The Occupations and Human Capital of U.S. Immigrants*

Todd Schoellman †

November, 2008

Abstract

Across Census occupations in 2000, foreign-born workers accounted for as little as 0.6% and as much as 45% of total employment. This paper addresses the interrelated questions of why immigrants select into occupations as they do, and how their selection process affects the distribution of wages across occupations for American-born workers. A model featuring multi-dimensional human capital is proposed. Occupations vary in their intensity of the different dimensions of human capital, while workers vary in their endowment, due to their different early-life backgrounds and the effects of selection for immigrants. Comparative advantage sorts workers into occupations intensive in the skills in which they are abundant. Estimation using information on the occupational choices of immigrants and the characteristics of occupations yields measures of the human capital of immigrants from 130 countries across 5 human capital dimensions. Immigrants are relatively abundant in cognitive ability and physical skills but scarce in communications and language skills. Counterfactual, general equilibrium experiments suggest that immigrants have a small, highly skewed effect on the wages of American-born workers. If all immigrants were removed, wage changes would range from at most a 1.2% decline up to an 88% increase, with the median absolute wage change just 0.7%. The effects of removing only unauthorized immigrants are also considered.

*Thanks to Curtis Simon, Kevin Murphy, and the Bag Lunch participants at Clemson University for helpful comments on early work. Also, thanks to Sherry Meador for research assistance and to Tom Mroz for generous advice and use of computational resources. The usual disclaimer applies.

†Address: John E. Walker Department of Economics, Clemson University, Clemson, SC 29642. E-mail: tschoel@clemson.edu.

1 Introduction

Immigrants in the United States tend to cluster into a relatively small set of occupations. Figure 1 plots the density of occupations by the fraction of their workforce that is foreign-born, taken from 2000 U.S. Census data.[1] For one-third of occupations, immigrants represent less than 5% of the workforce; for almost three-fourths, they represent less than 10%. But the tail of the distribution is long, and immigrants represent as much as 45% of the workforce in some occupations. A similar story applies to unauthorized (also called illegal) immigrants. Since workers do not self-report as unauthorized, I use as a proxy being born in one of the fifteen countries with the highest estimated rates of unauthorized immigration. For over 80% of occupations unauthorized immigrants represent less than 5% of the workforce, but they represent up to 43% of the workforce in some occupations. This paper takes as its starting point the skewed distribution of immigrant workers into occupations, and asks two questions. First, what determines the occupational choices of immigrant workers? Second, what are the consequences of this sorting for the wages paid by occupation for American workers?

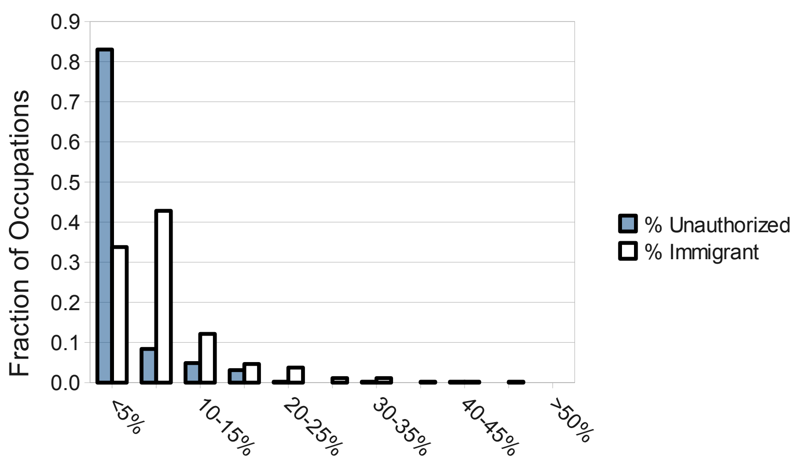

Figure 1: Fraction of Workers Foreign Born, by Occupation

To address these questions, consider a view of human capital and labor markets similar to recent work by Lazear (2003). Human capital is a vector of different attributes, such as physical skills and education, rather than a single scalar. Workers have heterogeneous endowments of human capital, so that they may be nimble and uneducated, nimble and educated, and so on. Occupations are differentiated by their technology for mapping human

[1]More details and documentation on these data is provided Section 4.

capital into output. For instance, professor is an occupation intensive in education, but not physical skills; acrobat is an occupation intensive in physical skills but not education. Educated workers earn higher wages as professors than as acrobats and, as long as they do not have too high a taste for being an acrobat, choose to be professors.

