{F_c(c(a))}_{\text{Hiring Firms}}. \quad (1.13)$$

where F_0 is the cumulative distribution function of $\hat{a}_{i0}$ for group A workers and F_{1a} is the cumulative distribution function of $\hat{a}_{i1}$ for group A veterans if all group A novices with $\hat{a}_{i0} \geq a$ are hired. The left-hand side of the equation is the mass of hired workers and the right-hand side is the mass of hiring firms.

The function $G(a)$ is the expected net surplus a group A novice worker with expected ability $\hat{a}_{i0} = a$ receives from working when the workers listed above are also employed. That is

$$G(a) = \underbrace{((\gamma_A + \gamma_B)a - c(a) - w_A)}_{\text{Net Wage in Novice Period}} + \underbrace{\Pr(\hat{a}_{i1} - c(a) \geq w_A) \times E[\hat{a}_{i1} - c(a) - w_A | \hat{a}_{i1} - c(a) \geq w_A]}_{\text{Veteran-Period Option Value}}.$$

The worker's novice-period wages equal her expected net marginal product at the marginal firm. Her expected ability is $(\gamma_A + \gamma_B)a$ because all three groups of novices with this expected ability work in the market. In her veteran period, her group affiliation is known and she will work in the market when her market wages exceed her outside option. Finally, I define $\bar{a}_A$ as the expected ability level at which a novice group A worker is indifferent between working in the market and accepting her outside option:

$$G(\bar{a}_A) \equiv 0. \tag{1.14}$$

Then, the two assumptions are:

$$(\gamma_A + \gamma_B)\bar{a}_A - c(\bar{a}_A) \;>\; 0 \tag{1.15}$$

$$(\frac{\gamma_B}{\gamma_B + \gamma_C})\bar{a}_A - c(\bar{a}_A) \;<\; 0. \tag{1.16}$$

The left-hand side of Equation (1.15) is the market wage received by the indifferent novice group A worker. I assume this wage is positive. The left-hand side of Equation (1.16) is the expected marginal product of group B and C workers with $\hat{a}_{i0} = \bar{a}_A$ at the marginal hiring firm. I assume it is negative. This ensures firms will not hire workers with $\hat{a}_{i0} < \bar{a}_A$: at these expected ability levels they would only hire group B and C workers who would not generate positive net output.

Proposition 3 *If Equations (1.15) and (1.16) are satisfied, then the Perfect Bayesian Equilibrium of this game is as follows. There exists a threshold, $\bar{c}$, such that all firms with $c_j \leq \bar{c}$, and only these firms hire a worker. There exists a threshold, $\bar{a}_A$, such that all novice workers with $\hat{a}_{i0} \geq \bar{a}_A$ are hired at a wage of $(\gamma_A + \gamma_B)\bar{a}_A - \bar{c} > 0$ and workers with $\hat{a}_{i0} < \bar{a}_A$ remain unemployed. At this wage, group A novice workers with expected ability $\hat{a}_{i0} = \bar{a}_A$ are indifferent to taking a job, but group B and C workers with $\hat{a}_{i0} = \bar{a}_A$ strictly prefer a market job to their outside options. Group B and C workers with $\hat{a}_{i0} < \bar{a}_A$ would be willing to work for lower wages, but firms will not hire them. Total social surplus is maximized when some novice workers with $\hat{a}_{i0} < \bar{a}_A$ are also hired.*

Proof. *See Appendix A.* ∎

1.3.5 Testable Predictions

The model makes predictions about the effects of the experiment on both treated workers and the market. In the first treatment, novice workers are hired at random from the applicant pool. I then compare the employment outcomes of the treatment and control groups in their veteran periods. The model makes the following testable predictions about the employment outcomes of veteran treatment group workers relative to veteran control group workers.

1. Veteran treatment group workers will have higher employment rates.

2. Veteran treatment group workers will have higher earnings.

3. Veteran treatment group workers will have higher reservation wages.

4. The treatment will have larger effects on the employment, earnings, and reservation wages of workers about whom the market is more uncertain, conditional on ability. Unconditional on ability, the model makes no predictions about the relative sizes of these treatment effects.

Treatment group workers have a range of values of $\hat{a}_{i0}$. Figure 2 is a schematic representation of the distribution of novice workers' expected abilities. (For clarity, it assumes that the market faces the same uncertainty about the abilities of all novice workers, so that the option value of being hired and the social planner's novice hiring threshold depend only on workers' expected abilities.) Workers in the unshaded region would be hired by both the market and the social planner, workers in the light gray shaded region would be hired only by the social planner, and workers in the dark gray shaded region would be hired by neither the market nor the social planner. Treatment group workers come from all three sections of the graph.

Control group workers in the unshaded region (with $\hat{a}_{i0} \geq \bar{c}$) will be hired by other employers in their novice periods, so there is no difference between the veteran-period outcomes of treatment and control group workers in this region. Their employment, earnings, and reservation wages will be the same, on average. However, no control group worker in either of the gray shaded regions (with $\hat{a}_{i0} < \bar{c}$) will be hired in her novice period. Thus, their expected abilities will remain below the hiring threshold and they will have no earnings or employment in their veteran period. On the other hand, some treatment group workers with $\hat{a}_{i0} < \bar{c}$ will perform well enough in their treatment jobs that their veteran-period expected abilities will exceed the hiring threshold and they will have positive veteran-period employment and earnings. Thus, as a group, veteran treatment group workers will have higher employment and earnings than veteran control group workers.

The reservation wages of veteran workers with expected ability below the hiring threshold ($\hat{a}_{i1} < \bar{c}$) is zero. These workers will not receive any positive wage offers, so they will be willing to accept any offer at least as high as their outside option. Veteran workers with expected ability above the hiring threshold have reservation wages equal to the positive wages they receive in the market ($\hat{a}_{i1} - \bar{c} > 0$). Receiving a treatment job affects the reservation wages of only workers in the gray shaded areas. By increasing some of these workers' expected abilities above the threshold, it increases their reservation wages above zero.

Novice workers in the gray shaded areas realize an increase in employment, earnings, and reservation wages of

$$\underbrace{\Pr(\hat{a}_{i1} \geq c^*)}_{\text{Increase in Employment}} \tag{1.17}$$

$$\underbrace{\Pr(\hat{a}_{i1} \geq c^*) \times E[\hat{a}_{i1} - c^* | \hat{a}_{i1} \geq c^*]}_{\text{Increase in Earnings and Reservation Wages}}. \tag{1.18}$$

Conditional on the novice worker's expected ability, $\hat{a}_{i0}$, both expressions are increasing in the variance of veteran expected ability, $\hat{a}_{i1}$. This simply says that the option value of being hired is increasing in the uncertainty about expected ability. However, these expressions vary directly with novice expected ability. If expected ability differs among workers about whom the market is more and less uncertain, the relative sizes of these expressions for workers about whom the market is more and less uncertain is indeterminate.

The second experimental treatment provides the market with more precise information about the performance of (randomly-selected) workers. This represents an increase in the precision of the signal $\hat{y}_i$ (an increase in h_y) for randomly-selected workers. There are two testable predictions from this treatment that compare the veteran-period employment outcomes of treatment group workers (those who received detailed evaluations) to control group workers (hired workers who received coarse evaluations).

5. Treatment group workers who receive positive detailed evaluations will have higher employment, earnings, and reservation wages in their veteran period than control group workers. Treatment group workers who receive negative detailed evaluations will have lower employment, earnings, and reservation wages.

6. Overall, veteran treatment group workers will have higher earnings and reservation wages than veteran control group workers. The treatment's effect on workers' employment is indeterminate.

Revealing positive information about workers' job performance increases their veteran-period expected abilities. Thus, they will all have higher employment, earnings, and reservation wages. Revealing negative additional information leads to the opposite results.

As a group, veteran workers' total earnings equal the total amount they produce. A more precise signal increases the market's ability to distinguish between high-ability and low-ability workers. This increases the fraction of high-ability workers employed, increasing the number of workers who generate positive marginal products, and decreases the fraction of low-ability workers employed, decreasing the number of workers with negative marginal products. The result is an increase in the total amount produced by these workers and thus their earnings. Since workers' earnings equal their reservation wages, the treatment increases workers' reservation wages as well.[9]

If the market received only an imprecise signal of their novice-period job performance, some low-ability workers would be hired in their veteran periods. A more precise signal of job performance can decrease these workers' veteran-period expected abilities sufficiently that they are unemployed in their veteran periods. However, other workers would not be hired in their veteran periods if the market only received an imprecise signal of their performance. A more precise signal can increase these workers' veteran-period expected abilities sufficiently that they are employed in their veteran period. The relative size of these two effects (and the net effect of detailed comment receipt on employment) depends on the distribution of ability.

Finally, I consider the market-level effects of employing many novice workers. Employing these novices increases the mass of veteran workers with $\hat{a}_{i1} \geq \bar{c}$ (the supply of workers recognized to be high ability). This generates two predictions.

7. Total market employment will increase as a result of the experiment by less than the increase in treatment group workers' employment.

8. Average wages will decrease, conditional on ability.

Because the experiment increases the fraction of veteran workers whose expected abilities exceed the hiring threshold, $\bar{c}$, there is an excess supply of labor after the experiment. To clear the market, the hiring threshold increases to c'. As the number of firms hiring workers increases, total employment must increase. However, as the hiring threshold increases, non-experimental workers with $\bar{c} \leq \hat{a}_i < c'$ become unemployed. Thus, total market employment will increase less than the increase in employment of treatment group workers. Wages, conditional on ability, equal workers' marginal product at the marginal firm:

$$w_{ij} = \hat{a}_{i1} - \bar{c}. \tag{1.19}$$

[9]Providing a detailed comment only affects workers' mean earnings because some workers are unemployed. If all workers were employed, the provision of a detailed comment would increase only the variance of earnings, not mean earnings.

When $\bar{c}$ increases to c', wages decrease.

1.3.6 Market Exit

This stylized model does not include the worker's choice to exit the market because including this choice adds little insight to the model. However, because market exit is an important, testable, outcome, I extend the model in this section to include this choice.

Each worker faces a distinct decision of whether to exit the market in each period of life. Before each period, a worker observes her period-specific outside option. With probability κ, her outside option is $w_1 > 0$; with probability $1 - \kappa$, her outside option is $w_0 = 0$. The worker first decides whether to accept the outside option. If she does, she exits the market. Otherwise, she remains in the market, costlessly signaling to firms she is available for jobs. After workers decide whether to exit the market, the market clears. Workers who remain in the market, but do not receive wage offers receive zero.

This section also introduces randomness in whether workers receive jobs. After the market clears, with probability ε, each worker who has not received a job receives a wage offer of $\delta > 0$.[10]

Proposition 4 *Veteran workers who have received a treatment job are less likely to exit the market than veteran control group workers. Veteran workers who receive a detailed evaluation may be either more or less likely to exit the market than workers who receive a coarse evaluation, depending on the relevant parameter values.*

Proof. *See Appendix B.* ■

1.4 Experimental Context and Design

1.4.1 Sample Selection

I recruited subjects for this experiment by posting hourly data-entry jobs to the marketplace and inviting workers to apply.[11] The job was expected to take approximately 10 hours and involved entering Census records from a PDF file into a Microsoft Excel spreadsheet. I invited applications

[10]The randomness in hiring reflects that some firms hire workers without reading workers' resumes. It only affects the model by ensuring that workers with outside option $w_0 = 0$ do not exit the market and, thus, removing the multiplicity of equilibria.

[11]I posted these jobs from the accounts of 23 different employers. Each employer posted 10 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 job.

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 workers who responded to the invitation and applied to the jobs, requesting a wage of $3 or less, form the experimental sample.

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 the 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.

The final column of Appendix Table 1 provides sample summary statistics. The majority of the sample lived in the Philippines, while India, Bangladesh, and Pakistan were also well-represented. Workers from the United States comprised 2.6% of the sample. The average worker had been in the marketplace for six months, sent 27 job applications, and passed 3.5 skills tests before the experiment. Workers with prior oDesk jobs had been in the marketplace twice as long and had sent seven times as many applications as inexperienced workers. The mean experienced worker had seven prior oDesk jobs, while the median experienced worker had four.

1.4.2 Experimental Protocol

There were three randomizations in this experiment. First, I randomized which workers I hired. Second, I randomized the amount of information I provided to the market about the job performance of hired workers. Third, three weeks after the initial hiring I invited randomly-selected experimental workers to apply to another data-entry job to assess their reservation wages.

In total, 952 workers, one quarter of the experimental sample, were selected to receive treatment jobs. The randomization was stratified on whether workers had a previous oDesk job: inexperienced workers were selected to receive treatment jobs with higher probability than were experienced workers. Hired workers were instructed to send an email to the employer (me) to receive a PDF file with the data to be entered, the data entry instructions, and a data entry template. They were given a maximum of 10 hours over one week to enter the data. They were told that, if they could not complete the task within 10 hours, they should send the file back unfinished.

When workers sent back the files, I recorded several measures of their job 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 used this information to rate all the workers on a one-to-five scale. The distribution of scores from my job was designed to match the distribution of scores of low-wage data entry workers in the marketplace, adjusted for the fact that a worker in my sample was more likely to be inexperienced than a typical oDesk worker.[12] Approximately 18% of workers did not return the file or log any hours. Under oDesk's protocol, these workers are not rated.

While all hired workers were rated using identical criteria, they did not all receive the same type of employer comment. Workers who earned a rating of least four were randomly selected to receive either an informative comment or an uninformative one. No workers with a rating below four received a detailed comment due to human subject concerns.[13]

The uninformative comment was chosen to be short and positive, like most of the comments in the marketplace. It read (where only the words in brackets vary by worker):

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

The detailed comment provided objective information on the worker's data entry speed, accuracy, ability to meet deadlines, and ability to follow instructions. Additionally, it repeated the uninformative comment, so the only difference between the two comment types was the additional objective information. 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%, top third, middle third, bottom third, bottom 10%] of providers in speed and the [top 10%, top third, middle third, bottom third, bottom 10%] in accuracy. It was a pleasure working with [x]."

Because such a high fraction of oDesk workers receive a rating of five, many workers who received this rating were in the bottom third of speed, accuracy, or both. If employers, particularly those

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

[13]The human subjects committee was concerned that giving workers negative evaluations would harm workers. They allowed me to give low numerical ratings (which were essential to the experiment). However, they permitted me to provide detailed information about the performance only of 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 these workers (e.g., they were in the bottom 10% in speed or accuracy). All analyses assessing the effect of the detailed comment condition on workers having received a rating of at least four and, thus, being eligible to receive a detailed comment.

new to the oDesk market, did not realize that a high fraction of oDesk workers receive a rating of five, these detailed comments would appear very negative. Workers have the option to remove the detailed comment, but in the marketplace as a whole only 4% of comments are removed. Workers who received a rating between three and four all received uninformative comments, while those who received a score less than three received no comment at all.

I also randomized whether the comment revealed that the employer had extensive experience hiring workers on oDesk. This was to test whether any effect of the detailed comment could be attributed to the market believing that firms who left a detailed comment were larger and more competent. The uninformative comment that provided information about the employer read:

> "It was a pleasure working with [x]. Our organization has hired hundreds of providers on oDesk."

The analogous informative 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%, top third, middle third, bottom third, bottom 10%] of providers in speed and the [top 10%, top third, middle third, bottom third, bottom 10%] in accuracy. This is based on our experience with hundreds of providers on oDesk. It was a pleasure working with [x]."

Figure 3 gives a schematic representation of these treatments. Dashed connecting lines indicate random assignment while solid connectors indicate non-random assignment.

To elicit workers' reservation wages, three weeks after the initial randomization, I randomly selected 630 workers from my sample to contact again from the account of a different employer. I contacted both workers who received treatment jobs and those who had not. These workers were invited to apply to another data-entry job requiring skills similar to the initial job. Unlike the previous job, however, these jobs had fixed wage rates: some paid $0.75 per hour, some paid $1 per hour, and some paid $2 per hour. The wage of the job each worker was invited to apply to was randomly selected. I recorded who accepted the invitation and offered a job to a randomly-selected 5% of applicants.

1.4.3 Data Collection

I hand-collected data on workers' job performance and whether they accepted the invitation to apply to the subsequent jobs. 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' employment, job applications, and profiles. The primary worker-level outcomes of interest 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). In addition to assessing workers' reservation wages directly by testing their willingness to apply to subsequent jobs, I use the wages workers post to their profiles as a measure of their reservation wages. All workers must post a wage on their profiles, so this measure is free from selection concerns. In a fully competitive market, the wage workers advertise would be their reservation wage. However, even if workers post wages above their reservation wages, there is no reason to believe the treatment affects the relationship between workers' posted wages and their reservation wages. Finally, I evaluate the effect of the treatments on workers' earnings from all jobs.

1.5 Worker-Level Effects

1.5.1 Effect of Job Receipt on Employment Outcomes

In this section, I evaluate the effect of the first treatment – providing workers with employment in their "novice period" – on their earnings, employment, and posted wages in the two months after the experiment: their "veteran period." The model predicts that employing these workers allows the market to identify high-ability workers, generating an option to hire the high-ability workers as veterans. Because these newly-identified high-ability workers will be rehired, the treatment increases workers' employment rates. Since workers receive the option value of being hired, the treatment also increases workers' earnings and reservation wages.

Appendix Table 1 shows that the randomization produced similar treatment and control samples on pre-experiment observable characteristics (conditional on the stratification). Because the randomization was stratified on whether workers had prior oDesk experience, all results are conditional on this variable.

Table 2 compares the employment, earnings, and posted wages of the treatment and control groups *excluding* the experimental jobs in the two months following the experiment. It shows that obtaining a treatment job substantially increased the future employment of both experienced and inexperienced workers. The treatment increased the probability that an inexperienced worker obtained any job in the subsequent two months from 12% to 30%. It also almost tripled the number of jobs the average inexperienced worker obtained from 0.28 to 0.81. Workers with prior oDesk experience generally worked much more following the experiment than did inexperienced workers. Obtaining a treatment job increased the probability an experienced worker was employed in the subsequent two months from 55% to 61% and increased the average number of jobs she obtained by 25% from 2.0 to 2.5.

Receiving a treatment job increased the average wage inexperienced workers posted on their profiles, a proxy for their reservation wages, from \$2.03 to \$2.28. This is an increase of approximately 12%. Among experienced workers, the treatment group's average posted wage was \$0.12 larger than the control group's, but this difference is not significant. The treatment almost tripled the earnings of inexperienced workers from \$10.06 to \$28.43. This increase in earnings of \$18.37 exceeds the \$16.03 I spent, on average, to hire these workers. Among experienced workers, the treatment group's earnings were more than \$23 higher than the control group's, which also exceeds the \$18.72 I spent, on average, to hire them, but this treatment effect is imprecisely measured.

The treatment's effect on workers' earnings corresponds, in the model, to the option value of hiring them. It is striking that it is so large given that the treatment was not targeted towards workers thought to have particularly high option values. Some treatment group workers had such low expected abilities and option values that they were in the dark gray region of Figure 2 and would not have been hired by the social planner. Others had expected abilities in the unshaded region and were already recognized as high ability before the experiment. The large average option value of hiring these workers suggests that some of them may have been inefficiently employed before the experiment. However, workers are only inefficiently unemployed if the *social* benefit of hiring them exceeds the *social* cost. The increase in the treatment group's earnings overstates the benefits to workers if some of the jobs they obtained as a result of the experiment would have otherwise gone to other workers or if treatment group workers face a non-zero cost of effort. However, it does not include any increase in firm profits. I estimate the experiment's effect on market surplus in Section 1.6.

The model predicts that the treatment's effects should be larger for workers about whom the market is more uncertain, conditional on expected ability. The market should be more uncertain about the ability of workers lacking experience, suggesting that the treatment should have larger effects for these workers. While the treatment effects are much larger in percentage terms for inexperienced workers, in absolute terms, the effect for inexperienced workers is only statistically different (larger) for one of the five outcomes: the probability that workers obtained an additional job. This is most likely because workers with and without previous jobs do not have the same expected abilities. The model predicts the treatment will have the largest effects for workers with expected abilities close to, but below, the hiring threshold. Workers with previous jobs had expected ability above the hiring threshold when they were first hired. After their jobs, if their expected abilities are no longer above the hiring threshold, they will be relatively close to it. On the other hand, workers without prior oDesk experience may have expected abilities relatively far away from the hiring threshold.[14]

Table 3 estimates the effect of job receipt on the employment outcomes of the entire sample. It displays estimates of the regression

$$y_i = \alpha + \beta job_i + \gamma_1 prevjob_i + X_i\gamma_2 + \varepsilon_i \tag{1.20}$$

where i indexes the worker, y_i is one of the five employment outcomes analyzed in Table 2, and job_i is an indicator for whether the worker received the treatment job. An indicator for whether the worker had a previous job, $prevjob_i$, is included because of the stratification, while ε_i is the mean-zero error term. The vector X_i contains worker characteristics measured before the experiment. The regressions in Panel A do not include any worker characteristics except for a dummy for whether the worker was in the second (randomly-selected) wave of the experiment. Regressions in Panel B add controls for the wage the worker requested when applying to the treatment job, controls for how long the worker had been on oDesk before the experiment, and country fixed effects.[15] Panel C additionally includes dummies for whether workers took each of the six most popular oDesk skills tests among low-wage data-entry workers and whether they scored above average on each of these tests.[16]

[14] I have attempted to test this explanation by estimating the treatment effect conditional on expected ability. However, the worker characteristics in the data are not very predictive and, thus, the standard errors are too large to meaningfully test this explanation.

[15] These regressions include dummies for whether the worker offered a wage of \$1 to \$1.99, a wage of \$2 to \$2.99, and a wage of exactly \$3. The excluded category is offering a wage less than \$1. They also include dummies for whether the worker had been on oDesk from 60 to 120 days, 121 to 180 days, 181 to 240 days, or more than 240 days. The excluded category is having joined oDesk in the last 60 days.

[16] The six tests were each taken by over 1,000 of the experimental workers. They include two tests about how oDesk works, two English tests, a Windows XP test, and a Microsoft Office test.

The table shows that the treatment significantly increases the employment, earnings, and reservation wages of experienced and inexperienced workers taken together. The estimates are very similar across panels, which is consistent with random assignment and the balanced covariates of the treatment and control groups. While the treatment effects are large, they are smaller than the coefficients on having had a previous job, γ. This is as expected given that this coefficient includes both the treatment effect of having had several previous jobs and the effect of differential selection into employment.

The employment outcomes of treatment group workers vary greatly with the rating I gave them. After two months, workers who received ratings of five earned \$34 more than the control group, while workers who received ratings of one and two earned \$23 less. This is consistent with employers learning about worker quality from the ratings. Alternatively, it could be that workers who received scores of five are more talented than the average member of the control group and would have earned more even without the rating, while workers who received ratings of one or two are less talented.

1.5.2 Effect of Detailed Evaluations on Employment Outcomes

This section analyzes the effect on workers' future employment outcomes of providing the market with more detailed information about their job performance. If the content of the comment affects the market's expectation of worker ability, revealing positive information about workers' performance will improve their employment outcomes, while revealing negative information will worsen their employment outcomes. Because providing this information allows employers to identify and hire the high-ability workers, the model predicts that workers who received detailed comments should have higher average earnings and reservation wages after the experiment than workers who received coarse comments.

Appendix Table 2 shows that the sample chosen to receive detailed comments was similar on observables to the sample chosen to receive uninformative comments. It considers only workers who received a rating of four or five because only these workers were eligible to receive detailed comments.

I determine the effect of receiving a particular type of detailed comment by comparing the employment outcomes of workers with the same performance who received coarse and detailed comments. I estimate the following regression:

$$y_i = \alpha + \beta_1(x_i \times detailed_i) + \beta_2 x_i + \beta_3 detailed_i + \varepsilon_i. \tag{1.21}$$

The dependent variable, y_i, is one of the five employment outcomes examined in Table 2. The variable $detailed_i$ indicates that a worker received a detailed comment while x_i is a positive measure of worker performance. The regression is limited to workers who received a score of at least four because receipt of a detailed comment was randomized only among these workers. The model predicts that β_3, the effect of the detailed comment for workers without the (positive) characteristic (e.g., workers who did not meet the deadline), is negative. It predicts that β_1, the additional effect of the detailed comment for workers with these (positive) characteristics, is positive as is $\beta_1 + \beta_3$, the overall effect of the comment for workers with each characteristic.

Table 4 displays the results of estimating this equation where x_i is an indicator for meeting the deadline. It shows that revealing that workers did not meet the deadline decreased earnings by \$35 on average. Revealing that a worker met the deadline increased earnings by \$59 relative to revealing she did not and by \$24 relative to leaving only a coarse comment. These are large effects; they are even bigger than the effect of receiving a job. There are large effects consistent with the model on the other employment outcomes as well, though except for hours worked, they are imprecisely measured.

This is the only piece of information that appeared to matter to employers. Appendix Table 3 shows that revealing that workers followed instructions, were in the top third of workers in speed, or were in the top third of workers in accuracy did not affect employment outcomes. This is most likely because meeting deadlines was the first piece of information revealed in the comment. Because of the length of the comment, the parts about speed and accuracy were not immediately visible on most workers' profiles; one had to click on the continuation to see them.

To test whether receiving a detailed comment increased average earnings and reservation wages, I estimate the following regression:

$$y_i = \alpha + \beta_1 detailed_i + \beta_2 rate45_i + \beta_3 job_i + \gamma_1 prevjob_i + X_i\gamma_2 + \varepsilon_i \tag{1.22}$$

where the variable $rate45_i$ indicates that workers received a rating of at least four. The worker characteristics, X_i, are the same controls included in Table 3 and β_1 is the coefficient of interest. Table 5 displays the results. The coefficient estimates are similar in all three panels as one would expect. The table shows that receiving a detailed comment led to a large increase in earnings ($23.85) for the average worker within two months. It also led to an increase in workers' reservation wages measured by the wages posted on their profiles.[17] The measured effect of the treatment on the number of jobs obtained is large and positive, but far from significant.[18]

1.5.3 Reservation Wage Treatment

In addition to using the wages posted on workers' profiles as a proxy for their reservation wage, I directly measured workers' reservation wages by inviting a randomly-selected sample to apply for additional jobs three weeks later. These jobs had a stated hourly wage, either $0.75, $1, or $2, and workers were randomly assigned to invitation pools with these different wages. To test the prediction that receiving a treatment job increased workers' reservation wages, I estimate the following regression:

$$apply_i = \alpha + \beta_1 job_i + \beta_2 prevjob_i + \gamma_w + \gamma_w \times job_i + \gamma_w \times prevjob_i + \varepsilon_i. \tag{1.23}$$

Here, $apply_i$ is a dummy for whether the worker applied for the job.[19] The variables γ_w are fixed effects for the job's wage and $\gamma_w \times job_i$ are interactions of this wage with an indicator for receiving a treatment job. Because I stratified job receipt on having a previous job, I must also interact these wage dummies with whether workers had a previous job.

Table 6 shows the results of this regression. Receiving a treatment job made workers 14 percentage points more likely to apply to these jobs. This is consistent with the model extension that predicts that receiving a treatment job will induce workers to remain in the market. Workers with previous jobs were also more likely to apply to these jobs. Unsurprisingly, workers were more likely to apply to higher wage jobs.

If receiving a treatment job increased workers' reservation wages, treatment group workers would be relatively more likely to apply to higher wage jobs than control group workers. The coefficients show this pattern. Treatment group workers were (an insignificant) 6 percentage points

[17] Both of these coefficients are significant only at the 10% level. The increase in earnings after one month ($13.46) is significant at the 5% level.