Immigrants from different countries face a variety of different environments and incentive structures, leading them to accumulate different combinations of human capital than Americans. Further, the set of foreign-born workers who immigrate is determined by a process of mixed self-selection and U.S. policy selection. The human capital endowments of immigrants will be distributed differently than the endowments of Americans because of these processes; and they may reasonably be supposed to vary by country of origin, and whether they entered through legal channels. Differences in human capital explain the observed sorting of immigrants into certain occupations. For instance, compared to Americans, immigrants are scarce in communications and language skills, which explains why occupations intensive in these skills (such as public relations specialist or funeral director) have so few immigrants in their workforce.

To study the impact of immigrants on wages requires estimates of their human capital. For several of the interesting and relevant dimensions of human capital, little or no information is available. This model provides an estimation strategy that infers the unobserved human capital. While information on human capital is limited, information on occupations is available in the U.S. Censuses, and there is a great deal of information about the characteristics of occupations. The O*Net Database provides an enormous quantity of information on the tasks, skills, abilities, and activities of the occupations used in the 2000 U.S. Census. I use this information to measure the skill intensity of occupations along five dimensions of human capital: education, training and experience, cognitive ability, physical skills, and language and communcation skills.[2] I then use logit frameworks to estimate the probability that workers from a given country choose a particular occupation as a function of the skill intensity characteristics of that occupation. Since the model predicts matching between worker endowments and occupational skill intensities, the resulting coefficients can be interpreted as the human capital endowments of the workers along the five dimensions. Hence, I am able to estimate the human capital endowments of workers from 131 countries along the 5 dimensions using 2000 U.S. Census PUMS.

The estimates indicate that immigrants as a group are relatively abundant in cognitive ability and scarce in experience/training and communications skills. Aggregate estimates

[2]Information on the characteristics of occupations has been used elsewhere to measure the specificity or generality of skills to occupations (Spitz-Oener 2006, Gathmann and Schonberg 2008) and the effects of computerization on workers (Autor, Levy, and Murnane 2003).

hide large composition effects: unauthorized immigrants are actually relatively scarce in every factor except physical skills, while other immigrants are highly abundant in cognitive ability in particular. The impact of immigrants on the labor market experiences of American-born workers are driven by the net impact immigrants have on the aggregate supply of the different types of human capital.

I estimate these impacts using two counterfactual experiments: removing all immigrants and removing only unauthorized immigrants from the United States. The model is structured so that these experiments simulate long-run, general equilibrium outcomes: they assume full adjustment of all prices, wages, capital, and workers across occupations. It assumes that workers can switch into the occupations vacated by immigrants if the skill intensity of the occupation and the worker's tastes make doing so desirable. It also accounts for real wage effects as prices for some goods in the consumption basket may rise.

The estimated wage changes are highly skewed. The largest wage decline is just 1.2% for removing all immigrants and 0.2% for removing unauthorized immigrants; the typical worker sees a small change, with just 0.7% and 0.2% median absolute wage changes; but the largest wage increases are 88% and 26%. The largest determinant of which occupations gain are their cognitive intensity (large positive effect) and communication intensity (large negative effect). The number of immigrants in the workforce is a much smaller determinant of wage gains, since workers reallocate themselves to take advantage of the newly vacated occupations. Broadly, the results suggest that immigration leads Americans to substitute into communications-intensive alternative occupations. However, for some occupations there are no good communications-intensive substitutes; it is this small handful of occupations that gain most from removing immigrants.

There is a large related literature on the effects of immigrants on wages, but most previous studies focus on effects on aggregate wages, or on a world with two types of workers, skilled and unskilled. This paper is most closely related to four previous studies. Ottaviano and Peri (2006) shares with this paper a long-run framework where capital can eventually adjust; they also consider the short-run and adjustment dynamics. Peri and Sparber (2008b) shows that immigrants specialize in manual occupations and suggests that this may limit competition by immigrants, which is mirrored here by the ability of Americans to substitute into communications-intensive occupations when they are available. However, I find that for some occupations no effective substitute is available, which drives distributional consequences of immigration. Peri and Sparber (2008a) and Borjas (2005) show that immigration affects high-skilled Americans as well, and that in particular it causes American-born workers to specialize in communications-intensive advanced degrees. I find a similar effect that immigrants have large effects on science and technology occupations, since as a group they are not communications-intensive. The primary difference between the last three studies and mine is that they focus on measuring the employment reallocation effect in the data; my main focus is inferring the human capital of workers and simulating the wage consequences.

The paper proceeds as follows. Section 2 presents the model. Section 3 illustrates the main properties of the model and the assumptions under which it is estimable. Section 4 introduces the data and estimates the human capital endowments of immigrants. Section 5 conducts the experiments using measured skills. Section 6 concludes.

2 A Model of Labor Markets with Vector Human Capital

2.1 Occupations and Human Capital

The model is a static representation of labor markets. Human capital H is an array of S elements, $H = (h_1, h_2, ..h_S)$, with a representative element indexed by h_s. Each s denotes a specific type of human capital, which I call a skill, although it may also include abilities, training, or any of the other common notions of human capital. Human capital endowments are defined on $(0, \infty)^S$.

There are J occupations in the economy, indexed by j. Each occupation utilizes all of the available skills of workers, but occupations vary in how intensively they use each of the skills. An occupation is characterized by a set of technological parameters $(\omega_s^j)_{s=1}^S$. The effective labor input of a worker with human capital H who works $n^j(H)$ hours in occupation j is given by:

$$n^j(H)\Pi_{s=1}^S(h_s)^{\omega_s^j}$$

ω_s^j is occupation j's s-intensity.

2.2 Firms

For every occupation, there exists a continuum of firms that employ workers of only that occupation. Firms hire workers and rent capital from households, then sell the produced occupational outputs to a final goods producer. Firms post wage schedules $W^j(H)$ giving the wage that they would be willing to pay for each possible endowment of skills, and a number of hours $L^j(H)$ that they would like to hire at that wage. Given the large number of firms, the posted wage contracts will offer exactly each worker's marginal product. Firms also take the rental price of capital R as given and rent a quantity $K^j(H)$ of capital for each type of worker. They find it optimal to vary capital allocated to workers with different levels of human capital.

Output $Y^j(H)$ of workers H in occupation j is the usual Cobb-Douglas aggregate over capital $K^j(H)$, effective labor input, and labor-augmenting technology A that is general across occupations. Firms choose capital and hours for each skill type to maximize profits per skill type:

$$P^j(K^j(H))^\alpha(AL^j(H)\Pi_{s=1}^S(h_s)^{\omega_s^j})^{1-\alpha} - RK^j(H) - W(H)L^j(H) \tag{1}$$

The first-order conditions are given by:

$$\begin{aligned} P^j\alpha\frac{Y^j(H)}{K^j(H)} &= R \\ P^j(1-\alpha)\frac{Y^j(H)}{L^j(H)} &= W(H) \\ P^j(1-\alpha)\omega_s^j\frac{Y^j(H)}{h_s} &= \frac{\partial W(H)}{\partial h_s}L^j(H) \end{aligned}$$

There is a single price-taking final goods producer. The producer faces prices P^j and purchases quantities of occupational outputs X^j. It aggregates the occupational outputs using a CES production function with elasticity of substitution ψ. It sells its output Y to consumers. I normalize the price of the final good to be the numeraire of the economy. Then the final goods producer maximizes profits:

$$\left[\sum_{j=1}^J(X^j)^{1-1/\psi}\right]^{\psi/(\psi-1)} - \sum_{j=1}^J X^jP^j \tag{2}$$

2.3 Workers

Workers have additively separable preferences over consumption and time spent in the labor market. Their preferences over consumption c are given by a standard CRRA function. Their preferences over time spent in the labor market depend on hours of work and the occupation where those hours are worked. Thus, the disutility of a forty-hour workweek may vary depending on whether the forty hours are spent working as a lawyer or a landscaper. These preferences are specific to the worker. I denote by ε^j the idiosyncratic preference of a worker for occupation j.