[18] It is significant at the 10% level after one month.

[19] The invitation to apply to each job explained that the job paid a stated wage, either $0.75, $1, or $2 per hour, and that workers who requested a higher wage would not be considered. The invitation described how to comply with this requirement. Still, there were some workers who applied to the jobs with wages higher than specified. I consider these workers as choosing not to apply to the job. However, the results are very similar if these workers are counted as applying.

more likely to apply to the $0.75 job than control group workers and (a significant) 20 percentage points more likely to apply to the $2 job. However, the differences between these coefficients are not significant.[20]

1.5.4 Mechanism for the Effect

The previous sections show that the experimental results are consistent with the model's predictions. However, the interpretation of these results depends on the mechanism that generates them. This section analyzes whether the results could be generated by three alternative mechanisms. The first is that completing the treatment job provides workers with human capital and this additional human capital improves workers' employment outcomes. The second is that the act of hiring a worker or leaving a detailed comment may by itself cause the market to positively update its belief about worker quality, improving employment outcomes. The third is that the treatment affects outcomes by changing worker behavior, not employer beliefs. The first two explanations seem unlikely. I find that the third may account for some, but not all, of the treatment effect.

It is relatively unlikely that working in the treatment job substantially increased workers' human capital. Workers worked for a maximum of 10 hours; the average hire worked for only 7.6 hours. Given the workers' offline experience, this is 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, as all hired workers completed the same task, human capital accumulation cannot explain why the detailed comment increased average earnings.

Hiring a worker in itself would lead the market to positively update its beliefs about the worker's ability if employers received different signals about worker quality. This explanation is less likely here than in a traditional labor market because, on oDesk, all employers see exactly the same resume and there are no face-to-face interactions. However, employers can interview workers via online chat or telephone. Moreover, employers may place different values on the same information. For example, some employers may know Filipino universities well, while others may not.

An empirical test of this explanation relies on the fact that the market observes immediately that a worker has begun working, but cannot see the worker's rating until nine to 11 days later. Panel A of Table 7 evaluates the treatment's effect in the week workers were working, while Panel

[20] I don't display the effect of the detailed comment on workers' reservation wages as measured by this treatment because the standard errors of this regression are so large as to not be informative. These results are available upon request.

B evaluates the treatment's effects the following week. The table shows that there is no effect of the treatment before workers were rated, but there is a large effect afterwards.[21]

A final alternative is that the treatments do not change workers' expected abilities; rather, they change workers' willingness to apply for jobs. For example, while completing the treatment job, workers might realize that oDesk jobs are more desirable than they believed and apply to more jobs as a result. It is difficult to rule out the possibility that this explanation causes part of the effect, given that the model extension in Appendix B also predicts that workers who received a treatment job will apply to more oDesk jobs. In fact, workers who received a treatment job did apply to more jobs than control group workers. However, this alternative hypothesis cannot explain the fact that workers who received treatment jobs were more likely to obtain the jobs they applied to than control group workers or why they had higher reservation wages. It also cannot explain the effects of receiving a detailed comment which did not change application patterns.

Panel A of Table 8 analyzes the effect of receiving a treatment job on applications. The first two columns display the results of estimating Equation (1.20) where the dependent variables are (1) an indicator for sending at least one application within two months and (2) the number of applications a worker sent. The average treatment group worker was 24 percentage points more likely to send at least one application and sent 24 more applications on average as a result of receiving the treatment job. Both of these effects are much greater for workers who did not have a previous job.

The third column of the panel examines the effect of the treatment on the competitiveness of the jobs to which workers applied. The data contains several objective features of jobs that correlate with the jobs' competitiveness. Employers posting a job must indicate whether the job is an hourly job (as opposed to a fixed wage job), its job category, and whether the employer has a preference for workers with a given English ability, number of oDesk hours, level of oDesk feedback, and a maximum or a minimum wage. I also observe the number of applicants to each job.

Using these job characteristics, I predict how difficult each job is to obtain. I consider all jobs that these workers applied to in the month before the experiment. I then regress a dummy for whether the worker's application was successful (multiplied by 100 for ease of interpreting the coefficients) on these job characteristics and worker fixed effects. The results of this regression are

[21] This alternative explanation also cannot account for the effect of the detailed treatment, since the market observes that both workers who received detailed and coarse comments were hired. However, it is possible that the market inferred that the hiring firm was competent and credible from the fact that the hiring firm left a long, detailed comment. I find directly revealing, in the comment, that a worker was hired by firm that had hired hundreds of workers on oDesk (and was therefore presumably competent and credible) had no effect on workers' outcomes.

displayed in Appendix Table 4 and, as expected, hourly jobs, jobs with employer preferences, and jobs with more applicants are more difficult to obtain.

I then use these estimates to predict a worker's relative probability of obtaining every job she applied to in the two months after the experiment. This relative predicted probability is the "ease index" of the job. The last column of Panel A displays the estimate of the following regression:

$$\widehat{ease}_j = \alpha + \beta job_i + \gamma_1 prevjob_i + X_i\gamma_2 + \varepsilon_{ij} \tag{1.24}$$

where the unit of observation is an application that worker i sent to job j. The dependent variable, $\widehat{ease}_j$ is the index and the only control variable, X_i, included is the wave of the experiment. Standard errors are clustered at the worker level. This regression shows that the average job treatment group workers applied to was substantially more difficult to obtain. This effect is sizeable: 13% of the variable's mean.

Panel B of Table 8 evaluates the effect of receiving a detailed comment on the same three outcomes. The first two columns display the results of estimating Equation (1.22) where the dependent variables are, respectively, whether the worker sent any application in the ensuing two months and the number of applications she sent. The only control variable included is the wave of the experiment. The measured effects are very small, go in opposite directions, and are far from being significant. The third column displays the results of estimating the regression

$$\widehat{ease}_j = \alpha + \beta_1 detailed_i + \beta_2 rate45_i + \beta_3 job_i + \gamma_1 prevjob_i + X_i\gamma_3 + \varepsilon_{ij} \tag{1.25}$$

where observations are applications, not workers, and the only control variable included is the wave of the experiment. It shows that receiving a detailed comment did not affect the competitiveness of the jobs to which workers applied.

Table 9 evaluates the treatment's effect on workers' application success. The model extension that includes applications predicts that workers who receive the treatment group job will be more likely to obtain the jobs they apply to than control group workers, controlling for job characteristics (but not necessarily unconditionally). Panel A displays the result of the regression

$$success_{ij} = \alpha + \beta_1 job_i + \gamma_1 prevjob_i + X_{ij}\gamma_3 + \varepsilon_{ij} \tag{1.26}$$

where the unit of observation is worker $i's$ application to job j and $success_{ij}$ is an indicator for whether this application was successful multiplied by 100 (for ease of viewing the table). The first column of the panel shows that when no worker or job controls, aside from the wave, are included, there is actually a sizable (insignificant) negative correlation between receiving a treatment job and sending a successful application.

The regression in Column 2 controls for the job's predicted ease index because the treatment affects the competitiveness of jobs that workers apply to. It also includes employer fixed effects because this predicted ease index does not control for all aspects of a job. Once the difficulty of the job is (at least partially) controlled for, receiving a treatment job is estimated to have a sizable positive (but insignificant) effect on application success. Because receiving a treatment job affects both which workers apply to jobs and the wage they ask for, Column 3 adds controls for worker characteristics and the wage requested.[22] Once these worker characteristics are included, receiving a treatment job is estimated to significantly increase the probability that a worker's application is successful by 12%.[23]

Panel B estimates the effect of receiving a detailed comment on the probability that a worker obtained a given job she applied to by estimating the equation

$$success_{ij} = \alpha + \beta_1 detailed_i + \beta_2 rate45_i + \beta_3 job_i + \gamma_1 prevjob_i + X_{ij}\gamma_2 + \varepsilon_{ij}. \tag{1.27}$$

The controls in each column are the same as in Panel A. The estimates change across columns, but not very much, consistent with the findings that receiving a detailed comment did not significantly change application patterns. In all three columns, the effect of receiving a detailed comment on application success is sizable, but imprecisely measured. This is similar to the large, but imprecise effect of receiving a job on the number of jobs a worker obtained.

1.6 Market-Level Effects

1.6.1 Effect of the Experiment on Market Employment and Wages

The model predicts that employing many novice workers affects the market by increasing the supply of workers recognized to be high ability. It thereby increases future market employment and

[22] The worker characteristics included are dummies for the number of tests the worker has passed, the number of qualifications she has, and the number of applications she sent before the experiment, whether she took and scored above average on the most popular oDesk skills test, and the total number of jobs she had before the experiment.

[23] I have controlled for selection on observables. However, if there is also selection on unobservables that is positively correlated with the selection on observables, this estimate will underestimate the true treatment effect.

decreases future market wages, conditional on ability. The model predicts that market employment should increase by less than the increase in treatment group workers' employment because treatment group workers take jobs that would have been held by other workers in the absence of this experiment. These predictions can be tested utilizing the variation in the effect of the experiment on different job categories. Employers must choose one of 74 job categories when they post a job.

The experimental workers comprise a large fraction of the oDesk marketplace. In the month prior to the experiment, they sent 11% of all the applications on oDesk and obtained 8% of the jobs. They sent 34% of all applications to data entry jobs, 33% of all applications to web research jobs, and 26% of all applications to administrative support jobs. On the other hand, they sent less than 1% of applications to jobs in 20 other job categories, primarily ones that require specific computer skills such as web programming or game development.

First, I create a variable measuring the effective number of treatment jobs created in each job category. Since workers do not choose one of these job categories, I allocate the treatment jobs to these categories using the share of their applications that treatment group workers sent to jobs in these categories in the month prior to the experiment. That is

$$jobscreated_j = \frac{\text{treatment group applications to category j jobs}}{\text{total treatment group applications}} \times 952 \tag{1.28}$$

where 952 is the total number of treatment group jobs created.

I estimate the effect of having an additional experienced, treatment group worker in a given job category on overall employment in that category by estimating the equation

$$y_{jt} = \alpha + \beta_1(jobscreated_j \times after_t) + \delta_t + \delta_j \times t + \varepsilon_{jt} \tag{1.29}$$

where the unit of observation is job category j in week t. The dependent variable is the number of jobs created in week t in job category j and δ_t are week fixed effects. The coefficient of interest, β_1, measures the number of additional jobs created in a job category with an additional treatment group worker in a week after the experiment. Job category-specific time trends $\delta_j \times t$ are included to control for the fact that even in the absence of the experiment, these categories might have grown at different rates. The regression includes all sixteen weeks in 2010 until eight weeks after the experiment, excluding the weeks of the experiment itself. Observations are weighted by their size in a pre-period.[24] Standard errors are clustered by job category.

The first column of Table 10 shows the results of this regression. It shows that each additional treatment group worker in a job category led to an additional 0.051 jobs created per week. Treatment group workers obtained an extra 0.062 jobs per week as a result of the treatment.[25] The market-level increase in employment (0.051 jobs per week) is smaller than the worker-level increase (0.062 jobs per week) as the model predicts, but not substantially or significantly so. This suggests that treatment group workers crowded out relatively little employment of other workers.

I estimate the effect of the experiment on market wages using the same identification strategy. The second and third columns of Table 10 display the results of the regression

$$wage_{ijt} = \alpha + \beta_1(jobscreated_j \times after_t) + X_i\beta_3 + \delta_t + \delta_j \times t + \varepsilon_{ijt}. \tag{1.30}$$

Observations in this regression are hourly jobs and the dependent variable is the hourly wage in the job. Because the model's predictions are conditional on worker ability, it also includes controls for worker ability.[26] Column 2 shows that having an additional treatment group worker in a given job category decreases average hourly wages by a significant \$0.0015 on average. Including worker characteristics does not affect these estimates.

1.6.2 Market-Level Welfare Analysis

The experiment's effect on market surplus can be estimated using its effects on market employment and wages. Figure 4 shows the effect of the experiment on the oDesk market. The inelastic supply curve (S) represents the number of workers recognized to be high ability. The experiment shifts this supply curve outward, increasing employment and decreasing wages. Total market surplus increases by the dark gray shaded area. The dark gray shaded triangle is the increase in firm profits from the increased employment and the dark gray shaded rectangle is the increase in worker surplus from the increased employment. The light gray shaded rectangle is a transfer from workers to firms.

Assuming linearity of the demand curve, the area of the triangle is one half the change in

[24] Because one job category had no new jobs in the pre-period, it is excluded from the regression.
[25] This is calculated by dividing the treatment effect in Table 3 by the number of weeks.
[26] These are indicators for the number of tests the worker took, the number of qualifications the worker listed, and the number of portfolio items the worker had before the experiment.

(future) employment generated by the experiment times one half the change in wages. This is:

$$\frac{1}{2} \times \underbrace{(952 \text{ jobs} \times 8 \text{ weeks} \times 1.31 \frac{\text{hours}}{\text{job} \times \text{week}})}_{\text{Increase in Employment}} \times \underbrace{(952 \text{ jobs} \times \frac{\$0.0015}{\text{job} \times \text{hour}})}_{\text{Decrease in Wages}} \approx \$6,900. \tag{1.31}$$

This calculation measures the increase in employment using a change in hours to correspond to the hourly wages. Estimating Equation (1.29) where the dependent variable is total hours worked in a job category-week shows that each treatment job created led to 1.31 additional hours worked in hourly jobs in each of the eight weeks after the experiment. Multiplying this figure by the number of weeks in the calculation (8) and the number of treatment jobs created (952) gives the overall increase in market employment resulting from the experiment. Similarly, multiplying the market-level change in wages per experimental job ($0.0015) by the number of experimental jobs (952) yields the overall decrease in market wages. This calculation shows that the increase in firm profits as a result of this experiment, not including the transfer from workers, is $6,900.

The area of the dark gray shaded rectangle, the increase in worker surplus due to increased market employment, is the increase in market employment multiplied by the prevailing wage level:

$$\underbrace{(952 \text{ jobs} \times 8 \text{ weeks} \times 1.31 \frac{\text{hours}}{\text{job} \times \text{week}})}_{\text{Increase in Employment}} \times \underbrace{\$2.09}_{\text{Wage Level}} \approx \$20,800. \tag{1.32}$$

Here, $2.09 is the average hourly wage experimental workers received after the experiment. However, the above calculation gives the increase in worker surplus only if workers face zero cost of effort. Panel A of Table 11 estimates the increase in total market surplus under different assumptions about the cost of worker effort.

These calculations may underestimate the increase in surplus on oDesk because they only consider gains for eight weeks after the experiment. However, they do not include any losses in other markets due to the increased employment on oDesk. If firms decrease their offline hiring or workers decrease their offline work when oDesk employment increases, the experiment's effect on oDesk social surplus exceeds its effect on total social surplus.

The social cost of this experiment includes the fixed cost of hiring workers and the cost of worker effort. The money paid to the workers in the treatment jobs is not a social cost; it is a transfer. If this transfer is funded by the government, there is a deadweight loss of taxation.[27] In a traditional

[27] Part of the funding for the experiment came from government grants, so there is some deadweight loss of taxation

job, these costs would be offset by the value of the output produced. However, in this experiment, there was no direct value to the output produced. Because workers were not expected to produce usable output, the fixed cost of employing them was relatively low.

Panel B of Table 11 calculates the social cost of this experiment under different assumptions about the size of these costs. The most difficult piece to estimate is the cost of worker effort. Given that 49% of the treatment group was willing to apply to a job offering a wage of $1 in the reservation wage treatment, I use this as the "best-guess" estimate of workers' cost of effort. However, I also estimate the social cost of the experiment using both higher and lower costs of effort.

This intervention was not designed optimally, nor intended, to increase net market surplus. Nonetheless, unless workers' cost of effort is estimated to be their full wage, the benefits of this intervention exceed its costs. This suggests that novice employment on oDesk was inefficiently low before the experiment.

1.7 Conclusion

The market for entry-level workers is characterized by high unemployment and strong competition for jobs. This paper proposes that the high unemployment partially results from the fact that firms do not receive the full benefit of discovering talent. Hiring an inexperienced worker requires an investment from firms, but reveals information about worker ability, generating an option to hire a known, good worker in the future. To the extent that this information is public, workers receive the option value. If workers cannot compensate firms for their investments, too few inexperienced workers will be employed.

This paper formalizes this intuition into a model of the labor market and tests this model through a large field experiment. Consistent with the model's predictions, giving workers the opportunity to demonstrate their abilities through short jobs increased their future employment, earnings, and reservation wages. By increasing the supply of workers recognized to be high-ability, it increased market employment, decreased market wages, and increased total market surplus. Despite the fact that this intervention was untargeted, under plausible assumptions, it increased market surplus by more than its social cost. This suggests that, before the experiment, firms hired inefficiently few novices and discovered inefficiently little novice talent.

in this context. More generally, the deadweight loss of taxation is included to show the cost of this program if it were not run as an experiment, but was subsidized by the government.

Directly giving the market more information about workers' job performance also affected workers' employment outcomes in the way the model predicts. Providing the market with additional positive information increased workers' future earnings while providing it with additional negative information decreased their earnings. Overall, giving the market more information about workers' job performance increased their future earnings. The model predicts that, by allowing the market to more accurately identify high-ability workers, this additional information increases the option value of being hired and workers' future earnings.

Although these results come from a particular online marketplace, there are two settings to which these insights might generalize: traditional (offline) entry-level labor markets and international trade of goods and labor. Compared to traditional entry-level employers, oDesk employers have more uncertainty about the abilities of job applicants. They do not get to meet applicants, may not be familiar with workers' credentials from foreign schooling systems or employers, and have very limited ability to verify these credentials. They may also be hiring workers for more technical tasks where ability varies more among applicants. This suggests that the inefficiency may be larger in the oDesk marketplace than in a traditional labor market. On the other hand, the fixed costs of hiring an oDesk worker are particularly low. oDesk workers can be hired with the click of a mouse. The tasks they complete are typically well-defined, easy to explain, and require no training. If workers do not perform well, they can be fired instantaneously with no penalty. This suggests that the inefficiency may be larger in a traditional labor market.

Under the supposition that these insights apply to traditional entry-level labor markets, theory suggests several policies that would reduce the inefficiency. Subsidizing entry-level hiring compensates employers for discovering inexperienced workers' talent, inducing them to hire more novices. The market has already developed institutions that compensate employers for hiring novices. Internships allow firms to pay novices low wages while, in Europe, unions and industry consortiums directly pay some workers' initial salaries. Alternatively, workers could be hired for short-term employment by a separate, credentialing employer who would relay their job performance to the market. For this employer to improve market efficiency, its assessments of job performance would need to be credible. Workers would be willing to pay the employer directly to take this employment "test," but they may be credit-constrained, in which case public funds could subsidize the employer. A traditional test would not be sufficient to correct this inefficiency given the difficulty of determining workers' motivation and reliability through means other than employment.

As most oDesk jobs are offshored from U.S. employers to foreign workers, the experimental results in this paper 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, new plants and capital. On the other hand, 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.

1.8 Appendix A

This section proves Proposition 3. Only the most efficient firms will hire a worker, so there is a threshold, $\bar{c}$, such that all firms with $c_j \leq \bar{c}$ hire a worker and all firms with $c_j > \bar{c}$ do not. Firms must be indifferent to hiring any employed worker and the marginal firm must earn zero profit. Thus, the wages of employed workers equal

$$w_{ij} = \hat{\eta}_i - \bar{c}. \tag{1.33}$$

If all novice workers with $\hat{a}_{i0} \geq \bar{a}_A$ and only these novices are hired, then

$$\hat{\eta}_{i0} = (\gamma_A + \gamma_B)\hat{a}_{i0}. \tag{1.34}$$

Veteran workers who have worked have known types. Thus, experienced veteran group C workers will not receive wage offers, but experienced veteran group A and B workers will work if their market wages exceed their outside options:

$$w_{ij} = \hat{a}_{i1} - \bar{c} > w_0. \tag{1.35}$$

If only novices with $\hat{a}_{i0} \geq \bar{a}_A$ and veteran group A and B workers with $\hat{a}_{i1} - \bar{c} > w_0$ work in the market, then, by definition, $\bar{c} = c(\bar{a}_A)$.

Novice group C workers with $\hat{a}_{i0} \geq a_A$ strictly prefer any positive wage offer to their outside options. Novice group A and B workers will work in the market when

$$\underbrace{((\gamma_A + \gamma_B)\hat{a}_{i0} - \bar{c} - w_0)}_{\text{Net Wage in Novice Period}} + \underbrace{\Pr(\hat{a}_{i1} - \bar{c} \geq w_0) \times E[\hat{a}_{i1} - \bar{c} - w_0 | \hat{a}_{i1} - \bar{c} \geq w_0]}_{\text{Veteran-Period Option Value}} \geq 0. \tag{1.36}$$

By the definition of $\bar{a}_A$ and the fact that both terms of this expression are increasing in $\hat{a}_{i0}$, group A workers will work exactly when $\hat{a}_{i0} \geq \bar{a}_A$. Since $w_B < w_A$, Group B workers with $\hat{a}_{i0} \geq \bar{a}_A$ strictly prefer working in the market to their outside option.

By the assumption in Equation (1.16) of the text, only these workers: novices with $\hat{a}_{i0} \geq \bar{a}_A$ and veteran group A and B workers with $\hat{a}_{i1} - \bar{c} > w_0$ work in the market. Firms will not hire novice workers with $\hat{a}_{i0} < \bar{a}_A$. Group A workers with $\hat{a}_{i0} < \bar{a}_A$ do not work in the market, so firms could only hire group B and C workers at these ability levels. But, these workers' net marginal products are negative. Firms will also not hire veterans who have not worked in their novice periods. These workers all have

$$\hat{a}_{i1} = \hat{a}_{i0} < \bar{a}_A. \tag{1.37}$$

Novice workers benefit more from market work than do veterans because novices receive the option value from talent discovery. Group A workers with these expected abilities prefer their outside option to market work as novices, so they also prefer their outside options to market work as veterans. Firms will not hire only group B and C workers with these expected ability levels.

This equilibrium does not maximize total surplus. Consider a social planner who has the same information as employers; that is, he cannot distinguish group A, B, and C novices. He must, therefore, choose which novices to hire solely on the basis of their expected abilities. He will set an ability threshold $\hat{a}_{i0} = a^*$ such that all novices with $\hat{a}_{i0} \geq a^*$ and only these will be employed. Denote the fixed cost of the marginal firm who hires a worker in the social planner's solution by c^*. For notational simplicity, denote the marginal worker's expected net product at this marginal firm by w^*:

$$w^* = (\gamma_A + \gamma_B)a^* - c^*. \tag{1.38}$$

The total social surplus from hiring a worker with $\hat{a}_{i0} = a^*$ is a weighted average of the surplus from hiring group A, B, and C workers:

$$\begin{aligned}
\gamma_A[w^* - w_A + \Pr(\hat{a}_{i1} - c^* &\geq w_A) \times E[\hat{a}_{i1} - c^* - w_A | a^* - c^* \geq w_A]] \\
+\gamma_B[w^* - w_B + \Pr(\hat{a}_{i1} - c^* &\geq w_B) \times E[\hat{a}_{i1} - c^* - w_B | a^* - c^* \geq w_B]] \\
+\gamma_C[w^*] &= 0.
\end{aligned}$$

If the social planner used the same hiring threshold as the market, $a^* = \bar{a}_A$, then the same number of firms would hire a worker in the social planner's solution as the market equilibrium $c^* = \bar{c}$. Under these parameters, group A workers are indifferent to taking this wage offer, so the value of the first line is zero. However, group B and C workers prefer this wage offer to their outside option, so the values of the second and third lines are positive. Thus, the left-hand side of this equation is positive. For this equation to hold, the social planner's hiring threshold must be below the market's: $a^* < \bar{a}_A$. The market hires too few novice workers.

1.9 Appendix B

This section proves Proposition 4. No worker with $w_0 = 0$ will exit the market. If she exits the market, she receives 0, while if she remains in the market, her expected gain is at least $\varepsilon\delta$. Veteran workers with outside option w_1 will exit the market if

$$\hat{a}_{i1} < w_1 + \bar{c}. \tag{1.39}$$

They will earn

$$w_{ij} = \max(\hat{a}_{i1} - \bar{c}, 0). \tag{1.40}$$

if they remain in the market and will exit if their outside option is greater than this wage.

The first treatment, job receipt, only affects novice workers with expected abilities

$$\hat{a}_{i0} < \bar{c} \tag{1.41}$$

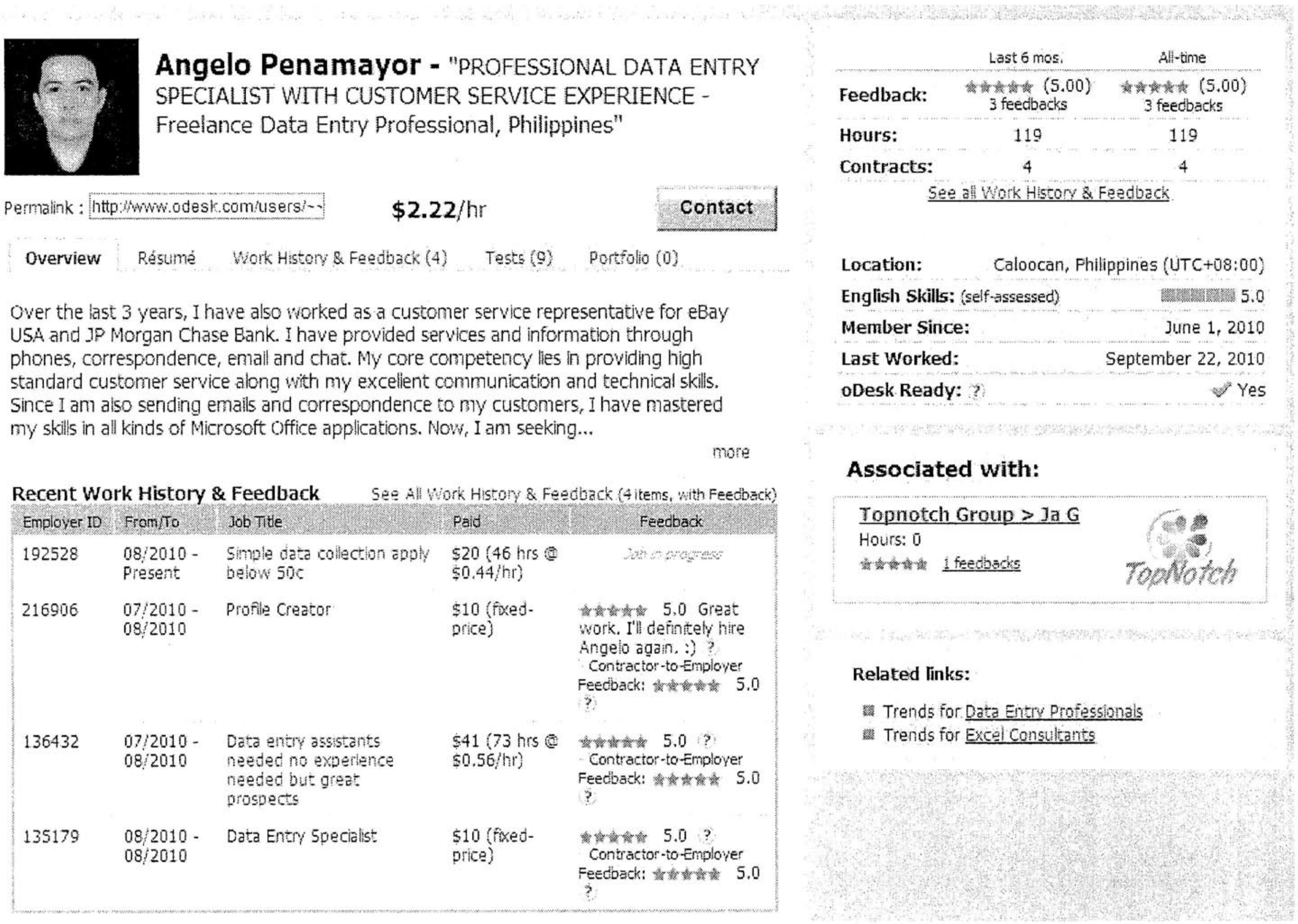
because control group novices with higher expected abilities are employed by other firms. All control group workers with $\hat{a}_{i0} < \bar{c}$ have

$$\hat{a}_{i1} = \hat{a}_{i0} < \bar{c} < w_1 + \bar{c}. \tag{1.42}$$

They will all exit the market if offered outside option w_1. On the other hand, some treatment group workers with $\hat{a}_{i0} < \bar{c}$ will perform well enough in the treatment job that their veteran-period expected ability exceeds the market-exit threshold, $w_1 + \bar{c}$. They will not exit the market even if offered outside option w_1.

The second experimental treatment, providing workers with a more detailed evaluation, has an indeterminate effect on the fraction of workers whose expected ability exceeds the market-exit threshold. Some workers' veteran expected abilities would exceed the market-exit threshold if they received only a coarse comment. Receipt of a detailed evaluation decreases some of these workers' expected abilities below the market-exit threshold. Other workers' veteran expected abilities would be below the market-exit threshold if they only received a coarse comment. Receipt of a detailed evaluation increases some of these workers' expected abilities above the market-exit threshold. The distribution of workers' expected abilities determines the relative sizes of these two groups and the net effect of the treatment on market exit.

Figure 1. Example oDesk Profile



Angelo Penamayor - "PROFESSIONAL DATA ENTRY SPECIALIST WITH CUSTOMER SERVICE EXPERIENCE - Freelance Data Entry Professional, Philippines"

Permalink : http://www.odesk.com/users/~~ $2.22/hr Contact

Overview Résumé Work History & Feedback (4) Tests (9) Portfolio (0)

Over the last 3 years, I have also worked as a customer service representative for eBay USA and JP Morgan Chase Bank. I have provided services and information through phones, correspondence, email and chat. My core competency lies in providing high standard customer service along with my excellent communication and technical skills. Since I am also sending emails and correspondence to my customers, I have mastered my skills in all kinds of Microsoft Office applications. Now, I am seeking...

more

Recent Work History & Feedback See All Work History & Feedback (4 items, with Feedback)

Employer ID	From/To	Job Title	Paid	Feedback
192528	08/2010 - Present	Simple data collection apply below 50c	$20 (46 hrs @ $0.44/hr)	Job in progress
216906	07/2010 - 08/2010	Profile Creator	$10 (fixed-price)	★★★★★ 5.0 Great work. I'll definitely hire Angelo again. :) ? Contractor-to-Employer Feedback: ★★★★★ 5.0 ?
136432	07/2010 - 08/2010	Data entry assistants needed no experience needed but great prospects	$41 (73 hrs @ $0.56/hr)	★★★★★ 5.0 ? Contractor-to-Employer Feedback: ★★★★★ 5.0 ?
135179	08/2010 - 08/2010	Data Entry Specialist	$10 (fixed-price)	★★★★★ 5.0 ? Contractor-to-Employer Feedback: ★★★★★ 5.0 ?

	Last 6 mos.	All-time
Feedback:	★★★★★ (5.00) 3 feedbacks	★★★★★ (5.00) 3 feedbacks
Hours:	119	119
Contracts:	4	4

See all Work History & Feedback

Location: Caloocan, Philippines (UTC+08:00)
English Skills: (self-assessed) 5.0
Member Since: June 1, 2010
Last Worked: September 22, 2010
oDesk Ready: ? ✓ Yes

Associated with:

Topnotch Group > Ja G
Hours: 0
★★★★★ 1 feedbacks
TopNotch

Related links:

- Trends for Data Entry Professionals
- Trends for Excel Consultants

Figure 2. Expected Novice Ability

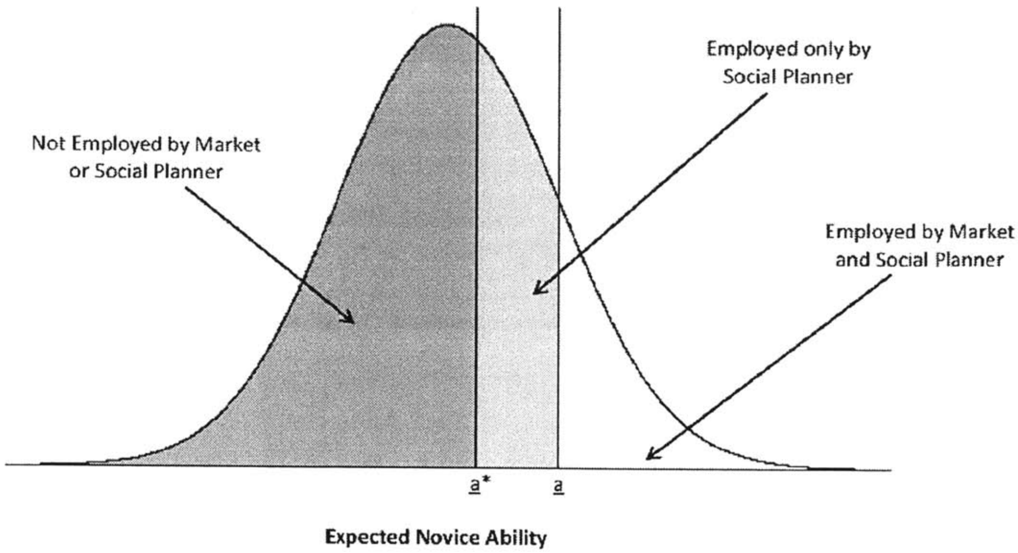


Figure 3. Experimental Design

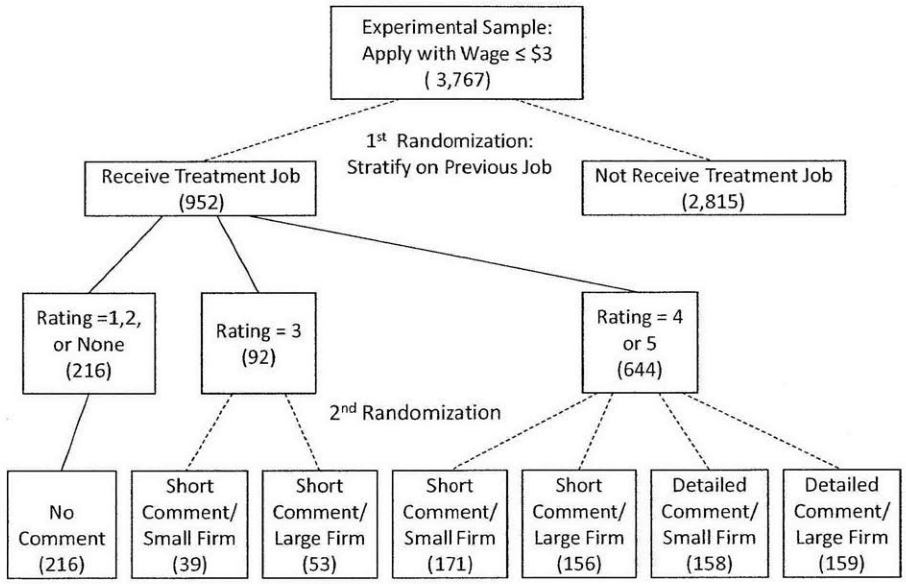


Figure 4. Experiment's Market-Level Effects

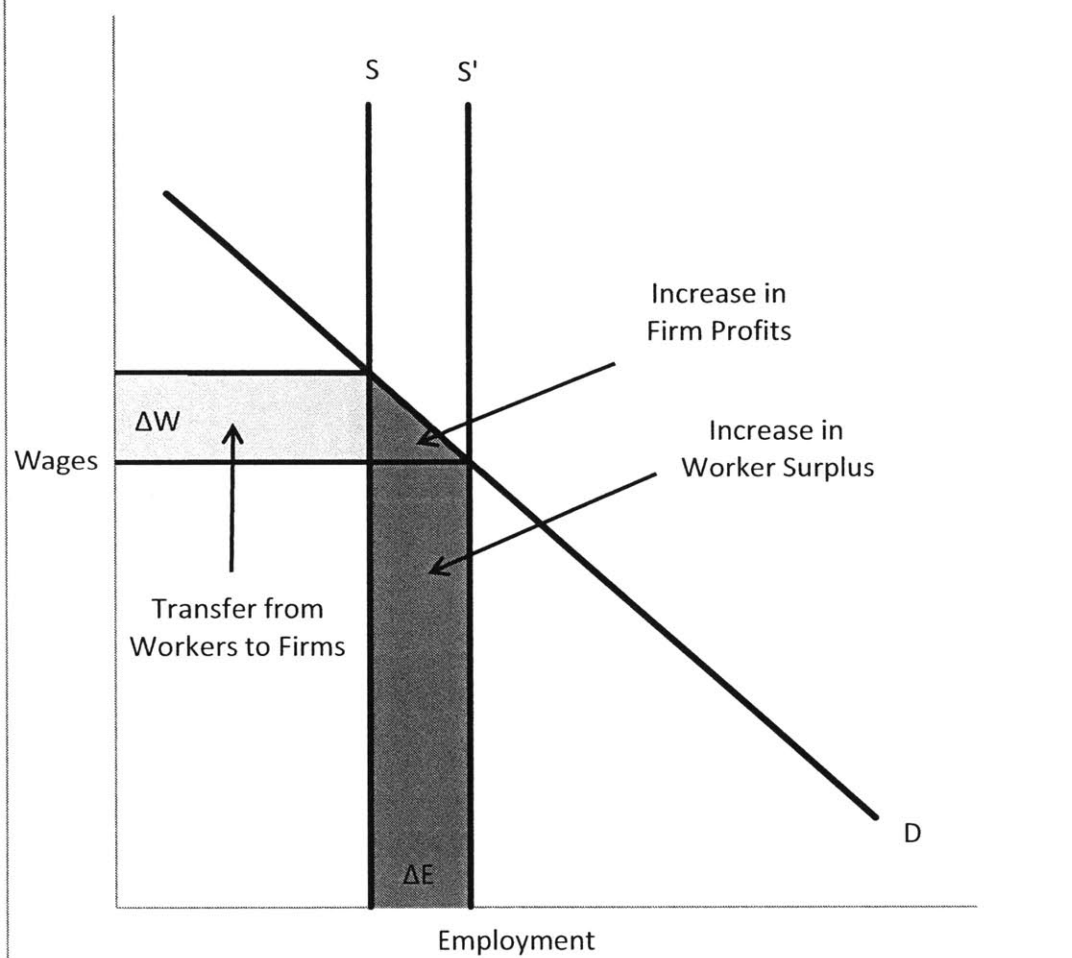


Table 1. Sample Selection

	No Previous Job	Any Previous Job	Total	Percentage
Contacted	7,136	2,826	9,962	100%
Applied	2,324	1,528	3,852	39%
Applied with Wage ≤ \$3	2,298	1,469	3,767	38%
Received Treatment Job	736	216	952	25% of experimental sample

Notes: The first row enumerates the workers invited to apply to the job while the second counts those that accepted that invitation. The third row is the experimental sample: workers who applied to the job requesting a wage less than or equal to \$3. The final row counts the workers who were randomly selected to receive a treatment job. Unless otherwise indicated, percentages refer to the percentage of contacted workers.

Table 2. The Effect of the Treatment Job on Employment Outcomes
During the Two Months After the Experiment

	A. Workers with No Previous Job			B. Workers with Previous Jobs		
	Treatment	Control	Difference	Treatment	Control	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Total Jobs	0.807	0.284	0.523**	2.463	1.958	0.505**
Standard Error of Difference			(0.073)			(0.244)
p-value: Equality of Differences						[0.947]
Any Job	0.299	0.117	0.182**	0.611	0.545	0.066*
Standard Error of Difference			(0.017)			(0.037)
p-value: Equality of Differences						[0.004]
Hours Worked	12.40	5.36	7.05**	60.99	47.80	13.19*
Standard Error of Difference			(1.59)			(7.12)
p-value: Equality of Differences						[0.45]
Posted Wage	2.28	2.03	0.25**	2.50	2.38	0.12
Standard Error of Difference			(0.05)			(0.08)
p-value: Equality of Differences						[0.25]
Earnings	28.43	10.06	18.37**	144.02	120.60	23.41
Standard Error of Difference			(3.57)			(18.14)
p-value: Equality of Differences						[0.80]
Observations	736	1,562	2,298	216	1,253	1,469

Notes: Cells in Columns 1, 2, 4, and 5 display the mean value of the employment outcome indicated by the row heading for workers indicated by the column heading within two months of the experiment. All experimental jobs and earnings are excluded. Columns 3 and 6 provide the difference in mean outcomes for treatment and control workers without and with previous jobs, respectively. The standard error of this difference is in parentheses. One asterisk indicates the difference is significant at the 10% level and two asterisks indicate the difference is significant at the 5% level. Column 6 presents, in brackets, the p-value from a test that the treatment effects for workers with and without previous jobs are equal.

Table 3. The Effect of the Treatment Job on Employment Outcomes
During the Two Months After the Experiment

	Total Jobs (1)	Any Job (2)	Hours Worked (3)	Posted Wage (4)	Earnings (5)
			A. No Controls		
Treatment Job	0.498**	0.150**	9.62**	0.21**	20.36**
	(0.092)	(0.017)	(2.59)	(0.06)	(6.29)
Previous Job	1.674**	0.406**	43.50**	0.33**	111.42**
	(0.094)	(0.015)	(2.68)	(0.04)	(6.69)
		B. Controls for Wage Offered, Date Joined oDesk, and Country			
Treatment Job	0.477**	0.148**	9.15**	0.21**	19.22**
	(0.093)	(0.017)	(2.58)	(0.05)	(6.26)
Previous Job	1.778**	0.416**	46.87**	0.30**	117.96**
	(0.103)	(0.016)	(2.96)	(0.04)	(7.48)
		C. Additional Controls for Test Scores			
Treatment Job	0.483**	0.147**	9.14**	0.21**	19.47**
	(0.092)	(0.017)	(2.59)	(0.05)	(6.27)
Previous Job	1.658**	0.388**	44.89**	0.28**	110.71**
	(0.102)	(0.017)	(3.02)	(0.04)	(7.45)
Observations	3,767	3,767	3,767	3,767	3,767

Notes: This table displays the results of estimating Equation 20. The regressions in Panel A contain no controls except for an indicator for being in the second experimental wave. The regressions in Panel B additionally include controls for the wage requested when applying for the treatment job, the date the worker joined oDesk and country fixed effects. (See footnote 15 for more information.) The regressions in Panel C add indicators for whether workers took each of the six most popular skills tests among low-wage data-entry workers and scored above average on these tests. All experimental jobs are excluded. Huber-White standard errors are in parentheses. Two asterisks indicate the coefficient is significant at the 5% level.

Table 4. Effect of Receiving a Detailed Comment: Meeting Deadlines During the Two Months After the Experiment

	Total Jobs (1)	Any Job (2)	Hours Worked (3)	Posted Wage (4)	Earnings (5)
Met Deadline × Detailed Comment	0.442	0.138	31.57**	0.52	59.27**
	(0.446)	(0.334)	(15.19)	(0.61)	(21.08)
Met Deadline	0.791**	-0.013	-0.47	0.07	19.90
	(0.311)	(0.190)	(14.53)	(0.37)	(17.61)
Detailed Comment	-0.238	-0.095	-27.40**	-0.28	-34.88**
	(0.388)	(0.331)	(13.98)	(0.60)	(15.51)
Sum of Detailed Coefficients	0.204	0.043	4.17	0.24	24.39
p-value of F-test	0.350	0.277	0.48	0.07	0.09
Observations	644	644	644	644	644

Notes: This table displays the results of estimating Equation 21 where x_i is an indicator for meeting the deadline. It includes only workers who received ratings of at least four and were thus eligible to receive a detailed comment. All experimental jobs are excluded. Huber-White standard errors are in parentheses. One asterisk indicates the coefficient is significant at the 10% level and two asterisks indicate the coefficient is significant at the 5% level.

Table 5. The Effect of Receiving a Detailed Comment on Employment Outcomes During the Two Months After the Experiment

	Total Jobs (1)	Any Job (2)	Total Hours (3)	Posted Wage (4)	Earnings (5)
			A. No Controls		
Detailed Comment	0.209	0.041	3.73	0.24*	23.85*
	(0.208)	(0.038)	(5.58)	(0.13)	(13.43)
Rating = 4 or 5	0.612**	0.159**	11.88	0.26**	16.46
	(0.176)	(0.036)	(4.58)	(0.08)	(10.87)
			B. Controls for Wage Offered, Date Joined oDesk, and Country		
Detailed Comment	0.217	0.041	4.32	0.22*	25.15*
	(0.212)	(0.039)	(5.55)	(0.13)	(13.31)
Rating = 4 or 5	0.606**	0.161**	10.52**	0.23**	11.42
	(0.181)	(0.036)	(4.59)	(0.06)	(11.03)
			C. Additional Controls for Test Scores		
Detailed Comment	0.182	0.035	3.53	0.22*	22.92*
	(0.210)	(0.038)	(5.55)	(0.13)	(13.25)
Rating = 4 or 5	0.612**	0.161**	10.89**	0.22**	12.16
	(0.181)	(0.036)	(4.61)	(0.06)	(10.96)
Observations	3,767	3,767	3,767	3,767	3,767

Notes: This table displays the results of estimating Equation 22. The regressions in Panel A contain no controls except for an indicator for being in the second experimental wave. The regressions in Panel B additionally include controls for the wage requested when applying for the treatment job, the date the worker joined oDesk and country fixed effects. (See footnote 15 for more information.) The regressions in Panel C add indicators for whether workers took each of the six most popular skills tests among low-wage data-entry workers and scored above average on these tests. All experimental jobs are excluded. Huber-White standard errors are in parentheses. One asterisk indicates the coefficient is significant at the 10% level and two asterisks indicate the coefficient is significant at the 5% level.

Table 6. Effect of the Treatment Job on Reservation Wages
Dependent Variable: Indicator for Applying

	(1)	(2)
Treatment Job	0.141**	
	(0.041)	
Treatment Job × \$0.75 Wage		0.059
		(0.071)
Treatment Job × \$1 Wage		0.162**
		(0.066)
Treatment Job × \$2 Wage		0.195**
		(0.079)
\$1 Wage Job	0.141**	0.099
	(0.044)	(0.064)
\$2 Wage Job	0.292**	0.212**
	(0.049)	(0.073)
Previous Job	0.075*	
	(0.041)	
Interactions of Prev. Job with Wage	No	Yes
p-value: Equality of Treatment Coefficients		0.386
Mean of Dependent Variable	0.395	0.395
Observations	630	630

Notes: This table displays the results of estimating Equation 23. Only workers who were invited to apply to this job are included. Huber-White standard errors are in parentheses. One asterisk indicates the coefficient is significant at the 10% level and two asterisks indicate the coefficient is significant at the 5% level.

Table 7. Effect of the Treatment Job on Employment Before and After Rating

	A. Week Between Hiring and Rating		B. Week After Rating	
	Total Jobs	Any Job	Total Jobs	Any Job
	(1)	(2)	(3)	(4)
Treatment Job	-0.010	0.000	0.086**	0.057**
	(0.014)	(0.010)	(0.021)	(0.012)
Previous Job	0.252**	0.170**	0.247**	0.167**
	(0.018)	(0.011)	(0.020)	(0.011)
Second Wave	0.008	0.001	-0.023	-0.014
	(0.016)	(0.010)	(0.017)	(0.010)
Constant	0.037	0.033	0.046	0.034
	(0.010)	(0.007)	(0.012)	(0.007)
Observations	3,767	3,767	3,767	3,767

Notes: Both panels present the results of estimating Equation 20. Panel A only includes jobs obtained in the week immediately following the worker's hire. Panel B only includes jobs obtained in the week immediately following workers' evaluations. Huber-White standard errors are in parentheses. Two asterisks indicate the coefficient is significant at the 5% level.

Table 8. Effect of the Treatments on Application Patterns
During the Two Months After the Experiment

	A. Effect of Treatment Job			B. Effect of a Detailed Comment		
	Sent Any Application	Applications Sent	Ease Index	Sent Any Application	Applications Sent	Ease Index
	(1)	(2)	(3)	(4)	(5)	(6)
Detailed Comment				-0.008	0.53	0.052
				(0.019)	(5.33)	(0.132)
Rating = 4 or 5				0.266**	29.17**	-0.558**
				(0.028)	(4.45)	(0.171)
Treatment Job	0.238**	23.88**	-0.421**	0.060**	3.85	0.015
	(0.015)	(2.18)	(0.085)	(0.027)	(2.38)	(0.145)
Previous Job	0.337**	30.38**	0.599**	0.332**	29.91**	0.582**
	(0.013)	(1.89)	(0.080)	(0.013)	(1.89)	(0.081)
Mean of Dep. Var.	0.745	30.28	2.973	0.729	25.28	2.973
Observations	3,767	3,767	114,082	3,767	3,767	114,082

Notes: The first two columns of each panel present the results of estimating Equations 20 (Panel A) and 22 (Panel B) where the unit of observation is the worker and the dependent variables are an indicator for whether the worker sent any application and the number of applications sent in the two months following the experiment. Huber-White standard errors are in parentheses. Columns 3 and 6 present the results of estimating Equations 24 and 25 where the unit of observation is an application. In these regressions, standard errors are clustered by worker. All experimental jobs are excluded. Two asterisks indicate the coefficient is significant at the 5% level.

Table 9. Effect of the Treatments on Application Success
Dependent Variable: Indicator that an Application is Successful × 100

	A. Effect of the Treatment Job			B. Effect of a Detailed Comment		
	(1)	(2)	(3)	(4)	(5)	(6)
Detailed Comment				0.381	0.349	0.257
				(0.331)	(0.269)	(0.257)
Rating = 4 or 5				-0.062	0.524	0.582*
				(0.499)	(0.331)	(0.302)
Treatment Job	-0.379	0.234	0.416**	-0.482	-0.332	-0.158
	(0.244)	(0.198)	(0.188)	(0.450)	(0.305)	(0.274)
Previous Job	2.035**	0.752**	0.731**	2.045**	0.777**	0.727**
	(0.238)	(0.192)	(0.210)	(0.239)	(0.191)	(0.210)
Ease Index	No	Yes	Yes	No	Yes	Yes
Employer Fixed Effects	No	Yes	Yes	No	Yes	Yes
Worker Characteristics	No	No	Yes	No	No	Yes
Mean of Dep. Var.	3.59	3.59	3.59	3.59	3.59	3.59
Observations	114,082	114,082	114,082	114,082	114,082	114,082

Notes: Panels A and B display the results of estimating Equations 26 and 27, respectively. The regressions in Columns 1 and 4 include no controls. The regressions in Columns 2 and 5 control for the ease index and employer fixed effects. The regressions in Columns 3 and 6 add worker characteristics measured before the experiment: dummies for the number of tests the worker passed, the number of qualifications she had, whether she took and scored above average on the most popular oDesk skills test, the total number of jobs she had before the experiment, and the wage she requested when applying for the treatment job. All experimental jobs are excluded. Standard errors are clustered by worker. One asterisk indicates the coefficient is significant at the 10% level and two asterisks indicate the coefficient is significant at the 5% level.

Table 10. Market-Level Effects of the Experiment

	A. Jobs Created	B. Hourly Wages	
	(1)	(2)	(3)
Jobs Created x After	0.051**	-0.145**	-0.139**
	(0.023)	(0.046)	(0.049)
Worker-Level Effect	0.062		
Week Fixed Effects	Yes	Yes	Yes
Category-Specific Time Trends	Yes	Yes	Yes
Worker Characteristics	No	No	Yes
Observations	1,825	125,456	125,456

Notes: This table displays the results of estimating Equations 29 (Column 1) and 30 (Columns 2 and 3). In Column 1, observations are weighted by the number of jobs created in the category in a pre-period. These regressions contain outcomes from 16 weeks before the experiment to 8 weeks afterwards, excluding the weeks of the experiment. All standard errors are clustered by job category. Two asterisks indicate the coefficient is significant at the 5% level. See the text for the creation of the "Worker-Level Effect."

Table 11. Estimated Effect of the Experiment on Market-Level Efficiency

	High Benefit, Low Cost	Medium Benefit, Medium Cost	Low Benefit, High Cost
		A. Increased Market Surplus	
Increased Firm Profit	$6,860	$6,860	$6,860
Increased Worker Surplus Not Including Cost of Effort	$20,772	$20,772	$20,772
Cost of Effort for Additional Hours Worked	-$4,970 ($0.50/hour)	-$9,939 ($1 per hour)	-$20,772 (Full wage)
Total Surplus	$22,662	$17,693	$6,860
		B. Social Cost	
Fixed Cost of Employing ($10 per hour spent)	$476 3 min/worker	$793 5 min/worker	$1,587 10 min/worker
Deadweight Loss of Taxation (Total Cost: $15,842)	$0 (0%)	$3,168 (20%)	$3,961 (35%)
Worker Effort	$3,622 ($0.50/hour)	$7,245 ($1 per hour)	$15,842 (Full wage)
Total Social Cost	$4,098	$11,206	$21,390
		C. Overall Welfare Change	
Total Surplus - Total Social Cost	$18,564	$6,487	-$14,530

Notes: The “Total Surplus” row sums the three prior rows in Panel A. The “Total Social Cost” row sums the three prior rows in Panel B. The "Total Surplus - Total Social Cost" is the difference between these two summations.

Appendix Table 1. Verification of Randomization: Treatment Job

	No Previous Job		At Least 1 Previous Job		All Workers
	Treatment Job	No Job	Treatment Job	No Job	
	(1)	(2)	(4)	(5)	(6)
Posted Wage	1.97	1.94	2.01	2.04	1.98
Days Since Joining oDesk	137	126	251	257	179
Number of Applications Sent	25*	22*	160	167	27
Self-Assessed English Score	4.5	4.5	4.7	4.7	4.56
Number of Tests Passed	2.7	2.7	4.5	4.7	3.48
Philippines	63%	61%	63%	64%	63%
India	10%	11%	10%	12%	11%
Bangladesh	10%	10%	15%**	10%**	10%
Pakistan	6.3%	7.0%	5.1%	4.6%	5.9%
United States	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	1562	216	1253	3,767

Notes: Each cell presents the mean value of the characteristic indicated by the row heading for workers indicated by the column heading before the experiment. One asterisk indicates the difference between treatment and control workers is significant at the 10% level and two asterisks indicate the difference is significant at the 5% level.

Appendix Table 2. Verification of Randomization: Detailed Comment

	Detailed Comment (1)	Uninformative Comment (2)
Posted Wage	1.99	2.03
Days Since Joining oDesk	163	164
Number of Applications Sent	55	53
Self-Assessed English Score	4.5	4.5
Number of Tests Passed	3.2	3.4
Philippines	67%	67%
India	9%	8%
Bangladesh	11%	9%
Pakistan	4.7%	4.9%
United States	1.6%	2.8%
Fraction with Previous Job	25%	25%
Workers with Previous Jobs Only		
Number of Previous Jobs	6.2	6.7
Average Feedback Score	4.6*	4.3*
Observations	317	327

Notes: Each cell presents the mean value of the characteristic indicated by the row heading for workers indicated by the column heading before the experiment. One asterisk indicates the difference between treatment and control workers is significant at the 10% level. Only hired workers who received a rating of four or higher are included.