Workers have two forms of heterogeneity: in their skill endowments, H, and in their preferences for occupations $\varepsilon = (\varepsilon^j)_{j=1}^J$. I describe a worker by her skills and preferences, (H, ε), so $c(H, \varepsilon)$ is the consumption of such a worker, and so on. (H, ε) is a random draw from a joint distribution with pdf $\xi(H, \varepsilon)$ which is defined on $(0, \infty)^S \times (0, \infty)^J$. I assume that ξ is integrable and well-behaved so that expected utility exists; in the next section I restrict ξ so that this is true.

Given a worker's skill endowment and her draw of occupational preferences, her utility function is given by:

$$U(H, \varepsilon) = \frac{c(H, \varepsilon)^{1-1/\sigma}}{1 - 1/\sigma} - \log\left(\sum_{j=1}^{J} \frac{n^j(H, \varepsilon)}{\varepsilon^j}\right) \tag{3}$$

Workers are also endowed with $a(H)$ units of capital. Their income comes from renting out their endowment of capital at rate R and from their wages $\sum_{j=1}^J w^j(H) n^j(H, \varepsilon)$. They spend their income on consumption, $c(H, \varepsilon)$, so their budget constraint is:

$$c(H, \varepsilon) = Ra(H) + \sum_{j=1}^{J} w^j(H) n^j(H, \varepsilon) \tag{4}$$

Workers choose consumption, hours worked, and occupations to maximize their utility, subject to their budget constraint and the time restriction $\sum_{j=1}^J n^j(H, \varepsilon) \leq \bar{N}$. One key feature of the problem is that the log preferences guarantee that each worker chooses a single occupation based on wage offers and their occupational tastes ε.

Proposition 1 – ***Independence of Occupational Choice***
The workers' choice problem can be analyzed in two separate pieces. First, they choose the occupation that maximizes $w^j(H)\varepsilon^j$. Second, they choose consumption and hours, which are independent of their taste realizations ε.

Proof: Combine the FOC for consumption and hours worked to find that for occupations with positive hours:

$$\sum_{j=1}^{J} \frac{n^j(H, \varepsilon)}{\varepsilon^j} = \frac{c(H, \varepsilon)^{1/\sigma}}{w^j(H)\varepsilon^j}$$

with $n^j = 0$ otherwise. As long as ξ is continuous, this equation will hold for only one occupation. Then substitute into equations (3) and (4) to find the equivalent problem:

$$
\begin{aligned}
\max \quad & U(H,\varepsilon) = \frac{(c(H,\varepsilon))^{1-1/\sigma}}{1-1/\sigma} - \frac{1}{\sigma}\log(c(H,\varepsilon)) + \log(w^j(H)\varepsilon^j) \\
\text{s.t.} \quad & c(H,\varepsilon) - (c(H,\varepsilon))^{1/\sigma} = Ra(H) \\
\text{s.t.} \quad & \sum_{j=1}^{J} n^j(H,\varepsilon) \le \bar{N}
\end{aligned}
$$

The only term that depends on occupational choice is $\log(w^j(H)\varepsilon^j)$. The optimal $c(H,\varepsilon)$ is independent of ε, and by the first-order conditions, so too is $n(H,\varepsilon)$. QED

For the rest of the paper, I omit the irrelevant ε when possible, using only $c(H)$. Hours worked $n(H)$ also does not vary with tastes ε, but the occupation chosen does. I let $d^j(H,\varepsilon)$ be an indicator function taking a value of 1 if worker (H,ε) chooses occupation j, and taking 0 otherwise. Given the functional form of this equivalent problem, ε^j represents the compensating wage differential across occupations. If a worker's draw ε^j for lawyer is twice that of her draw for fire fighter, her wage as a fire fighter needs to be twice her wage as a lawyer to make her indifferent between the two occupations.