Appendix Table 3. Effect of Detailed Comment: Instructions, Speed, and Accuracy During the Two Months After the Experiment

	Total Jobs (1)	Any Job (2)	Hours Worked (3)	Posted Wage (4)	Earnings (5)
			A. Instructions		
Follow All Instructions × Detailed Comment	0.579	0.097	-5.61	0.53**	10.67
	(0.485)	(0.092)	(15.78)	(0.26)	(37.72)
Follow All Instructions	0.022	0.024	4.43	-0.31	-6.95
	(0.316)	(0.065)	(9.77)	(0.18)	(21.95)
Detailed Comment	-0.234	-0.032	8.20	-0.17	15.82
	(0.415)	(0.080)	(14.49)	(0.20)	(34.57)
Sum of Detailed Coefficients	0.345	0.065	2.60	0.36	26.49
p-value of F-test	0.169	0.151	0.678	0.028	0.080
			B. Speed		
Top Third × Detailed Comment	0.594	0.035	-15.14	0.11	-2.95
	(0.471)	(0.081)	(11.60)	(0.33)	(29.08)
Top Third	0.108	0.017	12.91	0.14	17.28
	(0.276)	(0.056)	(7.95)	(0.11)	(16.16)
Detailed Comment	0.004	0.030	9.74	0.21*	25.82
	(0.256)	(0.049)	(7.70)	(0.12)	(17.96)
Sum of Detailed Coefficients	0.598	0.065	-5.40	0.32	22.87
p-value of F-test	0.130	0.316	0.534	0.310	0.318
			C. Accuracy		
Top Third × Detailed Comment	-0.075	0.063	-7.25	-0.16	-5.69
	(0.426)	(0.081)	(11.98)	(0.23)	(29.60)
Top Third	-0.238	-0.072	0.09	-0.03	-0.93
	(0.283)	(0.055)	(7.56)	(0.11)	(15.80)
Detailed Comment	0.227	0.016	6.45	0.30	26.06
	(0.287)	(0.050)	(7.60)	(0.19)	(17.71)
Sum of Detailed Coefficients	0.152	0.079	-0.80	0.14	20.37
p-value of F-test	0.628	0.212	0.931	0.296	0.391
Observations	644	644	644	644	644

Notes: This table displays the results of estimating Equation 21 where x_i is an indicator for following all instructions (Panel A), an indicator for being in the top third of workers in speed (Panel B), or an indicator for being in the top third of workers in accuracy (Panel C). It includes only workers who received ratings of at least four and were thus eligible to receive a detailed comment. All experimental jobs are excluded. Huber-White standard errors are in parentheses. One asterisk indicates the coefficient is significant at the 10% level and two asterisks indicate the coefficient is significant at the 5% level.

Appendix Table 4. Estimates of Ease Index

Dependent Variable: Indicator that an Application is Successful × 100	
Hourly Job	-0.755**
	(0.154)
Number of Applicants	-0.011**
	(0.000)
Data Entry	-1.693**
	(0.133)
Pref. for English Ability	-2.391**
	(0.134)
Pref. for Number of oDesk Hours	-0.588**
	(0.148)
Pref. for Level of oDesk Feedback	-1.457**
	(0.148)
Pref. for Maximum Wage below $5	-0.791**
	(0.130)
Pref. for Minimum Wage above $3	-0.007
	(0.169)
Mean of Dependent Variable	3.57
Observations	156,184

Notes: This table presents the results of an ordinary least squares regression of an indicator for whether an application was successful (multiplied by 100) on the job characteristics listed in the left-most column and worker fixed effects. The unit of observation is an application. All applications sent by experimental workers in the month before the experiment are included. Standard errors are clustered by worker. Two asterisks indicate the coefficient is significant at the 5% level.

Chapter 2

Small Differences that Matter: Mistakes in Applying to College

2.1 Introduction

Where a student applies to college affects whether she attends college and the type of college she attends, which in turn influences her human capital accumulation and future earnings. Yet, very little is known about how students decide where to apply.

This paper estimates the sensitivity of students' college application decisions to a $6 decrease in the cost of sending standardized test scores to colleges. I find that, in response to the $6 cost change, students sent substantially more score reports and applications. They also widened the range of colleges to which they sent their scores: sending scores to colleges that were both more- and less-selective than any they would have sent scores to otherwise. I present a simple accounting exercise that shows that the average low-income student received expected benefits from sending an additional score report that were much larger than $6. I then consider explanations for students' large reaction to such a small cost change. In the preferred explanation, I show that choosing which colleges to apply to is a complex problem. In the face of this complexity, students may rely on rules of thumb in deciding where to apply. The $6 cost decrease could have changed students' rules of thumb, explaining why it had such a large impact on students' application decisions.

Before the fall of 1997, students taking the ACT, a popular college entrance exam, could send their test scores to three colleges for free while each additional score report cost $6. Afterwards, students could send four score reports for free with the same $6 cost for each additional report beyond four. All students graduating high school in 1996 or earlier were eligible for only three free score reports while all students graduating in 2000 or later were eligible for four. Students in the high school class of 1998 were eligible for four free score reports if they took the ACT in their senior year, but only three if they took the test in their junior year.

Using micro data from the ACT Corporation detailing which colleges students sent their ACT scores to, I show that many students sent an additional score report when the fourth score report became free. Figure 1 shows the fraction of each graduating class that sent exactly three and exactly four score reports by whether they took the ACT or the SAT, a competing college entrance exam. In classes that received only three free ACT score reports, over 70% of ACT-takers sent three score reports and less than 5% sent four. However, once the fourth score report became free, less than 10% of ACT-takers sent three score reports while approximately 60% sent exactly four. Over this period, the fraction of SAT-takers that sent three score reports stayed relatively constant, while there was a small decrease in the fraction that sent four scores.

Since score reports sent are not a perfect proxy for applications, I use the Beginning Postsecondary Students Longitudinal Study (BPS) to show that the increase in score-sending translated into an increase in applications. The number of applications sent by ACT-takers over this period increased much more than did the number of applications sent by SAT-takers.

When students received four free score reports, they widened the range of colleges to which they sent scores. Some students sent scores to colleges that were more selective than any they would have sent scores to otherwise. This gave students an additional opportunity to be accepted at and attend a more-selective college. Some students also sent scores to colleges that were less selective and had higher admission rates than any they would have sent scores to otherwise. This increased the probability that they were admitted to at least one college they could afford. On average, students sent scores to colleges that were more competitive when they received four free score reports.

Sending an additional college application could have large welfare effects for low-income students by increasing the probability that they attend college and attend selective colleges. Low-income students are less likely to apply to and attend selective colleges than are their higher-income peers conditional on high school achievement (Bowen et al., 2005; Pallais and Turner, 2006; and Spies, 2001) even though they receive particularly high returns from attending selective colleges (Dale and Krueger, 2002 and Saavedra, 2008). Additionally, over a quarter of low-income students who say they would like to attend a four-year college and apply to at least one do not matriculate at one

(Avery and Kane, 2004) even though low-income students receive particularly large returns from attending college (Card, 1995).

I do not have application or matriculation data for the students in the ACT database. However, I use conservative assumptions, such as the assumption that students who sent score reports to colleges as a result of receiving the fourth free score report were only 50% as likely to be admitted as traditional applicants, to show that the average low-income student would have gained over $6,000 in expected future earnings by sending an additional score report.

To interpret students' large response to this $6 cost change, I derive the maximization problem that students should solve when choosing the colleges they apply to. I provide evidence that students' large response to the small cost change is inconsistent with optimizing behavior. However, it is not altogether surprising that students are not optimizing. Because students have almost innumerably many combinations of colleges to choose among, they cannot easily evaluate the utility they will get from attending each college (Avery and Kane, 2004), and they are likely uncertain about the probability they will be admitted to different colleges, students' maximization problem is very difficult to solve.

Thaler and Sunstein (2008) suggest that when faced with complex choices, individuals often rely on rules of thumb. In this context, students may interpret the ACT providing three (or four) free score reports as an indication that sending three (or four) score reports is recommended. They may then decide to send the recommended number of scores. In this way, this paper's findings are complementary to the findings in Madrian and Shea (2001), Choi et al. (2002), and Thaler and Sunstein (2008) that individuals are significantly affected by the default savings and health insurance plans.

The paper proceeds as follows. Section 2.2 describes the policy change and the datasets used. Section 2.3 discusses the literature on low-income students' college decisions and shows that low-income students of all ability levels send score reports to less-competitive colleges than do similarly-able higher-income students. Section 2.4 shows that ACT-takers sent more score reports and applications after the cost change while Section 2.5 shows that they also widened the range of colleges to which they sent scores. I provide an estimate of the earnings gain low-income students receive from sending an additional score report in Section 2.6. Section 2.7 models how students should make application decisions and examines why their application decisions deviate from optimality. Section 2.8 concludes.

2.2 Background Information

2.2.1 Policy Change

Before the fall of 1997, the ACT allowed students to send three free score reports. Starting in the fall of 1997, the ACT provided four free score reports. Thus, students graduating from high school in 1996 or earlier received only three free score reports. Students in the class of 1998 who took the ACT in their junior year or earlier received only three free score reports, while those who took the ACT in their senior year received four. All students in the classes of 2000 and 2004 received four free score reports.

The cost of each additional score report was constant at $6 from the fall of 1995 to the fall of 2001. Before the fall of 1995, each additional score report cost between $4 and $5.50 while after the fall of 2001, each additional score report cost $7. Students in the classes of 1991, 1992, and 1994 paid slightly less for each additional score report than did most students in the classes of 1996, 1998, and 2000, while students in the class of 2004 paid $1 extra. Focusing the results on the classes of 1996, 1998, and 2000 where most students paid $6 for each additional score report does not change the results.

While the monetary cost of sending four score reports increased in the fall of 1997, the non-monetary cost did not. Because students chose the colleges they sent scores to when they registered for the test, they had to provide payment regardless of the cost of sending score reports.[1] Both before and after the cost change, students were given six lines to record the colleges to which they wanted their scores sent.

2.2.2 Data

This paper uses three datasets: a large micro database from the ACT Corporation, the BPS, and the American College Survey (ACS).

ACT Database

While the SAT is better known, the ACT is a popular college entrance exam, especially in the Midwest. More students take the ACT than the SAT and almost all United States colleges, including the entire Ivy League, accept ACT scores.

[1] This is not necessarily true for low-income students who could waive the testing fee but not score-sending fees.

The database from the ACT Corporation includes information on one out of every four Caucasians, one out of every two minorities, and every student who did not provide a race who planned to graduate from high school in 1991, 1992, 1994, 1996, 1998, 2000, or 2004 when taking the ACT. This provides a large sample: 2,486,159 observations on students who attended a valid high school with over 287,000 in each year.[2] I observe each student's ACT score, high school GPA, race, gender, family income, high school, classes taken, and extracurricular activities. I also observe up to six colleges to which each student sent her ACT scores. Seeing only six colleges does not lead to a large censoring problem. As only 2% of students who sent any score reports sent six, very few students could have wanted to send more than six score reports.

Using score-sending data as a proxy for application data has become quite common in the literature.[3] By comparing applications sent to SAT score reports sent, Card and Krueger (2005) find that score-sending data are a good proxy for application data. However, score-sending data are not a perfect proxy. I see only the colleges students indicated they wanted their scores sent to at the time of test registration. Some students may not have applied to colleges to which they sent their scores. Others may have applied to colleges to which I do not observe them sending scores because they sent SAT scores instead or sent ACT score reports after test registration. Empirically, this latter concern is not large: in 2004, the only year for which data on score reports sent after registration are available, only 8% of students who sent score reports at registration sent additional score reports afterwards. However, because of these potential concerns, I also use the BPS to analyze application behavior directly.

Beginning Postsecondary Students Longitudinal Survey

The baseline BPSs of 1996 and 2004 report the number of applications sent by approximately 27,000 students who first entered postsecondary education between July 1, 1995 and June 30, 1996 and July 1, 2003 and June 30, 2004, respectively. The survey also reports whether the students took the ACT or SAT. The public-use BPS does not allow multivariate regression, only tabulations. However, the demographics of the SAT- and ACT-taking populations realized similar small changes

[2]For much of the analysis, I exclude students who sent no score reports. Since the ACT Corporation believes that people only take the ACT to apply to college, it seems likely that students who sent no score reports at the time of test registration sent score reports after viewing their scores or sent SAT score reports instead. However, Appendix Table 1 shows that the dramatic change in score-sending caused by the cost change is robust to including these observations. Excluding students who sent zero score reports still leaves a large sample: 2,049,389 observations with more than 253,000 in each year.

[3]See, for example, Card and Krueger (2005), Abraham and Clark (2006), Pope and Pope (2008), Long (2004), and Pallais and Turner (2006).

between the two waves of the survey.

American College Survey

I use the ACS to obtain information on the colleges to which students sent their scores. For each college, the ACS provides the 25th and 75th percentile ACT scores of the entering freshman class,[4] the number of applicants and admitted students, and the application fee.

I use test scores from freshmen matriculating in a base year, 1993, as my measure of college selectivity so that the analysis is not confounded by colleges becoming more competitive over time. I only discuss the results using colleges' 25th percentile ACT scores because the results are very similar for the 25th and 75th percentiles. However, results for the 75th percentile are available from the author. As many readers may not be familiar with the meaning of different ACT scores, I convert the ACT scores of colleges' freshmen into percentiles of the distribution of colleges the class of 1996 sent scores to. For example, if a college has a normalized 25th percentile ACT score of 30, 30% of score reports students sent in 1996 went to colleges with lower 25th percentile ACT scores. I also report the results using the actual ACT scores in footnotes for readers who are familiar with the metric.

2.3 College Choices of Low-Income Students

Low-income students are less likely to attend college than are their higher-income peers, conditional on high school achievement. Ellwood and Kane (2000) find that in the high school class of 1992, 66% of students with family incomes in the top quartile attended a four-year college, while only 28% of students from the bottom quartile did so. They estimate that 40% of this gap remains after controlling for 12th grade test scores. This is troubling as Card (1995) finds that the return to a year of college is higher for disadvantaged students.

Low-income students are also underrepresented at selective colleges. Hill et al. (2005) find that in the 2001-2002 school year only 10% of students at 28 elite private colleges (COFHE colleges)[5] were from the bottom 40% of the income distribution. Winston and Hill (2005) find that this was

[4]Many colleges do not provide both SAT and ACT scores of matriculating freshmen. For schools that only provide their freshman classes' 25th and 75th percentile SAT scores, I impute the corresponding ACT scores using a concordance produced by the College Board. The ACS does not provide information on the freshmen classes' median ACT or SAT scores.

[5]The COFHE (Consortium on Financing Higher Education) colleges are 31 elite private schools that include the entire Ivy League. Hill et al. (2005) and Winston and Hill (2005) have data on only 28 of these colleges.

35% less than the fraction of high-achieving high school students that were from the bottom 40% of the income distribution.

Many studies have found a large return to college quality for students of all income levels (e.g. Hoxby, 1998; Black et al., 2005; Zhang 2005; Brewer et al., 1999; and Black and Smith, 2006). There is no consensus in this literature, however, because Dale and Krueger (2002) find there is no return to college selectivity for most students when they compare the earnings of students who were admitted to the same colleges, but chose to attend different ones. Yet, Dale and Krueger do find large returns to college selectivity for low-income students. Many other studies (e.g. Saavedra, 2008; Monks, 2000; Behrman et al., 1996; and Loury and Garman, 2000) find that low-income students and minorities have particularly high returns to attending selective colleges.

Low-income students' application choices may play a role in their underrepresentation in college and at selective colleges. In their evaluation of the Boston COACH program, Avery and Kane (2004) argue that providing disadvantaged high school students with help in choosing which colleges to apply to increased the fraction of these students who attended college. Even with substantial college counseling, however, 27% of the disadvantaged students in Avery and Kane's study who wanted to attend a four-year college and applied to at least one were not admitted to one at which they decided to matriculate, including a large number of students who had high GPAs. This suggests that if low-income students apply to more appropriate sets of colleges, they may be more likely to attend college.

Similarly, Bowen et al. (2005), Spies (2001), and Pallais and Turner (2006) find that low-income students are less likely to apply to elite colleges than are their higher-income peers, conditional on high-school achievement. Yet, Bowen et al. (2005) finds that, conditional on applying, low-income students are no less likely to gain admission to or matriculate at elite colleges than are their higher-income peers.

2.3.1 Score-Sending by Family Income

Among those who sent score reports, low-income students sent fewer score reports than did their higher-income peers before four free score reports were provided.[6] Table 1 displays the results from

[6] Students with higher family incomes were more likely to send zero score reports at the time of test registration, however. As the free score reports were available only at registration, this suggests higher-income students were more likely to pay $6 for the option of seeing their scores before sending them to colleges or more likely to take both the SAT and ACT.

estimating

$$y_i = \alpha + \beta_1(middle\_income_i) + \beta_2(high\_income_i) + X_i\beta_3 + \varepsilon_i \tag{2.1}$$

where y is the number of score reports individuals sent to colleges. The variables *middle_income* and *high_income* are dummies for having a family income between \$36,000 and \$80,000 per year and above \$80,000 per year, respectively. The dummy for being low-income, having a family income less than \$36,000 per year, is omitted. The vector X contains many controls for student demographics, high school achievement, and extracurricular participation,[7] as well as high school fixed effects. The first column contains no controls, while the second column adds all of the controls in footnote seven, and the third column adds high school fixed effects.[8] Throughout the paper, standard errors are clustered at the state level.

Including all of the controls and high school fixed effects, I find that high-income students sent 0.13 more score reports on average than did observationally equivalent low-income students. Middle-income students also sent more score reports than did their low-income peers, though substantially fewer than did high-income students.

Low-income students also sent scores to less-competitive colleges on average than did their higher-income peers. Figure 2 plots the average selectivity of colleges that high- and low-income students with each ACT score sent scores to. To compute the selectivity of each college, I normalize the 25th percentile ACT score of the college's incoming freshmen as described in Section 2.2. Students' ACT scores, plotted on the x-axis are converted to percentiles of the 1996 ACT distribution for ease of interpretation. The graph shows that virtually throughout the ability distribution low-income students sent their scores to less-selective colleges than did their higher-income peers, though this difference is larger for high-ability students.

Table 2 shows that low-income students were less likely than their higher-income peers to send scores to selective "reach" colleges and more likely to send scores to "safety schools" whose students

[7] I control for state dummies, race dummies, a dummy for United States citizenship, an indicator for English being the primary language spoken in the home, the number of siblings the student has under the age of 21, the size of the community the student lived in, and the student's gender. I also control for her high school GPA, dummies for each ACT score, and the number of years of English and math classes she had taken as well as dummies for attending a private high school, being on a college preparatory track, having any college credit, and having taken honors English or math. Finally, I include indicators for ever having been elected to a student office, working on the staff of a school paper or yearbook, earning a varsity letter for sports participation, and holding a regular part-time job.

[8] It is not clear that I should control for high school fixed effects. If students' high schools are partially a function of their incomes, regressions that include high school fixed effects likely underestimate the effect of family income on application choices. However, if students' high schools are a function of characteristics correlated with income, omitting high school fixed effects would likely overestimate the effect of family income. Because I would rather conservatively underestimate the gap between the income groups, I prefer the specification with high school fixed effects. However, I almost always report the results both ways.

were below their ability level. Panels A and B show the results of estimating equation (2.1) where the dependent variables are the selectivities of the most- and least-selective colleges students sent scores to, respectively. These are the largest and smallest 25th percentile ACT scores of the colleges students sent scores to. Columns 1 through 3 contain the same controls as in Table 1. Columns 4 through 8 contain all the controls and high school fixed effects and are limited to students in different parts of the ACT score distribution.

Without controlling for students' differential high school achievement or demographics, low-income students' most-selective colleges were 12 percentage points less selective than those of high-income students. This is the difference between the University of Pennsylvania and the University of Nevada, Reno. The background characteristics in footnote seven explain approximately 60% of this disparity and adding high school fixed effects on top of these controls explains approximately 35% of the remaining gap. Conditional on background characteristics and high school fixed effects, the disparity was 3 percentage points: the difference in selectivity between the University of Pennsylvania and Kenyon College. The differences between low- and high-income students' least-selective colleges were slightly smaller.[9] Similar to Figure 2, columns 4 through 8 show that these differences in college competitiveness existed throughout most of the ability distribution, but were larger for high-ability students.

2.4 Increase in Applications Sent

When the fourth score report became free, there was a dramatic increase in the fraction of ACT-takers sending exactly four score reports and a large decrease in the fraction sending exactly three, but very little change in the number of score reports sent by SAT-takers. Figure 1 plots the fraction of students in each high school class that sent exactly three and exactly four score reports by whether they took the ACT or the SAT. In each class in which all ACT-takers received only three free score reports over 70% of ACT-takers sent exactly three score reports while less than 5% sent exactly four. In the class of 1998 in which some ACT-takers were eligible for four free score reports, the fraction of ACT-takers sending four score reports jumped to over 40%, while the fraction of students sending three score reports plummeted to less than a third. In classes

[9]The difference in the actual (not normalized) 25th percentile ACT scores of low- and high-income students' most-selective colleges was 1.4 ACT points with no controls and 0.3 ACT points conditional on all of the background characteristics and high school fixed effects. The difference in the 25th percentile ACT scores of these students' least-selective colleges was 0.9 ACT points with no controls and 0.2 ACT points conditional on all of the controls and high school fixed effects.

where all ACT-takers were eligible for the fourth free score report, over 55% of students sent four score reports and less than 10% sent exactly three. Yet, among SAT-takers, there was actually a small decrease in the fraction of students sending four score reports and no change in the fraction of students sending three. There was a small decrease in the fraction of ACT-takers sending four score reports between 2000 and 2004. Appendix Figure 1, which replicates Figure 1 excluding students who sent no score reports, shows that some of this decrease resulted from an increase in the number of students sending no scores.

While there were large changes in the fraction of ACT-takers sending exactly three and exactly four score reports, there were very small changes in the fraction of students sending other numbers of scores. Aside from the percentage of students sending one score report in 2004, over the 13 years spanned by this data, the percentage of students sending one, two, five, and six score reports each varied by less than one percentage point, remaining almost unchanged after 1997.

Table 3 displays regression estimates of the effect of the cost change on the number of score reports sent. The first two columns report estimates of the equation

$$\begin{aligned} y_i &= \alpha + \beta_1(\text{class of } 1998_i) + \beta_2(post\_1998_i) + \beta_3 lowinc_i + \beta_4 t + \beta_5(t \times lowinc_i) \\ &\quad + X_i\beta_8 + \varepsilon_i \end{aligned} \tag{2.2}$$

The dependent variable, y, is the number of score reports sent. The variable class of 1998 is an indicator for the class of 1998 and post_1998 is an indicator for the classes of 2000 and 2004. I include separate indicators for the class of 1998 and the classes after 1998 because I expect the policy to have larger effects in years when all test-takers were eligible for the fourth free score report. The variable *lowinc* is a low-income dummy; t and $t \times lowinc$ represent a linear time trend and a separate linear time trend for low-income students, respectively. The vector X includes the covariates in footnote seven and high school fixed effects.

To show the differential response of low-income students to the cost change, the last three columns of Table 3 report the results of estimating

$$\begin{aligned} y_i &= \alpha + \beta_1(\text{class of } 1998_i) + \beta_2(post\_1998_i) + \beta_3 lowinc_i + \beta_4(lowinc_i \times \text{class of } 1998_i) \\ &\quad + \beta_5(lowinc_i \times post\_1998_i) + \beta_6 t + \beta_7(t \times lowinc_i) + X_i\beta_8 + \varepsilon_i \end{aligned} \tag{2.3}$$

where lowinc × class of 1998 and lowinc × post_1998 are the interactions of the two time indicators with the low-income dummy.

The estimates show that on average, students sent 0.79 more score reports in the classes of 2000 and 2004 than in classes in which students only received three free score reports. On average, middle- and high-income students sent 0.78 additional score reports and low-income students sent 0.81 additional reports. Predictably, the class of 1998 did not increase its score-sending as much as did later classes.[10] Estimating these regressions separately for students in different quartiles of the ACT distribution shows that once all test-takers received four free score reports, students of all ability levels increased their score-sending by similar amounts.

These results change very little with the exclusion of the controls and high school fixed effects. Appendix Table 1 shows that the results are also similar when students who sent no score reports are included. Excluding the time trends attenuates the coefficients (though they are still large and significant) as, on average, middle- and high-income students sent 0.02 fewer score reports each year between 1991 and 2004 and low-income students sent 0.01 fewer score reports.

The ACT database does not include information on applications, so I use the BPS to show that the additional score reports translated into additional applications. Table 4 reports the average number of applications sent by ACT-takers and SAT-takers who entered college during the 1995-1996 school year and by those who entered during the 2003-2004 school year. It shows that the average ACT-taker in the 2003-2004 entering class sent 0.33 more applications than did the average ACT-taker in the 1995-1996 entering class, while the average SAT-taker in the entering class of 2003-2004 sent only 0.13 more applications than her 1995-1996 counterpart. Low-income ACT-takers sent 0.39 more applications each over this period, while low-income SAT-takers sent only 0.21 more.

The ACT and SAT experienced similar changes in their test-taking populations between the entering college classes of 1995-1996 and 2003-2004.[11] Assuming that without the cost change ACT-takers would have realized the same increase in applications over this period as did SAT-takers, ACT-takers sent 0.20 additional applications as a result of the cost change. Regressions

[10]In the class of 1998, middle- and high-income students sent an additional 0.46 score reports on average, while low-income students sent an additional 0.56 reports. If low-income students were more likely than their higher-income peers to take the ACT in their senior year as opposed to their junior year, then a higher fraction of low- than high-income students in the class of 1998 would have received four free score reports. This would explain why low-income students realized a larger increase in score-sending over their higher-income peers in the class of 1998 than in later classes.

[11]Slightly smaller fractions of both test-taking populations were white in the class of 2003-2004 than in the class of 1995-1996, but there were no significant changes in the percentage of students of any other race or either gender over this period. As one would expect, the average nominal parental income of both sets of test-takers increased.

indicate that they sent 0.79 more score reports on average, implying that 25% of the additional score reports they sent turned into applications. Low-income students sent 0.18 additional applications and 0.81 additional score reports per capita as a result of the cost change, implying that 23% of their additional score reports turned into applications.[12]

2.5 Changes in the Selectivity of Colleges To Which Scores Were Sent

When students sent more score reports, they sent scores to a wider range of colleges, sending scores to colleges that were both more- and less-selective than any they would have sent scores to otherwise. Table 5 displays the results of estimating equations (2.2) and (2.3) where the dependent variable is the selectivity of the most- and least-competitive colleges each student sent scores to.

The most-selective colleges middle- and high-income students sent scores to were 4.1 points more selective in classes where all students received four free score reports than in classes where all students received only three. Their least-competitive colleges were 3.6 percentage points less competitive. Low-income students saw a larger, 5 percentage point, increase in the selectivity of their most-competitive colleges, and a decrease in selectivity of their least-competitive colleges statistically indistinguishable from that of higher-income students. A 5 percentage point selectivity difference is the difference between Cornell and Furman Universities, while a 3.6 percentage point difference is the difference between Princeton and Washington and Lee Universities. The average selectivity of colleges students sent scores to increased by 0.9 percentage points for middle- and high-income students and a not-significantly-different 1.0 percentage points for low-income students.[13] This is approximately the difference between the University of Pennsylvania and Emory University. These results are quite robust.[14] These selectivity changes are smaller for the class of 1998 in which

[12]This figure is slightly different than the $\frac{0.18}{0.81} = 22\%$ that would be expected from the text due to rounding to the two-digit figures 0.18 and 0.81 in the text.

[13]Middle- and high-income students increased the selectivity of the most-competitive colleges they sent scores to by 0.51 (not normalized) ACT points, while low-income students increased the selectivity of these colleges by 0.57 ACT points. The selectivity of the least-competitive colleges low- and higher-income students sent scores to both decreased by 0.37 ACT points. The average selectivity of the colleges students sent scores to increased by 0.07 ACT points for higher-income students and 0.11 points for low-income students.

[14]Adding a quadratic time trend, a separate quadratic time trend for low-income students, and controls for the student's interest in science, social service, business, and arts, whether the student had done community service, and her grades in algebra 1, biology, and United States history does not change the coefficients of interest. Additionally, I estimate counterfactual regressions as in Table 5 while assuming that, instead of being implemented in the fall of 1997, the new cost structure was first implemented for the class of 1992, 1994, or 1996. I drop students graduating after 1996. Only one of the six coefficients on the dummy for graduating after the counterfactual policy change is

not all students were offered four free score reports.[15]

Sending scores to less-selective colleges entailed sending scores to colleges with higher admissions rates and less-able applicant pools. As a result of the cost change, the highest admissions rate among colleges students sent scores to increased by 1.3 and 1.8 percentage points for higher- and low-income students, respectively. The average ACT scores of students who, in 1996, sent score reports to students' least-selective colleges decreased by 1.0 and 1.1 percentile for higher- and low-income students, respectively.

Table 6 re-estimates equation (2.3) for students in different parts of the ability distribution. It shows that all ability groups realized significant changes in the most- and least-competitive colleges to which they sent scores. Higher-ability students realized smaller increases in the selectivity of their most-competitive colleges. This may be because higher-ability students were sending their scores to more-competitive colleges than were lower-ability students before the cost change, so they had relatively smaller untapped pools of more-competitive colleges to which they could send their scores. The ability groups realized similar decreases in the selectivity of their least-competitive colleges.

2.6 Assessing Benefits to Students

This section provides a benchmark calculation of the benefit low-income students receive from sending an additional score report. I calculate the return that comes through two channels. First, sending an additional score report may increase the probability that a student attends college by increasing the probability that she is admitted to a college she can afford. Second, sending an additional score report may increase the probability that a student attends a more-selective college by increasing the probability that she is admitted to one.

There may be many other benefits and costs of sending an additional score report that, without a good deal more data, I cannot evaluate. Sending an additional score report may give the student more college options, allowing her to attend a college that offers her a better financial aid package

significant and this effect is small, less than 15% of the actual policy effect, with a t-statistic of 2.5.

[15]The fact that students sent their scores to a wider range of colleges after they received four free score reports does not imply that they systematically sent their fourth score reports to colleges that were either more- or less-selective than the three colleges they would have otherwise sent scores to. Many other methods of selecting colleges would have also led to this pattern. This pattern would even result if, after the cost change, students simply chose four colleges from the same distribution from which they had previously chosen three. However, students could not have chosen colleges from the same distribution before and after receiving the fourth free score report because they sent scores to more-selective colleges afterwards.

or that is a better "fit." At the same time it takes time and sometimes money[16] to apply to an additional college and costs an admissions officer time to read the application. With more applications to complete and read, students and admissions officers may spend less time on each application, potentially leading to worse matches between students and colleges. However, the calculations in this section suggest that the benefit to sending an additional score report is so large that it is likely to greatly outweigh these costs.

2.6.1 Attending College

Avery and Kane (2004) find that 27% of students from disadvantaged high schools who stated that they wanted to attend a four-year college and applied to at least one nonetheless did not end up matriculating at one. They argue that this number would have been higher if they had not provided the students with extensive help in selecting which colleges to apply to. To be conservative, I assume that only 10% of Avery and Kane's non-matriculators, 2.7% of students, who sent an additional application would be admitted to a college they could afford and thus induced to attend four years of college by doing so. Given that only 23% of low-income students who sent an additional score report sent an additional application, this implies that only 0.61% of students who sent an additional score report attended college because they did so.

To calculate the earnings gain these low-income students who were induced to attend college receive, I use Carniero et al.'s (2003) estimate that, on average, students on the margin of going to college would have lifetime earnings 51% higher as college graduates than as high school graduates. This includes college graduates' forgone earnings while in college, their lower experience at any given age than high school graduates, and the cost of college tuition. This is approximately a 10.8% return for each year of college, unsurprisingly smaller than Card's (1995) estimates of the return to a year of college that do not include forgone earnings and experience or tuition costs. I use Day and Newburger's (2002) estimate that the lifetime earnings of high school graduates from these cohorts will be approximately $1.2 million in 1999 dollars. Under these assumptions, a low-income student receives a benefit of over $3,700 from sending an additional score report through this channel. The benefit is

$$0.61\% \times \$1,200,000 \times 51\% \approx \$3,700 \tag{2.4}$$

[16]Most colleges allow low-income students to waive their application fees.

where 0.61% of low-income students who sent an additional score report were induced to attend college because of their extra score report and \$1,200,000 × 51% is the return these students get from attending college.

2.6.2 Attending a Preferred College

Dale and Krueger (2002) estimate that low-income students receive a 4% wage premium for attending a college whose students score 100 points higher on the SAT.[17] Day and Newburger (2002) estimate that the average college graduate will earn \$2.1 million in 1999 dollars over her lifetime. Under these assumptions and the assumptions detailed below, the return to sending an additional score report is approximately \$2,600. That benefit is

$$\$2,100,000 \times 4\% \times 0.44 \times 0.07 \approx \$2,600 \tag{2.5}$$

Here, \$2,100,000 × 4% is the estimated return a low-income student gets from attending a college with students who have SAT scores 100 points higher and 0.44 is the conversion factor between 100 SAT points and one (non-normalized) ACT point calculated from a concordance produced by the College Board. I estimate that, on average, students who sent an additional score report would have attended a college whose students had (non-normalized) ACT scores 0.07 points higher as a result of the cost change. For comparison, low-income students who sent an additional score report sent their scores to colleges whose students had non-normalized ACT scores 0.12 points higher as a result of the cost change.

To calculate the 0.07 figure I make four assumptions. First, I assume that score reports sent to students' most-competitive colleges translated into applications at the same rate as did other score reports: 23%. Second, to be conservative, I assume that these new applicants were only 50% as likely to be admitted to these colleges as the average applicant. Third, I assume that before the cost change, students attended the most-selective colleges they sent scores to. This is a very conservative assumption since students may not have been admitted to or even applied to these colleges. Finally, I assume that after the cost change, students who were admitted to the most-selective colleges they sent scores to attended those schools, while those who were not admitted attended the same schools

[17] As discussed in Section II, there are many studies that estimate the return to college selectivity. I use estimates from the Dale and Krueger (2002) study because it examines low-income students separately and because Dale and Krueger's methodology generally finds smaller returns to college quality than do other approaches, allowing my statements about the benefit of sending an additional score report to be more conservative.

they would have attended without the fourth free score report. While probably not every student attended the most-competitive college that admitted her, Dale and Krueger (2002) find that the vast majority of students in their highly-able sample attended the most selective college to which they were admitted.[18] Even if all students attended the most-competitive college they sent scores to before the cost change, but only half of students attended the best college they were admitted to after the cost change, the return to sending an additional score report through this channel would be over $1,250, much larger than the $6 cost change. Appendix Table 2 lays out each of the steps to calculating the 0.07 figure.

2.7 Interpretation of Behavior Change

As $6 is much smaller than the estimated return low-income students receive from sending an additional score report, much less an additional application, it seems unlikely that it could be optimal for 18% of low-income students to send an additional application as a result of a $6 cost change.

Since students only benefit from sending score reports to colleges they apply to, middle- and high-income students should treat score-sending costs and other application costs equivalently. Yet, they do not appear to do so. While applications increased by 8% in response to the $6 decrease in score-sending costs, I find that there is no relationship in the ACS data between changes in application fees and changes in the number of applications colleges received from 1993 to 2002. This lack of relationship could result from the endogeneity of application fees if colleges decrease their fees when they expect to have few applicants and raise them when they expect to have many. However, if colleges believed that changing their application fees led to large changes in the number of applications they received, colleges that go to great lengths to encourage applications by sending representatives to high schools and purchasing radio and television advertisements would likely eliminate or greatly reduce their application fees to encourage more applications.

It is not surprising that students do not optimally choose the colleges to which they apply given that it is almost impossible to do so. Students must choose one of over $2^{2,400}$ combinations of colleges to apply to, while determining the value of applying to even one combination is not

[18] The fact that students did not send score reports to these more-competitive colleges when they only received three free score reports does not indicate that they did not want to attend these colleges. It may simply reflect the fact that they thought they were unlikely to be admitted. The return to college selectivity literature suggests that these students would receive substantially higher earnings from attending these more-selective colleges.

straightforward. This value depends on the utility the student would get from attending each college, the probability that she would be admitted to each combination of the colleges in the set she applies to, and the cost of applying to the colleges. Specifically, the utility student i gets from applying to a set of colleges θ is

$$\left[\sum_{I} p_{i\varphi\phi} \sum_{j\in\varphi} \Pr(A_{j\varphi}) \times E_1[u_{ij}|A_{j\varphi}]\right] - E_1[c_{i\theta}] \tag{2.6}$$

where event

$$A_{j\varphi} = \{E_2[u_{ij}] > E_2[u_{ik}] \forall k \neq j, k \in \varphi\} \tag{2.7}$$

Here I is the set of all sets $\{\varphi, \phi\}$ that partition the set of colleges θ. In this expression, $p_{i\varphi\phi}$ is the probability that student i is admitted to all of the colleges in φ, but none in ϕ (conditional on applying), u_{ij} is the utility student i gets from attending college j, and $c_{i\theta}$ is the cost to student i of applying to the colleges in θ. The operator E_1 indicates the expectation when the student chooses where to apply while E_2 indicates the expectation when the student decides where to matriculate.

The intuition behind this expression is that with probability $p_{i\varphi\phi}$, the student is only admitted to the colleges in φ. She will then attend the college in the set φ which gives her the largest expected utility. This is college j when event $A_{j\varphi}$ occurs. When she attends college j, she will, on average, obtain the expected utility from attending that college, conditional on her having chosen to attend. This will be different from her expected utility from attending at the time she applies if she learns more about the college after she applies but before she decides where to matriculate.[19]

Expression (6) is difficult to evaluate because students face uncertainty about u_{ij}, $p_{i\varphi\phi}$, and often even $c_{i\theta}$. The utility student i gets from attending college j, u_{ij}, depends on her financial aid package (which is often not revealed until after admissions decisions), her earnings after attending college j, her earnings if she does not attend a four-year college, and the utility derived from experiences she would have at the school. Avery and Kane (2004) show that students have great difficulty estimating even part of u_{ij}. The high school seniors Avery and Kane surveyed overestimated college graduates' average earnings by 50% and tuition at local colleges by 100% to 200% on average. Twenty-five percent of the students surveyed estimated that the net present value of college was negative.

While a student may know a college's admissions rate and statistics on its entering students,

[19]Howell (2004) proposes a similar model of students' college application decisions which she estimates structurally using data from the Department of Education.

she may not know the admissions rate of applicants with her grades and test scores or how the school's admissions officers will weight her extracurricular activities, recommendation letters, and essays. It is even less likely that the student would know the strength of the correlation between different colleges' admissions decisions which is necessary to determine $p_{i\varphi\phi}$. Finally, she might not even know her cost of applying to a set of colleges because this depends on how long it takes to complete the applications.

Several behavioral explanations can explain students' large response to the fourth free score report. Thaler and Sunstein (2008) argue that when faced with uncertainty, people often follow rules of thumb. In the current context, students may interpret the ACT providing three (or four) free score reports as a signal that sending three (or four) applications is recommended. They may then use that signal as a rule of thumb about how many colleges to apply to. Under this explanation, when the cost of score-sending changed, students responded to the change in their rule of thumb, not the actual cost change. This explanation is consistent with the Madrian and Shea (2001), Choi et al. (2002), and Thaler and Sunstein (2008) results that default 401(k) plans and Medicare Part D plans have large effects on savings and health care plan choices. Deciding which financial investments and health care plan best suit an individual family is exceedingly complex. When individuals are unsure which to choose, they may follow the rule of thumb of simply sticking with the default plan.

College application guides show that many students are looking for an authority to provide a rule of thumb on how many colleges they should apply to. "How many applications are enough?" is the first frequently asked question on the College Board's website for college counselors[20] and is prominently featured in many other college guides. The College Board suggests sending five to eight applications, many more than students send on average.

An additional consideration is that the cost of the fourth score report did not just decrease by \$6, it decreased to \$0. Several recent studies have found that demand is discontinuous at a price of zero. For example, Kremer and Miguel (2007) find that the take-up of deworming drugs in Kenya decreased from 75% to 19% when students were charged \$0.30 per pill instead of receiving the drugs for free, even though deworming provides large benefits. Demand was not sensitive to changes in price once the price was above \$0. In the current context, even though the fourth score report cost \$0, sending an additional application was still costly for higher-income students because of application fees. Ariely (2008) documents, however, that people buy much more of a product

[20]See http://professionals.collegeboard.com/guidance/applications/how-many (accessed January 22, 2009).

when part of the item (such as shipping-and-handling) costs $0 than they would from an equivalent reduction in the total price of the item where each part retained a positive price.

2.8 Conclusion

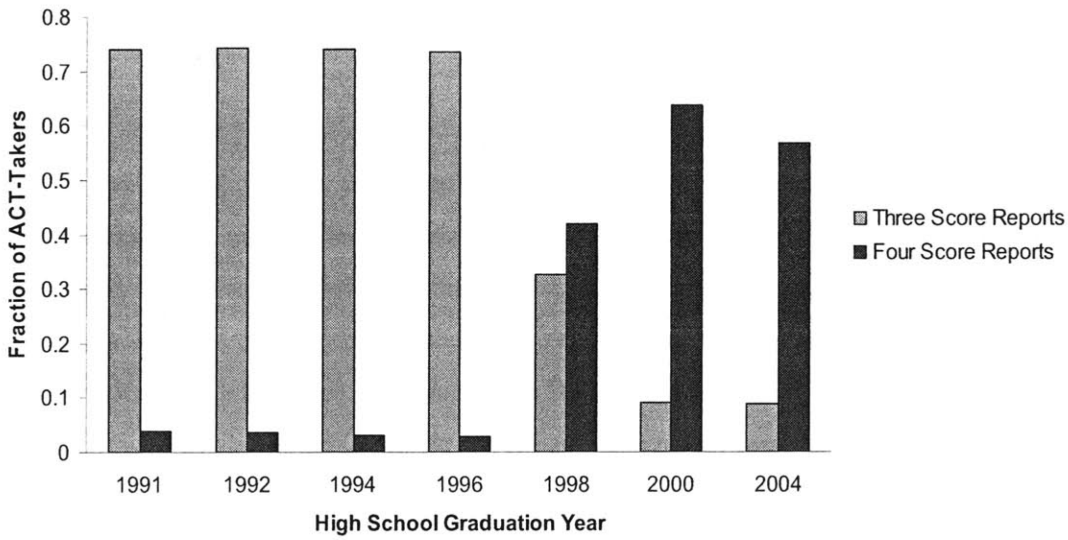
The colleges a student applies to greatly affect whether she attends college, the type of colleges she attends, and her future earnings. Yet, little is known about how students decide where to apply. This paper analyzes the effect of a small, $6 decrease in the cost of sending ACT scores to colleges on the number and selectivity of colleges to which students sent their scores. Before the fall of 1997, students taking the ACT could send their test scores to three colleges for free while each additional score report cost $6. Afterwards, students could send four score reports for free with the same $6 marginal cost for each additional report.

The paper finds that many students sent an additional score report and application as a result of the $6 cost change. Students also sent their scores to a wider range of colleges, sending scores to colleges that were both more- and less-selective than any they would have otherwise sent scores to. I estimate that by increasing the probability that she attended college and attended a selective college, sending an additional score report increased each low-income student's expected future earnings by thousands of dollars.

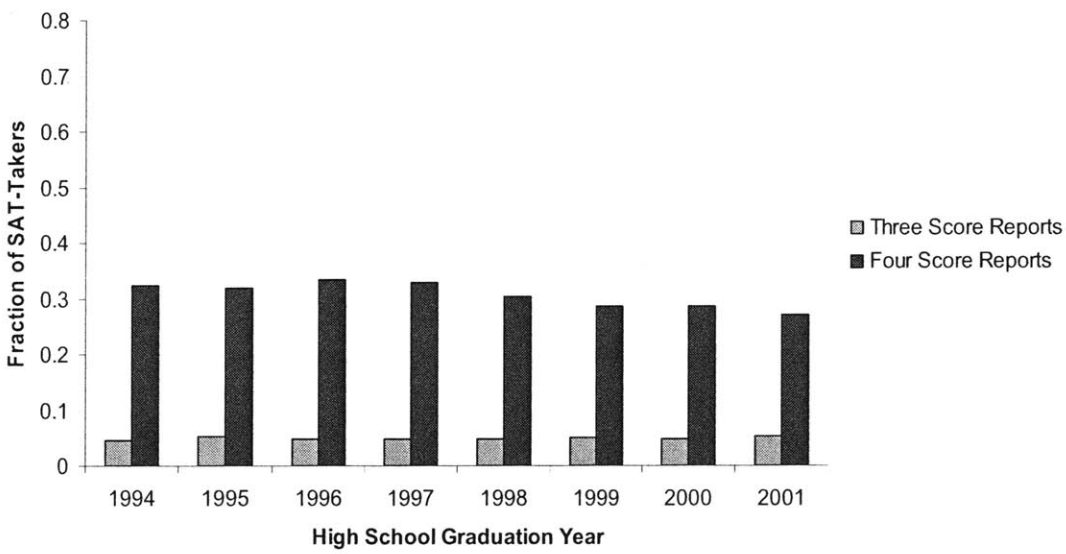
I provide evidence that the large increase in the number of applications sent in response to this $6 cost change is inconsistent with optimal decision-making. Moreover, I show that optimally choosing which colleges to apply to is almost impossible because students' maximization problem is so difficult to solve. Faced with great uncertainty, students may rely on rules of thumb: in this case interpreting the ACT's provision of three (or four) score reports as a recommendation about how many scores to send. Under this explanation, students responded to the change in their rule of thumb instead of the actual change in the cost of sending score reports. Given students' apparent reliance on rules of thumb, providing them with rules of thumb based on data as opposed to the pricing structure of the ACT could lead to large changes in application behavior, facilitating higher college attendance and better student-college matches.

Figure 1. Number of Scores Sent by High School Graduation Year

1a: Students Who Took the ACT

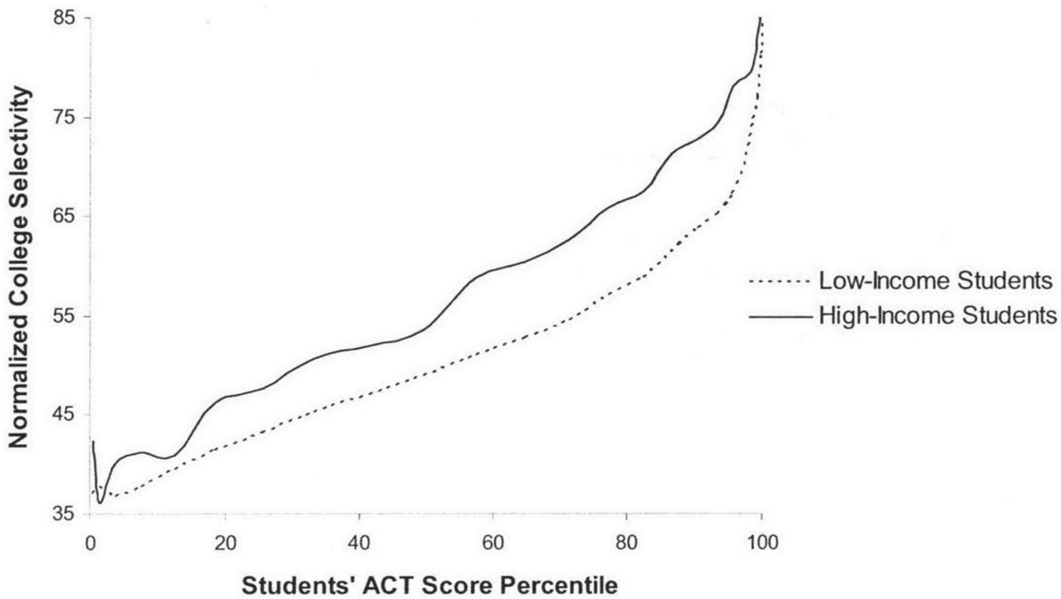


1b: Students Who Took the SAT



Notes: The bars indicate the fraction of each high school class that sent either exactly three or exactly four score reports. The analysis is not limited to students who sent at least one score report. Data in Panel A come from the ACT Database and data in Panel B come from a database of SAT-takers produced by the College Board.

Figure 2. College Selectivity by Family Income in the Class of 1996



Notes: The y-axis measures the average selectivity of colleges that students in the high school class of 1996 sent scores to. The x-axis measures students' own ACT score in percentile terms. Low-income students have family incomes less than $36,000 per year, while high-income students have family incomes greater than $80,000 per year. The data come from the ACT database and the American College Survey.

Table 1. Number of Scores Sent by Family Income in the Class of 1996

	(1)	(2)	(3)
Middle-Income	0.107	0.047	0.039
	(0.005)	(0.005)	(0.005)
High-Income	0.238	0.160	0.134
	(0.013)	(0.013)	(0.012)
Background Characteristics	No	Yes	Yes
High School Fixed Effects	No	No	Yes

Notes: Each column presents results from a separate regression of the number of score reports sent on income dummies. The omitted dummy is the low-income indicator. All regressions are limited to students in the class of 1996 who sent at least one score report. Standard errors, clustered at the state level, are in parentheses. Low-income students have family incomes less than $36,000 per year, middle-income students have family incomes between $36,000 and $80,000, and high-income students have family incomes above $80,000. The controls included in "background characteristics" are listed in footnote seven. Data come from the ACT database.

Table 2. College Selectivity by Family Income in the Class of 1996

A. Most-Selective College

	All Students			By ACT Quartile				
	(1)	(2)	(3)	Bottom Quartile	2nd Quartile	3rd Quartile	Top Quartile	Top 10%
Middle-Income	7.83	1.42	0.75	0.85	0.31	0.96	0.68	1.33
	(0.70)	(0.20)	(0.16)	(0.37)	(0.25)	(0.29)	(0.35)	(0.42)
High-Income	12.03	4.83	3.09	1.53	2.45	3.58	3.34	3.42
	(0.94)	(0.31)	(0.25)	(0.69)	(0.48)	(0.43)	(0.53)	(0.91)

B. Least-Selective College

	All Students			By ACT Quartile				
	(1)	(2)	(3)	Bottom Quartile	2nd Quartile	3rd Quartile	Top Quartile	Top 10%
Middle-Income	5.46	1.09	0.32	0.39	0.20	0.30	0.33	1.13
	(0.68)	(0.25)	(0.15)	(0.32)	(0.27)	(0.30)	(0.30)	(0.59)
High-Income	8.59	3.99	2.19	0.11	1.31	2.74	3.27	3.84
	(0.78)	(0.39)	(0.32)	(0.70)	(0.56)	(0.49)	(0.58)	(0.99)
Background Characteristics	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High School Fixed Effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Each column presents results from a separate regression where the dependent variable is the largest (Panel A) or smallest (Panel B) normalized 25th percentile ACT score of freshmen at the colleges a student sent scores to. The dependent variables are regressed on income dummies where the omitted dummy is the low-income indicator. All regressions are limited to students from the class of 1996 while the regressions in the last five columns are additionally limited to students in the parts of the ACT distribution indicated. Standard errors, clustered at the state level, are in parentheses. Low-income students have family incomes less than $36,000 per year, middle-income students have family incomes between $36,000 and $80,000, and high-income students have family incomes above $80,000. The controls included in "background characteristics" are listed in footnote seven. Data come from the ACT database and the American College Survey.

Table 3. Change in the Number of Scores Sent by Family Income

	(1)	(2)	(3)	(4)	(5)
Low-Income	-0.049	-0.058	-0.084	-0.110	-0.059
	(0.007)	(0.007)	(0.005)	(0.011)	(0.009)
Class of 1998	0.429	0.492	0.388	0.505	0.459
	(0.027)	(0.022)	(0.028)	(0.031)	(0.022)
Low-Income × Class of 1998			0.114	0.084	0.098
			(0.011)	(0.011)	(0.010)
Post-1998	0.612	0.791	0.587	0.808	0.780
	(0.037)	(0.013)	(0.037)	(0.035)	(0.015)
Low-Income × Post-1998			0.069	0.006	0.029
			(0.011)	(0.021)	(0.012)
Time Trends	No	Yes	No	Yes	Yes
Background Characteristics	No	Yes	No	No	Yes
High School Fixed Effects	No	Yes	No	No	Yes

Notes: Each column presents results from a separate regression of the number of score reports sent on a low-income dummy, an indicator for graduating high school in 1998, and an indicator for graduating after 1998. In the last three columns, interactions of these two graduation-year indicators with the low-income dummy are included. All regressions are limited to students who sent at least one score report. Standard errors, clustered at the state level, are in parentheses. Low-income students have family incomes less than $36,000 per year. The controls included in "background characteristics" are listed in footnote seven. The time trends included are a linear time trend and a linear time trend interacted with the low-income indicator. Data come from the ACT database.

Table 4. Change in the Number of Applications Sent by Standardized Test Taken

	ACT		SAT	
	Everyone	Low-Income	Everyone	Low-Income
Average Applications Sent by Class Entering College in 2003-04	2.93	2.92	3.27	3.19
	(0.04)	(0.08)	(0.05)	(0.08)
Average Applications Sent by Class Entering College in 1995-96	2.60	2.54	3.14	2.98
	(0.09)	(0.10)	(0.08)	(0.11)
Difference	0.33	0.39	0.13	0.21
	(0.10)	(0.13)	(0.09)	(0.13)

Notes: Low-income students are dependent students with family incomes less than $36,000. Standard errors are in parentheses. The data come from the Beginning Postsecondary Students Longitudinal Studies of 1996 and 2004.

Table 5. Changes in College Selectivity by Family Income

	A. Most-Selective College				
	(1)	(2)	(3)	(4)	(5)
Low-Income	-7.62	-1.84	-7.98	-8.20	-1.75
	(0.62)	(0.17)	(0.60)	(0.58)	(0.16)
Class of 1998	1.61	2.70	1.23	2.56	2.37
	(0.20)	(0.22)	(0.19)	(0.18)	(0.20)
Low-Income × Class of 1998			1.09	0.91	1.01
			(0.24)	(0.21)	(0.20)
Post-1998	1.50	4.44	1.24	3.76	4.13
	(0.28)	(0.25)	(0.25)	(0.21)	(0.25)
Low-Income × Post-1998			0.75	0.31	0.87
			(0.33)	(0.26)	(0.22)
	B. Least-Selective College				
	(1)	(2)	(3)	(4)	(5)
Low-Income	-5.73	-1.29	-5.87	-5.97	-1.21
	(0.54)	(0.14)	(0.54)	(0.56)	(0.13)
Class of 1998	-2.52	-2.21	-2.55	-2.12	-2.17
	(0.23)	(0.13)	(0.22)	(0.18)	(0.14)
Low-Income × Class of 1998			0.03	-0.09	-0.16
			(0.27)	(0.20)	(0.19)
Post-1998	-4.48	-3.45	-4.63	-3.82	-3.58
	(0.43)	(0.20)	(0.38)	(0.23)	(0.20)
Low-Income × Post-1998			0.49	0.24	0.38
			(0.35)	(0.22)	(0.21)
Time Trends	No	Yes	No	Yes	Yes
Background Characteristics	No	Yes	No	No	Yes
High School Fixed Effects	No	Yes	No	No	Yes

Notes: Each column presents results from a separate regression where the dependent variable is the largest (Panel A) or smallest (Panel B) normalized 25th percentile ACT score of freshmen at the colleges a student sent scores to. The dependent variables are regressed on a low-income dummy, an indicator for graduating high school in 1998, and an indicator for graduating after 1998. In the last three columns, interactions of these two graduation-year indicators with the low-income dummy are included. Standard errors, clustered at the state level, are in parentheses. Low-income students have family incomes less than $36,000 per year. The controls included in "background characteristics" are listed in footnote seven. The time trends included are a linear time trend and a linear time trend interacted with the low-income indicator. Data come from the ACT database and the American College Survey.

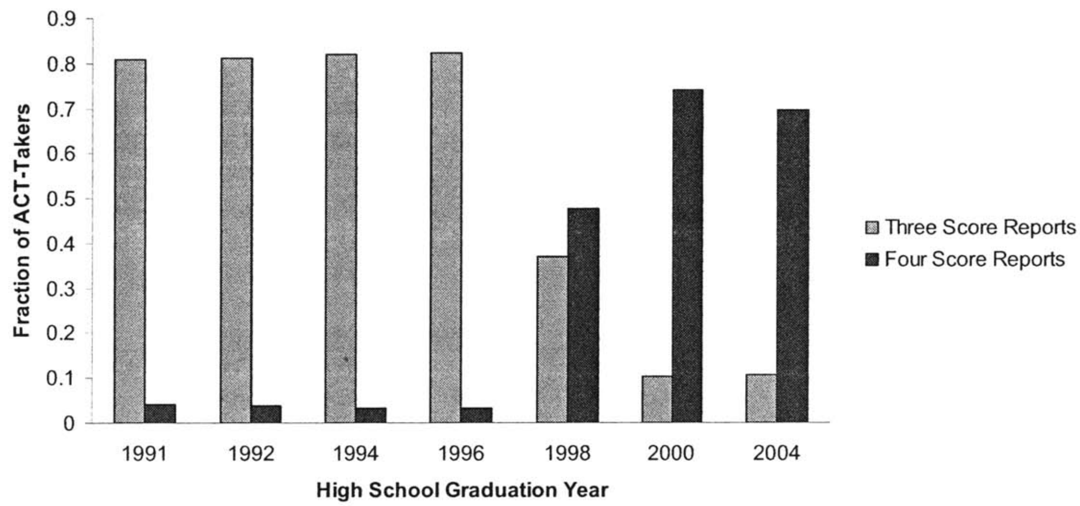
Table 6. Changes in College Selectivity by Family Income and Ability

	A. Most-Selective College				
	Bottom Quartile	2nd Quartile	3rd Quartile	Top Quartile	Top 10%
Low-Income	-1.17	-1.54	-1.78	-1.76	-1.00
	(0.25)	(0.19)	(0.19)	(0.21)	(0.25)
Class of 1998	3.26	3.15	2.26	1.58	1.03
	(0.26)	(0.26)	(0.22)	(0.21)	(0.24)
Low-Income × Class of 1998	1.53	0.27	0.29	0.50	0.75
	(0.38)	(0.27)	(0.34)	(0.30)	(0.51)
Post-1998	4.66	4.83	4.09	3.49	2.64
	(0.37)	(0.33)	(0.29)	(0.26)	(0.28)
Low-Income × Post-1998	1.19	0.18	0.53	0.00	0.88
	(0.37)	(0.38)	(0.42)	(0.35)	(0.51)
	B. Least-Selective College				
	Bottom Quartile	2nd Quartile	3rd Quartile	Top Quartile	Top 10%
Low-Income	-0.48	-0.71	-0.95	-1.51	-1.73
	(0.18)	(0.17)	(0.15)	(0.19)	(0.29)
Class of 1998	-1.78	-2.17	-2.29	-2.08	-1.47
	(0.25)	(0.19)	(0.22)	(0.20)	(0.34)
Low-Income × Class of 1998	-0.33	-0.06	-0.49	0.02	-0.42
	(0.28)	(0.27)	(0.32)	(0.49)	(0.63)
Post-1998	-3.04	-3.12	-3.90	-3.65	-2.89
	(0.27)	(0.27)	(0.32)	(0.28)	(0.49)
Low-Income × Post-1998	-0.07	-0.12	0.41	-0.12	0.19
	(0.38)	(0.35)	(0.41)	(0.47)	(0.94)
Time Trends	Yes	Yes	Yes	Yes	Yes
Background Characteristics	Yes	Yes	Yes	Yes	Yes
High School Fixed Effects	Yes	Yes	Yes	Yes	Yes

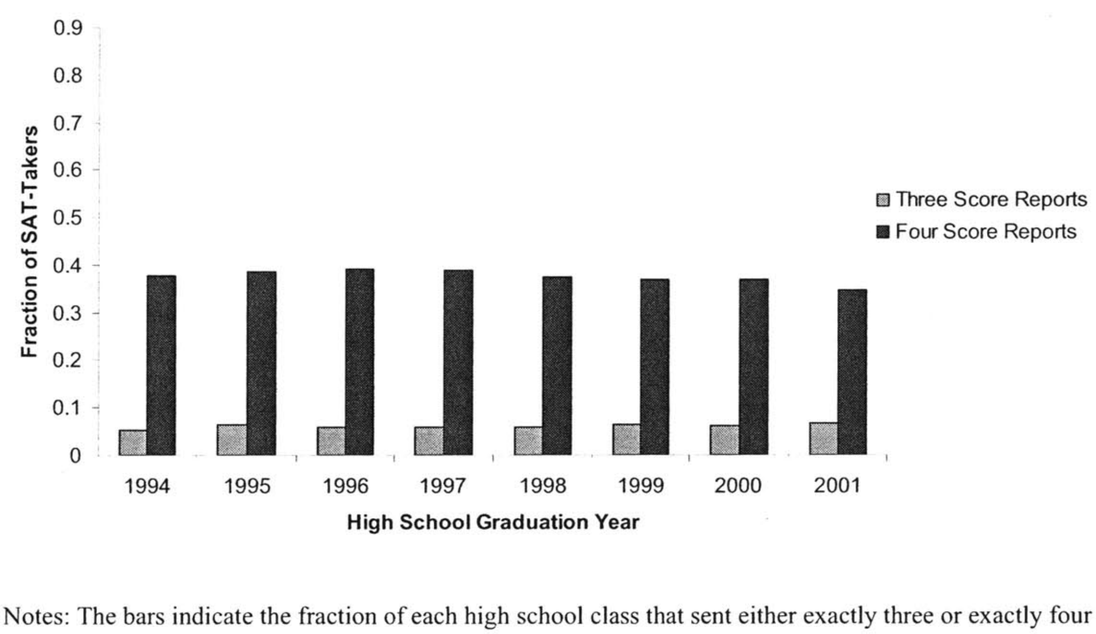
Notes: Each column presents results from a separate regression where the dependent variable is the largest (Panel A) or smallest (Panel B) normalized 25th percentile ACT score of freshmen at the colleges a student sent scores to. The dependent variables are regressed on a low-income dummy, an indicator for graduating high school in 1998, an indicator for graduating after 1998, and the interactions of these two graduation-year indicators with the low-income dummy. Standard errors, clustered at the state level, are in parentheses. The regressions are limited to students in the parts of the ACT distribution indicated by the column headings. Low-income students have family incomes less than $36,000 per year. The controls included in "background characteristics" are listed in footnote seven. The time trends included are a linear time trend and a linear time trend interacted with the low-income indicator. Data come from the ACT database and the American College Survey.

Appendix Figure 1. Number of Scores Sent by High School Graduation Year: Excluding Students Who Sent Zero Score Reports

Appendix 1a: Students Who Took the ACT



Appendix 1b: Students Who Took the SAT



Notes: The bars indicate the fraction of each high school class that sent either exactly three or exactly four score reports. The analysis is limited to students who sent at least one score report. Data in Panel A come from the ACT Database and data in Panel B come from a database of SAT-takers produced by the College Board.

Appendix Table 1. Change in the Number of Scores Sent by Family Income: Including Students Who Sent Zero Score Reports

	(1)	(2)	(3)	(4)	(5)
Low-Income	0.149	-0.065	0.075	0.016	-0.061
	(0.022)	(0.006)	(0.017)	(0.019)	(0.007)
Class of 1998	0.335	0.453	0.286	0.504	0.423
	(0.028)	(0.022)	(0.028)	(0.022)	(0.024)
Low-Income × Class of 1998			0.133	0.053	0.093
			(0.013)	(0.012)	(0.016)
Post-1998	0.357	0.689	0.294	0.712	0.667
	(0.036)	(0.015)	(0.036)	(0.024)	(0.018)
Low-Income × Post-1998			0.194	0.027	0.062
			(0.021)	(0.015)	(0.013)
Time Trends	No	Yes	No	Yes	Yes
Background Characteristics	No	Yes	No	No	Yes
High School Fixed Effects	No	Yes	No	No	Yes

Notes: Each column presents results from a separate regression of the number of score reports sent on a low-income dummy, an indicator for graduating high school in 1998, and an indicator for graduating after 1998. In the last three columns, interactions of these two graduation-year indicators with the low-income dummy are included. The regressions are not limited to students who sent at least one score report. Standard errors, clustered at the state level, are in parentheses. Low-income students have family incomes less than $36,000 per year. The controls included in "background characteristics" are listed in footnote seven. The time trends included are a linear time trend and a linear time trend interacted with the low-income indicator. Data come from the ACT database.

Appendix Table 2. Procedure for Determining the Average Increase in Selectivity of Colleges Students Attended

Step Number	Description	Output	Assumption Used[a]
1	Restricting the sample to low-income students, regress dummies for the student's most-competitive college having each level of selectivity on all of the independent variables in Equation (2) including high school fixed effects.	The coefficient on the post-1998 dummy in each regression equals the change in the probability that a student's most-competitive college had this level of selectivity as a result of the cost change.	
2	Multiply the positive coefficients from Step 1 by 23%, the probability that these additional score reports turned into applications.	Each product is the increase in the probability that the most-competitive college a student applied to had this level of selectivity as a result of the cost change.	1
3	Multiply each product from Step 2 by 50% times the admission rate of students at colleges with that level of selectivity in 2000. (The admission rates are calculated from the ACS.)	Each product is the increase in the probability that a student was admitted to the most-competitive college she sent scores to and that college had the indicated level of selectivity.	2
4	Sum the probabilities from Step 3. Call the sum Z.	This gives the total mass of students who sent scores to more-selective colleges than they otherwise would have and who were admitted to these colleges and thus (by Assumption 4) attended.	4
5	Compute a weighted average of college selectivity levels using the probabilities from Step 3 as weights.	This gives the average selectivity of the more-competitive colleges the students attended.	4
6	Consider the selectivity levels that had negative coefficients in Step 1. Starting from the highest selectivity level, sum the absolute value of the probabilities until reaching Z. Let the amount that each selectivity level contributed towards this sum be weights and form a weighted average of these selectivity levels.	This gives the average selectivity of the colleges that the students who were admitted to more-selective colleges would have sent scores to and thus attended (by Assumption 3) without the cost change.	3 If there had been no cost change, the students who attended more-competitive colleges because of the cost change would have attended more-selective colleges than students who sent score reports to more-selective colleges because of the cost change but did not end up attending these colleges.[b]

7	Subtract the average in Step 6 from the average in Step 5.	This is the average change in selectivity that students who attended more-selective colleges as a result of the cost change experienced.	
8	Multiply the result from Step 7 by the result from Step 4.	This gives the average change in selectivity that a low-income student experienced as a result of some students attending more-competitive colleges.	
9	Divide the result from Step 8 by 0.81, the fraction of low-income students who sent an additional score report as a result of the cost change.	This gives the average change in selectivity that a low-income student who sent an additional score report experienced.	Conditional on all the controls and the time trend, students who did not send an additional score report did not change their application behavior as a result of the cost change.

Notes: [a] The assumption numbers refer to those in Section VI of the text.
[b] This assumption is very conservative. Any alternative assumption would lead to larger estimates of the increase in the selectivity of colleges low-income students attended.

Chapter 3

Taking a Chance on College: Is the Tennessee Education Lottery Scholarship Program a Winner?

3.1 Introduction

Many policies implemented to improve American elementary and secondary education provide incentives to teachers and schools, not students. Those that do provide incentives to students typically do so by punishing students who perform poorly instead of rewarding those who do well. However, since 1991, more than a dozen states have enacted scholarship programs that award merit aid at in-state colleges to large fractions of the states' high school graduates for students' high school GPAs, scores on a standardized test, or both. This paper analyzes the effect of one of these programs, the Tennessee Education Lottery Scholarship (TELS), on high school achievement as measured by the ACT.

Approximately a quarter of high school seniors live in states offering these scholarship programs and these programs represent a large expense – Tennessee's program cost \$68 million for just one class in its first year – understanding the effects of these programs is important in its own right. To this end, this paper also analyzes the effect of Tennessee's scholarship program on students' college preferences.

The main prong of the TELS, the HOPE Scholarship, rewarded Tennessee residents who (1) scored at least 19 on the ACT (or 890 on the SAT) or (2) had a final high school GPA of 3.0

or higher, including a 3.0 unweighted GPA in all 20 credits of college core and university track classes. Winners received a renewable $3,000 per year to attend any four-year Tennessee college or a renewable $1,500 per year to attend any two-year college in the state. Using micro data on students' ACT scores, the colleges to which students sent their scores, and a rich set of background characteristics, I analyze the effect of the TELS on Tennessee students' ACT scores.

This scholarship increases the return to scoring 19 or higher on the ACT for students who were unsure of their ability to qualify for the scholarship through their GPAs and who were considering attending an in-state college. Because it does not strongly affect the return to increasing the ACT score for students who would have already scored 19 or higher or for students who cannot reach 19, I expect to (and do) find that students increased their scores from below 19 to 19 or just above, but there was very little change in the rest of the test score distribution.

Secondly, I analyze the effect of the TELS on students' college preferences as measured by their stated preferences and the colleges to which they sent their ACT scores. The TELS decreases the cost of attending in-state relative to out-of-state colleges and four-year in-state relative to two-year in-state colleges. I find no effect of the TELS on college preferences. While there were small changes in preferences in Tennessee in 2004, I show that the changes occurred primarily for students ineligible for the TELS and thus are extremely unlikely to have resulted from the scholarship.

The paper is organized as follows: Section 3.2 provides background information and discusses the relevant literature. Section 3.3 describes the dataset used; Sections 3.4 and 3.5 present the empirical results on changes in the test-score distribution and students' college preferences, respectively. Section 3.6 concludes.

3.2 Background

3.2.1 Tennessee Education Lottery Scholarships

The TELS, funded by a newly-created state lottery, first began awarding scholarships in the fall of 2004 to college freshmen from the high school class of 2004 and college sophomores from the high school class of 2003.[1] Scholarships were available to Tennessee residents who enrolled in Tennessee post-secondary institutions and required no application except for the FASFA, which was required

[1] The TELS encompassed five programs, the largest of which was the HOPE Scholarship which accounted for over 99% of scholarship winners attending a two- or four-year college. The other programs provided supplements to the HOPE Scholarship for low-income and high-achieving students, provided a smaller award to low-income students who very slightly missed HOPE eligibility, and provided aid to students entering technical or trade schools.

of all scholarship winners whether or not they were likely to be eligible for need-based aid.

The TELS is open to a larger percentage of students than many of the other state merit scholarship programs – approximately 65 percent of high school graduates – and, comparatively, is especially inclusive of African-Americans and low-income students (Ness and Noland 2004). It is the only program that allows students to qualify through their performance on a standardized test or through their grades.

Table 1 shows the size of the HOPE Scholarship. The state awarded 23,287 scholarships from the class of 2004, costing almost $68 million. Of those who qualified for and accepted the scholarship, 79 percent attended four-year colleges and the vast majority of these attended public colleges.

3.2.2 Tennessee Postsecondary Education

Even before the TELS was enacted, 85 percent of Tennesseans going straight to college attended one of the state's nine four-year public universities, 13 two-year public colleges, 35 independent institutions, or 27 technology centers.[2]

The four-year public universities were competitive enough that a student on the margin of HOPE eligibility would have found peers of similar ability in the university system, but Tennessee's "best and brightest" typically would not. Approximately 60 percent of 2004 Tennessee freshmen at Tennessee four-year public colleges received lottery scholarships. Academically, the four-year public colleges range from the historically black Tennessee State University, at which the middle 50 percent of students score between 16 and 21 on the ACT (equivalent to 760 to 1010 on the SAT) to the flagship, the University of Tennessee-Knoxville, at which the middle 50 percent of students score 20 to 27 on the ACT (equivalent to 940 to 1230 on the SAT).

While the size of the individual HOPE awards was comparable to many other states' scholarship programs, students planning to attend a public university could not have expected the HOPE Scholarship to cover their entire cost of attendance. In the year before the TELS was implemented, tuition and fees alone of the public four-year universities ranged from $500 to $1,500 more than the value of the HOPE Scholarship. For two-year colleges tuition and fees ranged from $550 to $600 more than the value of the scholarship.

[2] The statistics in this section are derived from the Statistical Abstract of Tennessee Higher Education: 2003-04, Dorie Turner's article "99% of UT Freshmen Qualify for Lottery Aid," the American College Survey, and personal correspondence with Robert Anderson at the Tennessee Higher Education Commission.

3.2.3 Research on Merit Aid in Other States

Many papers analyze the effects of broad-based merit scholarship programs on student behavior. Dynarski (2004) analyzes programs in seven Southern states (not including Tennessee's) using CPS data and finds large effects on college matriculation. In the aggregate, these programs increased the probability that students from these states would enroll in college by 4.7 percentage points, primarily by increasing the probability of enrolling at a public college.

Cornwell, Lee, and Mustard (2005 and 2006), Cornwell and Mustard (2006), and Cornwell, Mustard, and Sridhar (2006) find that the Georgia HOPE program increased students' desire to attend Georgia colleges, particularly elite colleges. There was a significant increase in the number of students attending Georgia colleges after the scholarship was implemented, specifically at four-year public and private Georgia colleges and HBCUs, while the acceptance rates of Georgia colleges and the yield rates of elite Georgia colleges decreased relative to control schools. They also find that to retain HOPE funding while in college, in-state University of Georgia students enrolled in fewer credits overall, withdrew from more classes, took fewer math and sciences classes, and switched to easier majors than their out-of-state peers.

Less attention has been paid to these scholarships' effects on high school achievement. However, Henry and Rubenstein (2002) do indirectly show that the Georgia HOPE program increased Georgia high school achievement. They argue that the HOPE program increased high school grades among Georgia students entering Georgia public colleges, but SAT scores did not decrease relative to grades for this group, so the increase in grades must have been a result of improved achievement, not grade inflation.

3.3 Data Description

The data for this analysis come from a unique set of individual records assembled from the administrative files of the ACT Corporation. The data include a broad cross section of observations on students who took the ACT and planned to graduate high school in 1996, 1998, 2000, and 2004: one out of every four Caucasians, one out of every two minorities, and every student who either listed their race as "multiracial" or "other" or failed to provide a race. I limit the sample to the 24 states[3]

[3]These are Alabama, Arkansas, Idaho, Illinois, Iowa, Kansas, Kentucky, Louisiana, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Mexico, North Dakota, Ohio, Oklahoma, South Dakota, Tennessee, Utah, Wisconsin, and Wyoming.

in which the flagship university received more ACT than SAT scores, because the students who take the ACT in states where the ACT is not the primary college entrance exam are not expected to be representative of the states' potential college-goers. In the analysis of college preferences, I use only data from 1998, 2000, and 2004 since students graduating high school in 1996 could not send complimentary score reports to as many schools as in later years. Finally, in my analysis of score-sending, I omit the 13.6 percent of students who did not send their scores to any colleges.

Considering only these 24 states, there is still a large sample: 997, 346 records for the four years and over 225,000 in each year. The sample from Tennessee is large – there are 58,595 observations in total, with over 14,400 individuals in each year – and representative of the state's college-goers – over 87 percent of the state's high school seniors took the ACT in 2004.

I observe each student's composite ACT score the last time she took the exam,[4] stated preferences regarding aspects of her desired college, demographic and other background information, and up to six colleges to which she sent her score.[5] Students indicate the colleges they'd like their scores sent to when they register for the test[6] and indicate the state, number of years, institutional control, and maximum tuition of their desired college on test day. The background information is very detailed and includes gender, race, family income, classes taken, extracurricular activities, and out-of-class accomplishments.

The analysis in this paper differs from studies that analyze improvements in high school grades, which may confound changes in student achievement and incentives for grade inflation, because the ACT is an objective, nationally-administered test. The extremely rich background information allows me to conclude that my results are not due to a changing pool of test-takers. However, I cannot conclude that the increases in ACT scores reflect human capital accumulation. Students could have increased their score as a result of developing ACT-specific human capital, getting more sleep the night before the exam, or exploiting the randomness in test questions by retaking the test until receiving their desired score. I cannot rule out the first two explanations, but I do use data on the traditional gains from retesting to show that the effect is too large to plausibly be generated by increased retesting alone.

[4] Despite my best efforts to procure data on how many times students took the ACT, their scores on previous test sittings, and even aggregate information on retesting rates, I was only able to obtain a student's score the last time she took the ACT.

[5] Observing only six colleges to which students sent their scores does not practically limit my knowledge of student preferences as over 98% of students who sent test scores sent five or fewer.

[6] Students could send four score reports for free, with each additional report costing \$6 in 1998 and 2000 and \$7 in 2004.

Card and Krueger (2005) suggest that score-sending data is a good proxy for application data. They find that the number of SAT scores sent to a particular college is very highly correlated with the number of applications it receives. Using score-sending data to measure college preferences has become common in the literature.[7] However, additionally analyzing students' stated college preferences allows me to detect any changes in preferences over colleges within their application portfolios that would not be evident from examining score-sending or application data alone.

3.4 Impact of the TELS on ACT Score

Tennessee students increased their ACT scores from below 19 to 19, 20, and 21 as a result of the TELS. This is evident in graphically comparing the distribution of test scores in Tennessee from 2004 with distributions in earlier years and in difference-in-difference regressions. The results are incredibly robust.

3.4.1 Graphical Analysis

Figures 1a, 1b, 1c, and 1d show the probability and cumulative distributions of test-scores in Tennessee and comparison states for each year. They show the 2004 distribution marks a clear departure from the other years in Tennessee, but not in the rest of the country.

The first analytic question is whether the change in the distribution of test scores captures changes in the performance of test-takers or changes in the composition of the pool of test-takers. To control for differential selection into test-taking in the different years, I use a procedure developed by Dinardo, Fortin, and Lemieux (DFL), reweighting the 1996, 1998, and 2004 distributions so that they represent the counterfactual test-score distributions that would have prevailed if the same pool of students had taken the ACT in those years as in 2000.[8] I construct the counterfactual distribution for each year separately, pooling the observations from 2000 and that year, running a probit regression of the year of the observation (2000 = 1) on control variables and reweighting that year's data by the ratio of the fitted value to the fitted value's additive inverse. Appendix A has a mathematical explanation of the reweighting procedure.

I use many control variables in addition to state and year dummies, including information

[7]For examples, see Card and Krueger (2005), Abraham and Clark (2006), Pope and Pope (2006), Long (2004), and Pallais and Turner (2006).

[8]This is very similar to the estimator proposed by Barsky et al. (2002) in their exploration of the black-white wealth gap.

on the students' academic and extracurricular activities and dummies for whether each of the control variables is missing.[9] Because students' academic and extracurricular records are potentially affected by the TELS, I reperform the analysis using only fully exogenous variables as controls. The results are very similar.

Figures 2a through 2d show how this reweighting procedure changes the test-score distributions for Tennessee and the rest of the country in 2004, while figures 3a through 3d are replications of figures 1a through 1d using the reweighted distributions. The graphs show that the change in the score distribution in 2004 was a result of actual score improvement and not solely a result of differential selection into test-taking. Figures 3a and 3b show that the only change in the Tennessee distribution from previous years was the change theory predicts: a shifting of mass from below 19 to 19 or just above. There is almost no change in the distribution of scores above 21.

To quantify the differences in the test-score distributions, I use Kullback and Leibler's (1951) divergence function:

$$D_{KL}(P||Q) = \sum_i P(i) \times \log\left(\frac{P(i)}{Q(i)}\right) \tag{3.1}$$

to calculate the divergence of the 2000 distribution (P) from the other years' actual distributions, reweighted distributions, and distributions reweighted using only exogenous controls (Q). The results are presented in Table 2. For states other than Tennessee, the divergences are small and fairly similar across years. For Tennessee, the divergences of the 2004 population from the 2000 population are more than double the divergences between 2000 and any other year.

3.4.2 Regression Analysis

Using an indicator variable for whether the student scored 19 or higher on the ACT as the dependent variable (yist), I estimate difference-in-difference regressions comparing the change in ACT scores in Tennessee in 2004 to the change in scores in other states. I estimate the effect using both an

[9] Background variables included are: income dummies, race dummies, a dummy for U.S. citizenship, an indicator for English being the primary language spoken in the home, the number of siblings the student has under the age of 21, the size of the community the student lives in, and the student's gender. The variables on the student's academic career are whether she attends a private high school, is on a college preparatory track, has any college credit, the number of years of English and math classes she has taken, and dummies for whether she has taken honors English or math. Also included are dummy variables about the student's extracurricular activities that record whether she was ever elected to a student office, worked on the staff of a school paper or yearbook, earned a varsity letter for sports participation, and held a regular part-time job.

OLS and probit specification, estimating equations (3.2) and (3.3) respectively:

$$y_{ist} = \beta_0 + \beta_1 TELS_{st} + \beta_2 X_{ist} + \delta_s + \delta_t + \varepsilon_{ist} \quad (3.2)$$

$$y_{ist} = \Phi(\beta_0 + \beta_1 TELS_{st} + \beta_2 X_{ist} + \delta_s + \delta_t) + \varepsilon_{ist} \quad (3.3)$$

Here $TELS_{st}$ is an indicator for whether the student lived in a state and year (Tennessee in 2004) when the TELS was in effect and Φ represents the standard normal cumulative distribution function. The coefficient of interest in both cases is β_1. The terms δ_s and δ_t are state and year fixed effects respectively and ε_{ist} is an idiosyncratic error term. I consider several different specifications for X_{ist}, the control variables. Standard errors are clustered at the state-year level.

Results from estimating equations (3.2) and (3.3) are presented in Tables 3 and 4 respectively. In Table 4, the first row in each cell contains the average of the marginal effects for each observation in the sample while the second and third rows in each cell are the estimated coefficient and standard error. The impact of the TELS is clear: every coefficient in the tables is positive and significant at the 5 percent level and over 95 percent are significant at the 1 percent level. As more controls are added, the measured impact of the program decreases, however it still remains large and significant. The coefficient also remains large and significant after the many robustness checks described in Appendix B. The results from these robustness checks are displayed in Appendix Table 1.

For the whole population, the OLS coefficient is 0.061 when all controls are included and 0.079 when only the strictly exogenous controls are included, signifying that 6.1 percent and 7.9 percent of Tennessee students increased their score to 19 or higher as a result of the TELS, respectively. These estimates are similar in magnitude to the actual 7.2 percent increase in the number of Tennessee students who scored 19 or higher before and after the TELS was implemented.

3.4.3 Assessing the Magnitude of the Effect

A 6.1 percentage point increase in the number of students scoring 19 or higher is a large increase in test scores.[10] Since only 42.1 percent of Tennessee test-takers scored below 19 in 2000, this

[10]The magnitude of this effect is roughly comparable to estimated magnitudes of effects of the Georgia HOPE program. Henry and Rubenstein (2002) find the percentage of Georgia high school students earning a GPA of "B" or better increased by 2.9 percentage points and Dynarski (2004) finds that college matriculation of Georgia students increased by 4.7 percentage points as a result of the Georgia HOPE program. These effects are approximately $\frac{1}{2}$ and $\frac{3}{4}$ of the size of the effect of the TELS respectively. Cornwell, Lee, and Mustard (2005) find that GPAs among University of Georgia freshmen increased by 0.18 standard deviations which is 50% larger than the 0.12 standard deviation increase in ACT scores caused by the TELS.

implies that one out of every seven Tennessee students who could have increased their score to 19 or higher did so. This may even understate the change in achievement as some students may have increased their scores but not all the way to 19 and some students who would have scored 19 or higher even without the TELS worked to increase their scores because of uncertainty over how they would perform.

Further analysis suggests that despite the decrease in mass at scores as low as 12 and 13 in Figure 3a, only students who would have scored 15 to 18 without the TELS increased their scores to 19 or higher. Some students who would have scored below 15 without the TELS did increase their scores, but fell short of 19.

To determine whether students who would have earned low ACT scores increased their scores to 19 or higher, I first predict the ACT score students would have received without the TELS using data from 1996, 1998, and 2000 and all of the control variables. Then I estimate equation (3.2) separately for students predicted to have different ACT scores. Table 5 shows that the coefficients for students predicted to score below 15 are small and insignificant, but students predicted to score 15 to 17 and 17 to 19 saw 5.3 and 7.6 percentage point increases in the probability of scoring 19 or higher respectively. Column 3, which attempts to account for the prediction error, suggests that 6.7 percent and 12.6 percent of students who would have scored in these ranges increased their scores above the threshold.

Though no information is available on changes in retesting rates, these effects are too large to be a function of students simply retaking the test. An ACT Research Report using data from 1993 (Andrews and Ziomek 1998) found that 36 percent of students retook the ACT nationally. Eleven percent of those scoring 15 or 16 and 43 percent of those scoring 17 or 18 increased their score to 19 or higher on their second attempt. A back-of-the-envelope calculation shows that to get effects as large as seen here, there would need to be a 61 percentage point increase in retesting among students who scored 15 to 16 on their first attempt and a 29 percentage point increase in retesting among those who scored 17 to 18. Assuming the fraction of students retaking the ACT was constant across ACT scores, 97 percent of students scoring 15 and 16 would have to retake the ACT to see effects this large. This is too large to be plausible.[11] Since students often study between test dates, even if this effect could be explained only by retesting, the gains would still

[11] In fact, the fraction that would need to retake the test to see effects this large would probably need to be even greater. This calculation assumes that students retaking the ACT as a result of the TELS would see score increases as large as those who retook the ACT without it. But this is unlikely as there was a reason some students chose to retake the ACT initially.

have probably partially been a result of increased human capital.

3.4.4 Impact of the TELS on Different Subgroups

Tables 3 and 4 also display the results from restricting the estimation of equations (3.2) and (3.3) to different subgroups. They show that African-Americans were significantly less responsive to the TELS than Asians and Caucasians, and males were slightly more responsive than females.

The results in Table 3 suggest an African-American who would have scored below 19 without the TELS was over five times less likely than an Asian and seven times less likely than a Caucasian to increase her score to 19 or higher: the point estimate for blacks is smaller than for other groups and the fraction scoring below 19 is larger. Part, but not all, of this disparity in apparent responsiveness is due to the fact that African-Americans scoring below 19 scored, on average, lower than the other racial groups, so they would have had to increase their score by more to reach the cutoff of 19. However, a higher percentage of Tennessee blacks score between 15 and 18 than Tennessee whites, so if blacks and whites were equally affected by the TELS, the point estimate for blacks should be higher than the point estimate for whites. Repeating the analysis in Table 5 separately for blacks and whites suggests African-Americans were at least three times less responsive to the TELS than Caucasians.

Males were slightly more responsive to this program than females. The coefficients of different specifications of equation (3.2) are 25 percent to 37 percent higher for males than females. In pooled data the interaction term between being male and the presence of the TELS is positive and significant in every specification while the score distributions of males and females before the TELS were very similar. This result is interesting in light of several papers that find larger effects of financial incentives on educational attainment and performance for females (for example Angrist, Lang, and Oreopoulos 2006, Angrist and Lavy 2002, and Dynarski 2005).

3.5 Impact of the TELS on College Applications

The TELS did not affect where students sent their ACT scores or students' stated college preferences. I analyze students' preferences for in-state versus out-of-state, four-year versus two-year, and four-year in-state versus two-year in-state colleges as well as the tuition students were willing to pay and find no robust effect of the TELS.

For each type of school listed above, I estimate equation (3.2) separately using the total number of scores the student sent to that type of college, a dummy variable for whether the student sent any score to that type of college, and the student's preference for that type of college as dependent variables.[12] I also use the maximum tuition the student reports being willing to pay and the average tuition of colleges she sent scores to as dependent variables. Tables 6 and 7 present the results for the whole sample, students who scored 19 or higher on the ACT, students who scored below 19, and students who scored below 19 and reported a GPA below 3.0. I include all of the control variables and cluster standard errors at the state-year level.

While these subgroups are endogenous, if the results are due to the TELS, we would expect students scoring 19 or higher to be much more responsive than students who are likely ineligible. The results are not driven by the endogeneity of the subgroups; they are the same when ACT score and GPA are predicted using data before the TELS and the subgroups restricted based on those variables.

Panel A of Table 6 shows that the number of scores sent did not change significantly either in the aggregate or for any of the subgroups examined, allowing us to more easily interpret changes (or the lack thereof) in score-sending as changes (or lack thereof) in preferences. Panel B shows that the TELS did not induce students to prefer in-state colleges more strongly. In fact, while not all significant, the point estimates all indicate students were less likely to prefer in-state as compared to out-of-state colleges: students sent fewer scores to in-state colleges, more scores to out-of-state colleges, and were less likely to say they wanted to attend college in-state. This is not due to preferences of students who were ineligible for the scholarship: the point estimates for students scoring 19 or higher are all signed in the "wrong" direction as well.

While results from the entire sample indicate that Tennessee students increased their preference for four-year as opposed to two-year schools, the point estimates for students scoring 19 or higher, while not significant, indicate these students sent fewer scores to four-year colleges. It was only students scoring below 19 (including those with GPAs below 3.0) who sent scores to more four-year colleges in 2004. Moreover, while students scoring 19 or higher did realize decreases in both the total number of two-year colleges they sent scores to and their preferences for four-year colleges,

[12] Average tuition is calculated using data from the Integrated Postsecondary Education Data System (IPEDS). I use tuition in 2004-2005 regardless of when the student graduated high school to pick up changes in preferences over types of schools as opposed to changes in tuition of more or less popular schools. For schools that have different in-state and out-of-state tuitions, I use in-state tuition if the student lives in the same state as the school and the out-of-state tuition if she doesn't. It is interesting to note that the maximum tuition students indicate being willing to pay is less than half of the average tuition of the colleges they send their scores to. This most likely indicates that students are poorly informed about the cost of college. As long as the nature of their misinformation is not changing parallel to the TELS, evaluating their stated preferences over college tuition is still instructive.

these changes were only about half and one-third as large, respectively, as those realized by students with ACT scores below 19 and GPAs below 3.0.

While preferences for four-year colleges could theoretically decrease if students began to prefer two-year in-state colleges over four-year out-of-state colleges (an admittedly very unusual response), the predictions indicate unambiguously that the TELS should increase the preference for four-year in-state colleges. However, the point estimates for students scoring 19 or higher indicate that these students were less likely to send scores or express preferences for attending four-year in-state colleges.

Finally, there was no effect of the TELS on the tuition students were willing to pay. In the aggregate, Tennessee students both said they were willing to pay higher tuition and sent their scores to more expensive schools in 2004. However, students scoring 19 or higher were the only group not to see a significant increase in the actual tuition of colleges scores were sent to and the point estimate for this group is approximately one third of those for the other two groups. They also did not report larger increases in the tuition they were willing to pay than the other groups.

While students could have changed their college preferences later in the application process, this analysis provides strong evidence that students had not changed their college preferences as a result of the TELS by the time they took the ACT. This section shows the value of using micro data with detailed background characteristics to analysis the difference in score-sending.

3.6 Conclusion

The Tennessee Education Lottery Scholarship Program increased the return to scoring 19 or above on the ACT for students who were unsure of their ability to win the scholarship based on their grades and were interested in attending a Tennessee college. It also decreased the cost of attending in-state as compared to out-of-state and four-year in-state as compared to two-year in-state colleges for scholarship winners.

The TELS did not induce scholarship winners to change their college preferences. Students did respond to the scholarship, however. Graphically, it is clear that students' scores increased sharply around the threshold of 19 while difference-in-difference regressions show that the probability a given student would score 19 or higher on the ACT increased by 6 percent to 8 percent in the first year of the program. This effect is extremely robust. It is also a very large increase, implying that one out of every seven students who could have increased their scores to 19 or higher did so as a result of the TELS. African-Americans responded very little to the scholarship incentive while males were more responsive than females.

The large increase in ACT scores induced by the TELS show that policies that reward students for their academic performance can potentially generate large improvements in high school achievement. The fact that this performance improvement is too large to result from students simply retaking the test suggests that it may likely indicate true human capital accumulation. However, it remains to be seen whether this is ACT-specific human capital, such as learning the directions for the test, or whether this is human capital that will positively affect other outcomes.

3.7 Appendix 1: Mathematical Explanation of the Multivariate Reweighting Procedure

Let $f(ACT|z, t, TELS)$ be the distribution of ACT scores for students with a set of observable characteristics z, in year t, in a state of the world where the TELS is in place ($TELS = 1$) or a state of the world where it is not ($TELS = 0$). Let $dF(z|t)$ be the distribution of attributes z in the pool of ACT-takers in year t. Then, the actual distribution of scores in 2004 is

$$\int f(ACT|z, t = 2004, TELS = 1) \times dF(z|t = 2004) \tag{3.4}$$

while the distribution of scores that would have prevailed in 2004 if the population of test-takers was the same in 2004 as it was in 2000, is

$$\int f(ACT|z, t = 2004, TELS = 1) \times dF(z|t = 2000). \tag{3.5}$$

I define the reweighting function

$$\varphi_z(z) = \frac{dF(z|t = 2000)}{dF(z|t = 2004)} \tag{3.6}$$

so that $\varphi_z(z)$ essentially measures how much more frequently a student with characteristics z is in the 2000 pool of test-takers than in the 2004 pool. This allows me to calculate the counterfactual density:

$$\begin{aligned}\int f(ACT|z,t &= 2004, TELS=1) \times dF(z|t=2000) \\ &= \int f(ACT|z,t=2004, TELS=1) \times dF(z|t=2004) \times \varphi_z(z) \qquad (3.7)\end{aligned}$$

which is practically estimated by reweighting each observation in the 2004 pool by the relevant value of $\varphi_z(z)$, estimated as described in the text.

3.8 Appendix 2: Robustness Checks for the Effect of the TELS on ACT Scores

The difference-in-difference results are extremely robust in terms of both the magnitude and significance of the effect. Tables 3 and 4 show that the results are not sensitive to the specific controls or specification chosen. In Appendix Table 1, I show they are not due to serial correlation of ACT scores within states over time or the fact that I only analyze the behavior resulting from one state's policy change.

Serial correlation is not as likely to be a problem in my analysis as many other differences-in-differences papers because I use a short time series with only four periods (see Bertrand et al. 2004). Moreover, regressing the mean residuals within a state for a given year on their lagged value produces negative coefficients, suggesting that in fact the reported standard errors may be too high.

The coefficient is of similar magnitude and still significant when state-specific linear time trends are added and standard errors are clustered at the state level. It remains large and significant when the data are collapsed down to the state-year level and even, in all but one specification, when data are collapsed to the state-year level when state-specific linear time trends are added.[13] Even collapsing the data into two observations by state, one before and one after the TELS was implemented, yields a highly significant estimate which indicates a 7.6 percentage point increase in Tennessee students scoring 19 or higher as a result of the TELS.

Conley and Taber (2006) show that if the number of states whose policy changes stays fixed even when the number of states used as controls and the number of students in a state approaches infinity, the program effect estimated by a difference-in-difference regression is not

[13]The only coefficient that isn't significant is a coefficient in a regression that estimates 89 coefficients with 96 observations.

consistent. However, the last two sets of rows in Appendix Table 1 show that this is not driving my results. I use the consistent estimator of p-values that they suggest on both the individual and state-year data and the p-values are all below 0.05. The procedure for constructing this estimator is as follows.

Conley and Taber start with the following model of data at the state-year level:

$$\tilde{Y}_{jt} = \alpha \tilde{d}_{jt} + \tilde{X}'_{jt}\beta + \tilde{\eta}_{jt}. \tag{3.8}$$

Here j indexes the state, t indexes the time period and tildes denote variables which are projections onto group and time indicators. For any variable Z, $\tilde{Z}_{jt} = Z_{jt} - \bar{Z}_j - \bar{Z}_t - \bar{Z}$ where $\bar{Z}_j, \bar{Z}_t$, and $\bar{Z}$ and are the state, year, and overall mean of Z respectively. The variables X_{jt} are control variables that vary at the state-year level and d_{jt} is the indicator for whether the policy was in effect at the time; η_{jt} is the idiosyncratic error.

When the number of states who do change their policy (N_0) is finite, but the number of states who do not (N_1) grows large, the differences-in-differences estimator $\hat{\alpha}$ converges in probability to $\alpha + W$, where

$$W = \frac{\sum_{j=1}^{N_0} \sum_{t=1}^{T} (d_{jt} - \bar{d}_j)(\eta_{jt} - \eta_j)}{\sum_{j=1}^{N_0} \sum_{t=1}^{T} (d_{jt} - \bar{d}_j)^2}. \tag{3.9}$$

The states that change their policies are indexed by j equal to 1 to N_0; states which do not realize a policy change are indexed by j equal to N_0 +1 to N_1.

Assuming that the state-year errors are independent of any regressors and identically distributed across state-year cells, a consistent analog estimator of the conditional cumulative distribution function of W given the entire set of d's is

$$\hat{\Gamma}(w) = \left(\frac{1}{N_1}\right)^{N_0} \sum_{l_1=N_0+1}^{N_0+N_1} \cdots \sum_{l_{N_0}=N_0+1}^{N_0+N_1} 1 \left(\frac{\sum_{j=1}^{N_0} \sum_{t=1}^{T} (d_{jt} - \bar{d}_j)(Y_{l_j t} - X'_{l_j t}\hat{\beta})}{\sum_{j=1}^{N_0} \sum_{t=1}^{T} (d_{jt} - \bar{d}_j)^2} \right) \tag{3.10}$$

where 1(.) is an indicator function.

I use this estimated cumulative distribution function and report $\Pr(\hat{\alpha} + W < 0) = \Pr(W < -\hat{\alpha}) = \Gamma(-\hat{\alpha})$ as the p-value directly for the regressions done at the state-year level as Conley and Taber suggest. For the individual level regressions, presented in the line above, I also follow the suggestion of Conley and Taber. The regression model

$$Y_i = \alpha d_{jt} + X_{jt}^{'}\beta + \theta_j + \gamma_t + Z_i^{'}\delta + \mu_i \tag{3.11}$$

where Z_i are controls that vary at the individual level and μ_i are idiosyncratic errors, can be estimated as

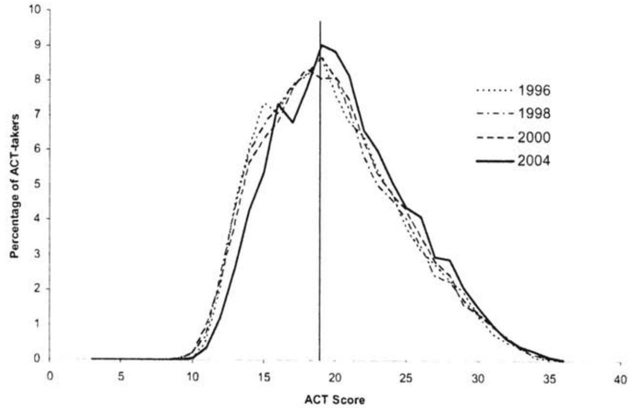
$$\begin{aligned} Y_i &= \lambda_{j(i)t} + Z_i^{'}\delta + \varepsilon_i & (3.12) \\ \lambda_{jt} &= \alpha d_{jt} + X_{jt}^{'}\beta + \theta_j + \gamma_t + \eta_{jt} & (3.13) \end{aligned}$$

where λ_{jt} are the individual state-year fixed effects and η_{jt} are errors. I estimate (3.12), use the estimates for λ_{jt} as the dependent variables in equation (3.13) and then calculate the p-value for each specification as in (3.10).

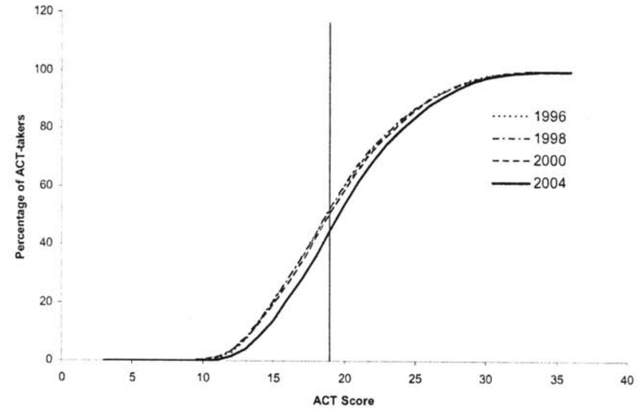
As a final robustness check, I compare the response of students who sent their test scores to at least one in-state college with those who did not. Sending a test score to an in-state college is endogenous (though as Section 3.5 shows, the TELS does not affect this outcome much). While the TELS may cause students who do not send their scores to any in-state colleges to increase their ACT score as a result of spillovers or uncertainty over whether they will want to attend an in-state college in the future, students who prefer to attend college in-state should have a larger response. This is the case. The coefficient from estimating (3.2) is 60 percent larger for students sending scores to in-state colleges. Adjusting for the distribution of test scores before the TELS shows even stronger results: the coefficients from the regressions restricted to students in predicted ACT ranges are four times larger for students sending scores in-state. For these students, the Kullback-Leibler divergences for 2004 are ten times larger than those for any other year, while for students who did not send any scores in-state, the divergences for 2004 are only 1.4 times larger than those in other years.

Figure 1. Actual Distributions of ACT Scores

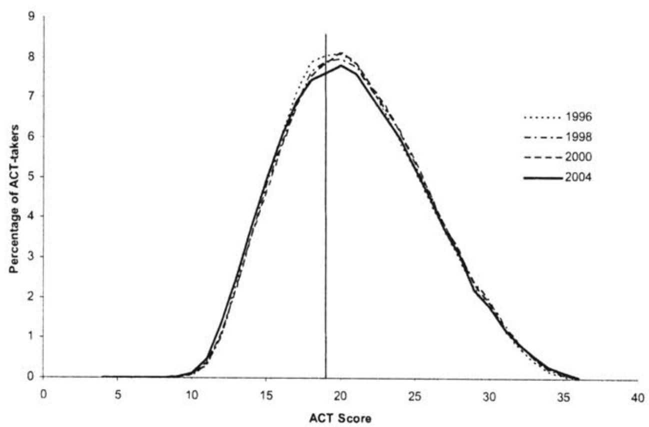
1a: Probability Distribution Function of ACT Scores in Tennessee



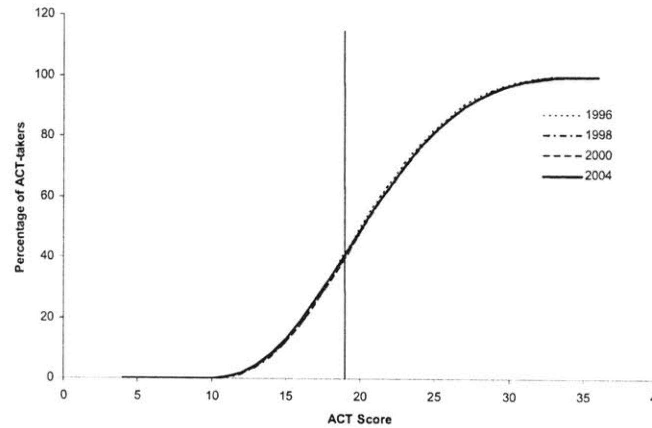
1b. Cumulative Distribution Function of ACT Scores in Tennessee



1c: Probability Distribution Function of ACT Scores in States Other than Tennessee



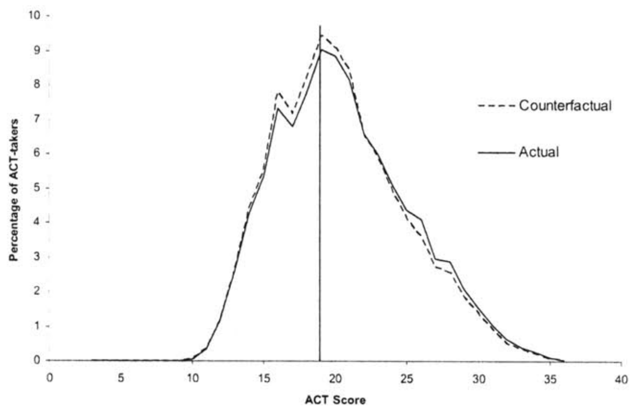
1d. Cumulative Distribution Function of ACT Scores in States Other Than Tennessee



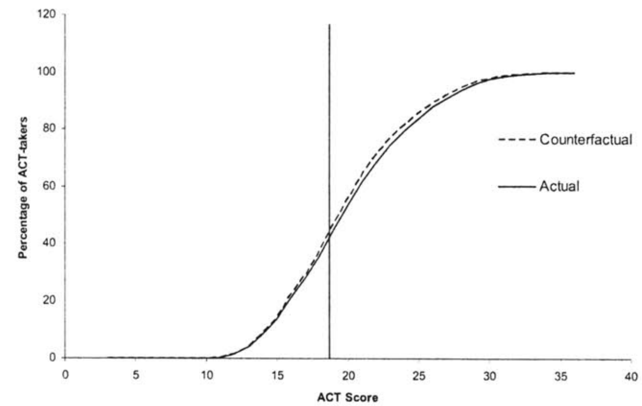
Data come from the ACT database. Vertical lines mark scores of 19.

Figure 2. Actual Distributions of ACT Scores Compared to Distributions Generated by Multivariate Reweighting

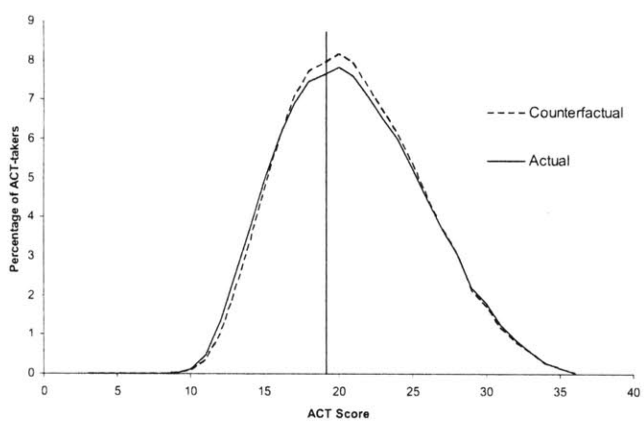
2a: Probability Distribution Function of ACT Scores in Tennessee



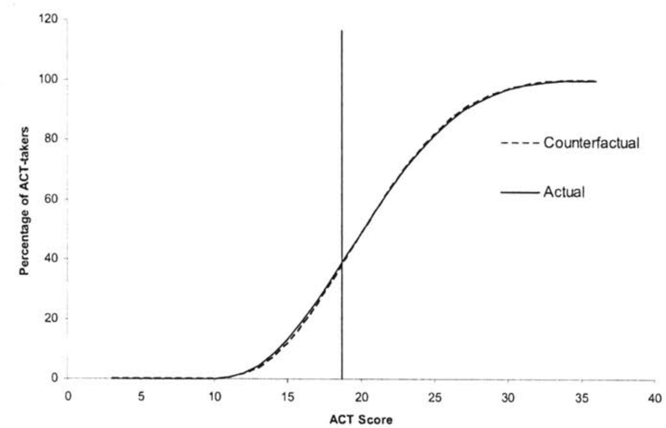
2b: Cumulative Distribution Function of ACT Scores in Tennessee



2c: Probability Distribution Function of ACT Scores in States Other than Tennessee



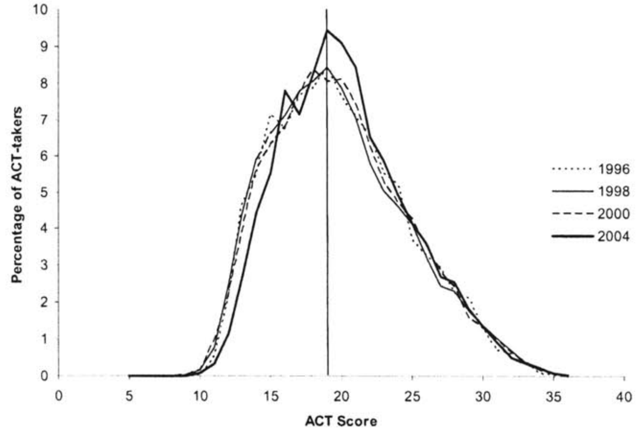
2d. Cumulative Distribution Function of ACT Scores in States Other Than Tennessee



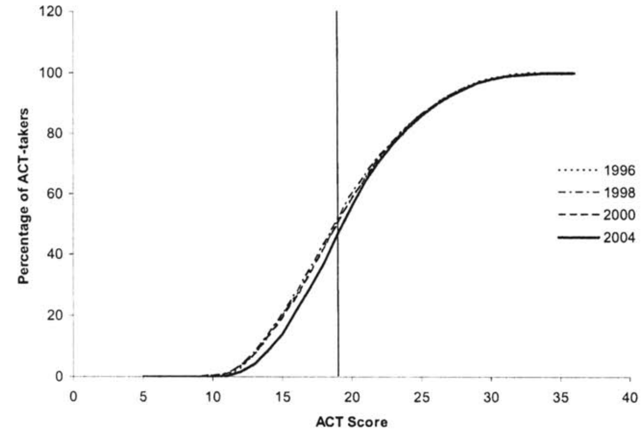
Data come from the ACT database. Vertical lines mark scores of 19. DFL distributions are produced using all of the controls described in footnote 9, state dummies (when applicable) and year dummies.

Figure 3. Distributions of ACT Scores Generated by Multivariate Reweighting

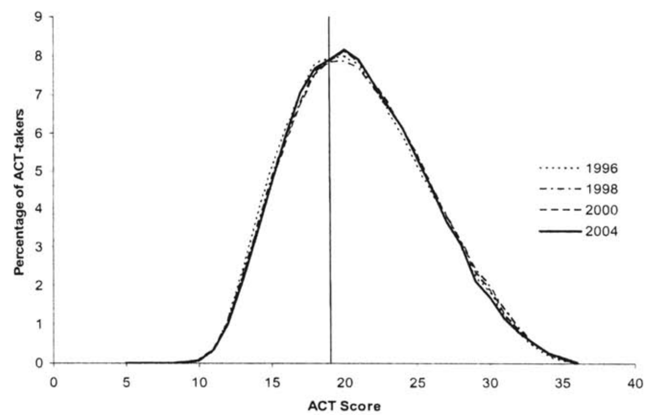
3a: Probability Distribution Function of ACT Scores in Tennessee



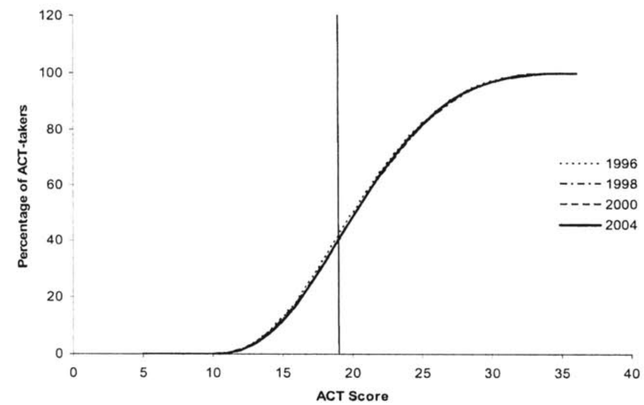
3b. Cumulative Distribution Function of ACT Scores in Tennessee



3c: Probability Distribution Function of ACT Scores in States Other than Tennessee



3d. Cumulative Distribution Function of ACT Scores in States Other Than Tennessee



Data come from the ACT database. Vertical lines mark scores of 19. Distributions are produced using all the controls described in footnote 9, state dummies (where applicable) and year dummies.

Table 1
College Choices of HOPE Scholarship Winners

Type of Institution	Number of Awards	Total Value
Two-Year Private	64	$119,000
Two-Year Public	4,827	$8,737,515
Four-Year Private	4,066	$12,799,297
Four-Year Public	14,330	$46,273,678
Total	23,287	$67,929,490

Notes: These figures only include students from the class of 2004. The data come from personal correspondence with Robert Anderson at the Tennessee Higher Education

Table 2
Kullback-Leibler Divergences Between ACT Distributions

Year	Actual Distribution	DFL Distribution	DFL Distribution with Only Exogenous Controls
		A. Tennessee	
1996	0.007	0.008	0.007
1998	0.003	0.003	0.002
2004	0.019	0.017	0.017
		B. Other States	
Year	Actual Distribution	DFL Distribution	DFL Distribution with Only Exogenous Controls
1996	0.001	0.002	0.001
1998	0.000	0.000	0.001
2004	0.002	0.001	0.001

Notes: Each cell in Panel A gives the Kullback-Leibler divergence between the actual Tennessee 2000 distribution of ACT scores and the either the actual Tennessee distribution for the year indicated, the Tennessee distribution created using the DFL technique and all the control variables, or the Tennessee distribution created using the DFL technique and only the fully exogenous controls. Panel B is the same for all states other than Tennessee. Exogenous controls are state and year fixed effects, income dummies, race dummies, and five background variables. The additional controls are eight aspects of the student's academic history and four aspects of the student's extracurricular participation. The specific controls are listed in footnote 9. The data come from the ACT database.

Table 3
The Effect of the TELS on Scoring 19 or Higher on the ACT: OLS Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	% Scoring < 19 in 2000
			A. Aggregate Results				
Everyone	0.086 **	0.086 **	0.078 **	0.063 **	0.072 **	0.061 **	42%
	(0.005)	(0.011)	(0.008)	(0.007)	(0.004)	(0.006)	
			B. Race				
Black	0.045 **	0.045 **	0.033 **	0.021 **	0.027 **	0.018 *	75%
	(0.009)	(0.006)	(0.006)	(0.008)	(0.007)	(0.008)	
Asian	0.075 **	0.075 **	0.066 **	0.048 **	0.063 **	0.048 **	38%
	(0.029)	(0.021)	(0.022)	(0.015)	(0.019)	(0.015)	
White	0.081 **	0.081 **	0.080 **	0.064 **	0.074 **	0.062 **	34%
	(0.006)	(0.008)	(0.007)	(0.007)	(0.004)	(0.006)	
			C. Gender				
Male	0.100 **	0.100 **	0.092 **	0.071 **	0.085 **	0.069 **	42%
	(0.008)	(0.014)	(0.010)	(0.008)	(0.004)	(0.005)	
Female	0.075 **	0.075 **	0.068 **	0.056 **	0.062 **	0.055 **	42%
	(0.007)	(0.010)	(0.007)	(0.007)	(0.004)	(0.006)	
Clustered SEs	No	Yes	Yes	Yes	Yes	Yes	
Income, Race, and Background	No	No	Yes	Yes	Yes	Yes	
Academic	No	No	No	Yes	No	Yes	
Extracurricular	No	No	No	No	Yes	Yes	

Notes: Each cell in the first six columns of data gives the coefficient and standard error of a separate regression limited to the individuals indicated by the leftmost column. The dependent variable is an indicator for whether the student scored 19 or higher on the ACT. The bottom rows indicate the control variables included in the regression and whether standard errors are clustered. When standard errors are clustered, it is at the state-year level and when they are not, they are White's robust standard errors. All regressions include state and year dummies. The specific control variables corresponding to each category are listed in footnote 9. One asterisk indicates the result is significant at the 5% level and two asterisks indicate the result is significant at the 1% level. The right-most column gives the percentage of test-takers in each subgroup scoring below 19 in 2000. Data come from the ACT database.

Table 4
The Effect of the TELS on Scoring 19 or Higher on the ACT: Probit Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	% Scoring < 19 in 2000
			A. Aggregate Results				
Everyone	0.080 **	0.080 **	0.070 **	0.060 **	0.070 **	0.060 **	42%
	0.228	0.228	0.228	0.208	0.213	0.205	
	(0.013)	(0.032)	(0.025)	(0.024)	(0.012)	(0.019)	
			B. Race				
Black	0.050 **	0.050 **	0.040 **	0.030 **	0.030 **	0.020 **	75%
	0.135	0.135	0.107	0.086	0.088	0.077	
	(0.027)	(0.018)	(0.019)	(0.026)	(0.021)	(0.027)	
Asian	0.070 *	0.070 **	0.060 **	0.050 **	0.060 **	0.050 **	38%
	0.209	0.209	0.210	0.172	0.205	0.174	
	(0.084)	(0.058)	(0.067)	(0.052)	(0.057)	(0.048)	
White	0.070 **	0.070 **	0.070 **	0.060 **	0.070 **	0.060 **	34%
	0.231	0.231	0.233	0.210	0.215	0.204	
	(0.018)	(0.025)	(0.022)	(0.022)	(0.010)	(0.018)	
			C. Gender				
Male	0.090 **	0.090 **	0.090 **	0.070 **	0.080 **	0.060 **	42%
	0.268	0.268	0.271	0.238	0.254	0.234	
	(0.020)	(0.038)	(0.031)	(0.026)	(0.013)	(0.019)	
Female	0.070 **	0.070 **	0.060 **	0.050 **	0.060 **	0.050 **	42%
	0.197	0.197	0.196	0.185	0.182	0.184	
	(0.018)	(0.027)	(0.020)	(0.022)	(0.012)	(0.020)	
Clustered SEs	No	Yes	Yes	Yes	Yes	Yes	
Income, Race, and Background	No	No	Yes	Yes	Yes	Yes	
Academic	No	No	No	Yes	No	Yes	
Extracurricular	No	No	No	No	Yes	Yes	

Each cell in the first six columns of data gives the average marginal effect, coefficient, and standard error of a separate regression limited to the individuals indicated by the leftmost column. The dependent variable is an indicator for whether the student scored 19 or higher on the ACT. The bottom rows indicate the control variables included in the regression and whether standard errors are clustered. When standard errors are clustered, it is at the state-year level and when they are not, they are robust standard errors. All regressions include state and year dummies. The specific control variables corresponding to each category are listed in footnote 9. One asterisk indicates the result is significant at the 5% level and two asterisks indicate the result is significant at the 1% level. The right-most column gives the percentage of test-takers in each subgroup scoring below 19 in 2000. Data come from the ACT database.

Table 5
Effect of TELS on Students with Different Predicted Academic Ability

Predicted Score	Coefficient (Standard		Percent Scoring Below 19 in	Coefficient/Fraction Scoring Below 19	Percent Accurately Predicted
Less than 11	0.032 (0.016)	*	97.3%	3.290	8.2%
Between 11 and 13	0.006 (0.008)		98.0%	0.612	30.3%
Between 13 and 15	0.010 (0.006)		90.1%	1.110	34.3%
Between 15 and 17	0.053 (0.007)	**	78.7%	0.067	25.8%
Between 17 and 19	0.076 (0.012)	**	60.2%	0.126	22.2%

Notes: ACT scores are predicted by an OLS regression using observations from 1996, 1998, and 2000, all the controls listed in footnote 16, and state and year dummies. The results in the four data columns are limited to students predicted to score in the range indicated in the leftmost column. The first data column gives the coefficient and standard error (clustered at the state-year level) from an OLS regression. The dependent variable is an indicator for whether the student scored 19 or higher on the ACT; all controls listed in footnote 9 plus state and year dummies are included. The far right-hand column indicates the percentage of students predicted to be in that range in 1996, 1998, and 2000 who did score in the range (e.g. for students predicted to be in the 11 to 13 range, the percentage of students scoring 11 or 12). One asterisk indicates the result is significant at the 5% level and two asterisks indicate the result is significant at the 1% level. Data come from the ACT database.

Table 6
Effect of the TELS on College Preferences and ACT Score-Sending:
Robustness Check, In-State vs. Out-of-State Colleges and Two-Year vs. Four-Year Colleges

	Full Sample		ACT at Least 19		ACT Below 19		ACT Below 19 and GPA		Mean in 2000
	A. Robustness Check								
Total Scores Sent	-0.043		-0.081		-0.002		0.017		2.99
	(0.047)		(0.053)		(0.032)		(0.035)		
	B. In-State vs. Out-of-State Colleges								
Total In-State Colleges	-0.084		-0.097		-0.031		0.005		2.08
	(0.052)		(0.057)		(0.039)		(0.042)		
Any In-State College	-0.005	*	-0.004		0.000		0.002		0.79
	(0.002)		(0.003)		(0.002)		(0.002)		
Total Out-of-State Colleges	0.041	**	0.015		0.029	*	0.012		0.91
	(0.014)		(0.015)		(0.014)		(0.016)		
Any Out-of-State College	0.023	**	0.013		0.018	*	0.011		0.45
	(0.008)		(0.008)		(0.009)		(0.008)		
Prefer to Attend College In-State	-0.012		-0.021	*	0.001		0.011		0.72
	(0.007)		(0.010)		(0.006)		(0.006)		
	C. Four-Year vs. Two-Year Colleges								
Total 4-year Colleges	0.029		-0.033		0.064	**	0.098	**	2.58
	(0.034)		(0.043)		(0.021)		(0.023)		
Any 4-year College	0.002	*	0.000		-0.001		-0.003		0.83
	(0.001)		(0.000)		(0.002)		(0.003)		
Total 2-year Colleges	-0.073	**	-0.048	**	-0.066	**	-0.081	**	0.41
	(0.016)		(0.012)		(0.019)		(0.023)		
Any 2-year College	-0.039	**	-0.030	**	-0.020		-0.017		0.32
	(0.011)		(0.009)		(0.012)		(0.015)		
Prefer to Attend 4-year College	0.033	**	0.013	**	0.029	**	0.041	**	0.84
	(0.007)		(0.004)		(0.011)		(0.014)		

Notes: Each cell in the first four columns of data gives the coefficient and standard error of a separate regression on the dependent variable listed in the left-hand column. Each regression is limited to the individuals indicated by the column heading. The first four specifications in Panels B and C relate to scores sent whereas the last relates to stated preferences. All of the controls listed in footnote 9 as well as state and year dummies are included and standard errors are clustered at the state-year level. One asterisk indicates the result is significant at the 5% level and two asterisks indicate the result is significant at the 1% level. The right-most column gives the mean of each variable for the sample of 2000 Tennessee test-takers. Data on score-sending, preferences, GPA, and ACT score come from the ACT database and data on the level and location of the colleges comes from the IPEDs.

Table 7
Effect of the TELS on College Preferences and ACT Score-Sending:
Four-Year In-state vs. Two-Year In-State Colleges and Tuition Preferences

	Full Sample	ACT at Least 19	ACT Below 19	ACT Below 19 and GPA Less Than 3.0	Mean in 2000 Tennessee
A. Four-Year In-State vs. Two-Year In-State Colleges					
Total 4-year In-State Colleges	-0.016	-0.053	0.033	0.084 **	1.72
	(0.038)	(0.048)	(0.024)	(0.024)	
Any 4-year In-State College	-0.003	-0.006 *	0.000	0.003	0.76
	(0.002)	(0.003)	(0.002)	(0.003)	
Total 2-year In-State Colleges	-0.068 **	-0.043 **	-0.064 **	-0.080 **	0.36
	(0.015)	(0.010)	(0.019)	(0.023)	
Any 2-year In-State College	-0.037 **	-0.029 **	-0.018	-0.014	0.29
	(0.011)	(0.008)	(0.013)	(0.015)	
Prefer 4-year In-State College	0.012 **	-0.009	0.015 **	0.033 **	0.58
	(0.003)	(0.007)	(0.005)	(0.008)	
B. Tuition					
Preferred Maximum Tuition	251.8 **	109.1	129.7 *	53.5	$ 4,016.97
	(31.6)	(61.0)	(56.6)	(54.5)	
Average Tuition of Colleges Scores Sent To	235.8 **	76.0	227.3 **	266.0 **	$9,141.88
	(62.9)	(42.7)	(46.1)	(52.2)	

Notes: Each cell in the first four columns of data gives the coefficient and standard error of a separate regression on the dependent variable listed in the left-hand column. Each regression is limited to the individuals indicated by the column heading. The first four specifications in Panel A relate to scores sent whereas the last relates to stated preferences. All of the controls listed in footnote 9 as well as state and year dummies are included and standard errors are clustered at the state-year level. One asterisk indicates the result is significant at the 5% level and two asterisks indicate the result is significant at the 1% level. The right-most column gives the mean of each variable for the sample of 2000 Tennessee test-takers. Data on score-sending, preferences, GPA, and ACT score come from the ACT database and data on the level, location, and tuition of the colleges comes from IPEDs. Tuition is the in-state tuition in 2004-05 if the student lived in the same state as the college and the out-of-state tuition in 2004-05 if the student lived in a different state.

Appendix Table 1
Robustness Checks

	(1)	(2)	(3)	(4)	(5)	(6)
	A. Residual Regressions					
Basic Specification	-0.270	-0.270	-0.254 **	-0.298	-0.269	-0.295
	(0.344)	(0.344)	(0.000)	(0.193)	(0.173)	(0.173)
Including State Time Trends	-0.403 **	-0.403 **	-0.427 **	-0.496 **	-0.443 **	-0.502 **
	(0.075)	(0.075)	(0.082)	(0.081)	(0.088)	(0.083)
	B. Individual-Level Regressions					
Including State Time Trends	0.085 **	0.085 **	0.075 **	0.074 **	0.070 **	0.074 **
	(0.011)	(0.014)	(0.011)	(0.013)	(0.010)	(0.013)
Clustering at State Level	0.092 **	0.092 **	0.078 **	0.063 **	0.072 **	0.061 **
	(0.021)	(0.021)	(0.011)	(0.008)	(0.005)	(0.005)
	C. State-Year Regressions					
Basic Specification	0.076 **	0.076 **	0.049 **	0.042 **	0.054 **	0.048 *
	(0.006)	(0.006)	(0.008)	(0.015)	(0.008)	(0.019)
Including State Time Trends	0.078 **	0.078 **	0.075 **	0.059	0.053 *	
	(0.013)	(0.013)	(0.018)	(0.035)	(0.025)	
	D. State Pre/Postperiod Regressions					
Basic Specification	0.076 **					
	(0.005)					
	E. Conley-Taber P-values					
Individual-Level Regressions	0.043	0.043	0.043	0.000	0.000	0.000
State-Year Regressions	0.043	0.000	0.000	0.000	0.000	0.000
Clustered SE's	No	Yes	Yes	Yes	Yes	Yes
Income, Race, and Background	No	No	Yes	Yes	Yes	Yes
Academic	No	No	No	Yes	No	Yes
Extracurricular	No	No	No	No	Yes	Yes

Notes: Each cell in the first four panels gives the coefficient and standard error of a separate regression, the specification of which is indicated by the leftmost column. In panels B, C, and D the dependent variable is an indicator for whether the student scored 19 or higher on the ACT. In Panel C the data is collapsed to the state-year level whereas in Panel D the data is collapsed to a pre-TELS and post-TELS observation for each state. Cells in Panel C and D are left empty when the model is not identified. In, Panel A, the dependent variable is the average state residual from estimating equation (2) using the controls indicated by the column and the independent variables are the average residual from the previous year and a constant. Panel E computes Conley-Taber.(2006) p-values which correct for the fact that I only analyze one policy change. Their construction is explained in Appendix B. The bottom rows indicate the control variables included in the regression and whether standard errors are clustered. When standard errors are clustered, it is at the state-year level and when they are not, they are White's robust standard errors. The specific control variables corresponding to each category are listed in footnote 9. One asterisk indicates the result is significant at the 5% level and two asterisks indicate the result is significant at the 1% level. Data come from the ACT database.

Bibliography

Abraham, K. and M. Clark (2006). Financial Aid and StudentsŠ College Decisions: Evidence from the District of Columbia Tuition Assistance Grant Program. *Journal of Human Resources 41*(3), 578Ű610.

Acemoglu, D. and J.-S. Pischke (1998). Why Do Firms Train? Theory and Evidence. *Quarterly Journal of Economics 113*(1), 79–118.

Acemoglu, D. and J.-S. Pischke (1999). Beyond Becker: Training in Imperfect Labour Markets. *The Economic Journal 109*(453), pp. F112–F142.

Altonji, J. and C. Pierret (2001). Employer Learning and Statistical Discrimination. *Quarterly Journal of Economics 116*(1), 313–350.

Andrews, K. and R. Ziomek (1998). Score Gains on Retesting with the ACT Assessment. *ACT Research Report Series 98*(7), 1–36.

Angrist, J., D. Lang, and P. Oreopoulos (2002). Lead Them to Water and Pay Them to Drink: An Experiment with Service and Incentives for College Achievement. *NBER Working Paper 12790.*

Angrist, J. and V. Lavy (2002). The Effect of High School Matriculation Awards: Evidence from Randomized Trials. *NBER Working Paper 9389.*

Ariely, D. (2008). *Predictably Irrational: The Hidden Forces that Shape Our Decisions.* New York, NY: HarperCollins.

Autor, D. H. (2001). Why Do Temporary Help Firms Provide Free General Skills Training? *Quarterly Journal of Economics 116*(4), 1409–1448.

Avery, C. and T. J. Kane (2004). Student Perceptions of College Opportunities: The Boston COACH Program. In C. M. Hoxby (Ed.), *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*, pp. 355–393. Chicago: University of Chicago Press.

Banerjee, A. V. and E. Duflo (2000). Reputation Effects and the Limits of Contracting: A Study of the Indian Software Industry. *Quarterly Journal of Economics 115*(3), 989–1017.

Barsky, R., J. Bound, K. Charles, and J. Lupton (2002). Accounting for the Black-White Wealth Gap: A Nonparametric Approach. *Journal of the American Statistical Association 97*(459), 663–673.

Becker, G. (1964). Human Capital. *Columbia University for the National Bureau of Economic Research.*

Behrman, J. R., M. R. Rosenzweig, and P. Taubman (1996). College Choice and Wages: Estimates Using Data on Female Twins. *Review of Economics and Statistics 78*(4), 672–685.

Black, D., K. Daniel, and J. Smith (2005). College Quality and Wages in the United States. *German Economic Review 6*(3), 415–443.

Black, D. and J. Smith (2006). Estimating the Returns to College Quality with Multiple Proxies for Quality. *Journal of Labor Economics 24*(3), 701–728.

Bowen, W. G., M. A. Kurzweil, and E. M. Tobin (2005). *Equity and Excellence in American Higher Education.* Charlottesville: University of Virginia Press.

Brewer, D. J., E. R. Eide, and R. G. Ehrenberg (1999). Does it Pay to Attend an Elite Private College? Cross-cohort Evidence on the Effects of College Type on Earnings. *Journal of Human Resources 34*(1), 104–123.

Card, D. (1995). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. In L. N. Christofides, E. K. Grant, and R. Swidinsky (Eds.), *Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp*, pp. 201–222. Toronto: University of Toronto Press.

Card, D. and A. B. Krueger (2005). Would the Elimination of Affirmative Action Affect Highly Qualified Minority Applicants?: Evidence from California and Texas. *Industrial and Labor Relations Review 58*(3), 416–434.

Carneiro, P., K. T. Hansen, and J. J. Heckman (2003). Estimating Distributions of Counterfactuals with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on Schooling Choice. *International Economic Review 44*(2), 67–113.

Choi, J. J., D. Laibson, B. Madrian, and A. Metrick (2002). Defined Contribution Pensions: Plan Rules, Participant Decisions, and the Path of Least Resistance. In J. Poterba (Ed.), *Tax*

Policy and the Economy, pp. 67–113. Cambridge, MA: MIT Press.

Cornwell, C., K. Lee, and D. Mustard (2005). Student Responses to Merit Scholarship Retention Rules. *Journal of Human Resources 40*(4), 895–917.

Cornwell, C., K. Lee, and D. Mustard (2006). The Effects of State-Sponsored Merit Scholarships on Course Selection and Major Choice in College. *IZA Discussion Paper 1953.*

Cornwell, C. and D. Mustard (2006). Merit Aid and Sorting: The Effects of HOPE-Style Scholarships on College Ability Stratification. *IZA Discussion Paper 1956.*

Cornwell, C., D. Mustard, and D. Sridhar (2006). The Enrollment Effects of Merit-Based Financial Aid: Evidence from GeorgiaŠs HOPE Scholarship. *Journal of Labor Economics 24*(4), 761–786.

Dale, S. B. and A. B. Krueger (2002). Estimating the Payoff to Attending a More Selective College: an Application of Selection on Observables and Unobservables. *Quarterly Journal of Economics 117*(4), 1491–1527.

Day, J. C. and E. C. Newburger (2002). The Big Payoff: Educational Attainment and Synthetic Estimates of Work-life Earnings. *Current Population Reports*, P23–210.

Dinardo, J., N. Fortin, and T. Lemieux (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica 64*(5), 1001–1044.

Dynarksi, S. (2004). The New Merit Aid. In C. M. Hoxby (Ed.), *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*, pp. 67–93. Chicago: University of Chicago Press.

Dynarski, S. (2005). Building the Stock of College-Educated Labor. *NBER Working Paper 11604.*

Ellwood, D. T. and T. J. Kane (2000). Who is Getting a College Education: Family Background and the Growing Gaps in Enrollment. In S. Danziger and J. Waldfogel (Eds.), *Securing the Future: Investing in Children from Birth to College*, pp. 283–324. New York, NY: Russell Sage Foundation.

Farber, H. and R. Gibbons (1996). Learning and Wage Dynamics. *The Quarterly Journal of Economics*, 1007–1047.

Henry, G. and R. Rubenstein (2002). Paying for Grades: Impact of Merit-Based Financial Aid on Educational Quality. *Journal of Policy Analysis and Management 21*(1), 93–109.

Hill, C. B., G. C. Winston, and S. A. Boyd (2005). Affordability: Family Incomes and Net Prices at Highly Selective Private Colleges and Universities. *Journal of Human Resources 40*(4), 769–790.

Howell, J. S. (2004). *A Structural Equilibrium Model of the Market for Higher Education: Assessing the Impact of Eliminating Affirmative Action.* Ph. D. thesis, University of Virginia.

Hoxby, C. M. (1998). The Return to Attending a More Selective College: 1960 to the Present. http://www.nyu.edu/classes/jepsen/hoxby-selective.pdf.

Kremer, M. and E. Miguel (2007). The Illusion of Sustainability. *Quarterly Journal of Economics 112*(3), 1007–1065.

Kreps, D. M. and R. Wilson (1982). Reputation and Imperfect Information. *Journal of Economic Theory 27*(2), 253 - 279.

Kullback, S. and R. Leibler (1951). On Information and Sufficiency. *Annals of Mathematical Statistics 22*(1), 76–86.

Loewenstein, M. A. and J. R. Spletzer (1998). Dividing the Costs and Returns to General Training. *Journal of Labor Economics 16*(1), pp. 142–171.

Long, M. C. (2004). College Applications and the Effect of Affirmative Action. *Journal of Econometrics 121*(1-2), 319–342.

Loury, L. D. and D. Garman (1995). College Selectivity and Earnings. *Journal of Labor Economics 13*(2), 289–308.

Madrian, B. C. and D. F. Shea (2001). The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior. *Quarterly Journal of Economics 116*(4), 1149–1188.

McMillan, J. and C. Woodruff (1999). Interfirm Relationships and Informal Credit in Vietnam. *Quarterly Journal of Economics 114*(4), 1285–1320.

Ness, E. and B. Noland (2004). Targeted Merit Aid: Tennessee Education Lottery Scholarships. Presented at the 2004 Annual Forum of the Association for Institutional Research, Boston MA.

Newman, K. (1996). Job Availability: Achilles Heel of Welfare Reform. *National Forum 76*(3), 20–24.

Pallais, A. and S. Turner (2006). Opportunities for Low Income Students at Top Colleges and

Universities: Policy Initiatives and the Distribution of Students. *National Tax Journal 59*(2), 357–386.

Pope, D. and J. Pope (2008). Understanding College Choice Decisions: Why College Sports Success Matters. *Social Science Research Network Working Paper 1307788.*

Resnick, P., R. Zeckhauser, J. Swanson, and K. Lockwood (2006). The Value of Reputation on eBay: A Controlled Experiment. *Experimental Economics 9*, 79–101.

Saavedra, J. E. (2008). The Returns to College Quality: A Regression Discontinuity Analysis. Harvard University.

Spies, R. R. (2001). The Effect of Rising Costs on College Choice: The Fourth and Final in a Series of Studies on This Subject. *Princeton University Research Report Series 117.*

Tervio, M. (2009). Superstars and Mediocrities: Market Failure in the Discovery of Talent. *Review of Economic Studies 76*(2), 829–850.

Thaler, R. H. and C. R. Sunstein (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness.* New Haven: Yale University Press.

Tirole, J. (1996). A Theory of Collective Reputations (with Applications to the Persistence of Corruption and to Firm Quality). *The Review of Economic Studies 63*(1), pp. 1–22.

Turner, D. (2005). Ninety-Nine Percent of UT Freshmen Qualify for Lottery Aid. *Chattanooga Times Free Press August 14*, B1.

Unknown. *Statistical Abstract of Tennessee Higher Education 2003-2004.* Tennessee Higher Education Commission.

Winston, G. C. and C. B. Hill (2005). Access to the Most Selective Private Colleges by High Ability, Low-Income Students: Are They Out There? *Williams Project on the Economics of Higher Education Discussion Paper 69.*

Zhang, L. (2005). Do Measures of College Quality Matter? The Effect of College Quality on Graduates' Earnings. *The Review of Higher Education 28*(4), 571–596.