2.4 Equilibrium

For the purposes of conducting counterfactual experiments, it is necessary to define the equilibrium conditions of the economy. There are four sets of market clearing conditions for this economy: one condition for output, one condition for capital, one condition for each of the occupational goods markets, and one condition for each type of human capital. They are given by:

$$Y = \int\int c(H)\xi(H,\varepsilon)dHd\varepsilon \tag{5}$$

$$X^j = \int\int Y^j(H)\xi(H,\varepsilon)dHd\varepsilon \quad \forall j \tag{6}$$

$$\sum_{j=1}^{J}\int K^j(H)dH = \int\int a(H)\xi(H,\varepsilon)dHd\varepsilon \tag{7}$$

$$L^j(H) = \int n(H)d^j(H,\varepsilon)\xi(H,\varepsilon)d\varepsilon \quad \forall H \tag{8}$$

An equilibrium in this economy is a set of prices $(P^j, R, W(H))$, allocations for the workers, $(c(H), n(H), d^j(H,\varepsilon))$, allocations for intermediate goods firms, $(K^j(H), L^j(H), Y^j(H))$, and allocations for the final goods producer (Y, X^j) that satisfy the following conditions:

1. Taking prices as given, workers maximize their utility (3) subject to their budget constraint (4) and time restriction $\sum_{j=1}^{J} n^j(H) \leq \bar{N}$.
2. Taking prices as given, intermediate firms maximize profits, (1).
3. Taking prices as given, the final goods producer maximizes profits, (2)
4. Markets clear, (5) - (8).

3 Equilibrium Predictions

The equilibrium has two main predictions which are useful for the results that follow. First, labor market outcomes are characterized by comparative advantage driven by endowments, similar to the Heckscher-Ohlin theory of trade. Workers who are more skill s-abundant are more likely to choose occupations that are s-intensive. The random draws of preferences make the results easier to characterize by changing discrete, binary outcomes to continuous probabilities of choosing occupations. Further, the random draws suggest a way to estimate the human capital endowments using logits. Second, aggregate prices and wages are affected by the aggregate supply of different combinations of human capital. This result gives the counterfactual experiments their interest, by linking the wage effects of immigrants to their impact on the aggregate skill distribution.

3.1 Allocation of Workers to Occupations

In equilibrium, the wage offered to worker H if she chooses occupation j is given by:

$$W^j(H) = \frac{(P^j)^{1/(1-\alpha)}}{R^{\alpha/(1-\alpha)}} \alpha^{\alpha/(1-\alpha)}(1-\alpha)A\Pi_{s=1}^{S}(h_s)^{\omega_s^j} \tag{9}$$

Workers choose the occupation j that maximizes the product of wages and the idiosyncratic preference for occupation j. I respecify this as maximization in logs for convenience:

$$\log(w^j(H)\varepsilon^j) = \log\left(\frac{A(1-\alpha)\alpha^{\alpha/(1-\alpha)}}{R^{\alpha/(1-\alpha)}}\right) + \frac{1}{1-\alpha}\log(P^j) + \sum_{s=1}^{S}\omega_s^j \log(h_s) + \log(\varepsilon^j)$$

The model can be estimated under a variety of assumptions on $\xi(H, \varepsilon)$. However, two assumptions are particularly helpful in making the estimation computationally tractable. First is an assumption over the joint distribution:

Assumption 1 – ***Independence of Tastes and Endowments***
H and ε are independently distributed with pdf's $f(H)$ and $g(\varepsilon)$.

Note that this assumption does not rule out skill endowment affecting occupational choice. It merely constrains the effects to come through the wage channel. The second assumption is over the functional form of the idiosyncratic preferences:

Assumption 2 – ***Distribution of Preferences***
ε^j is distributed i.i.d according to the Type-2 Gumbel distribution or, equivalently, $\log(\varepsilon^j)$ is distributed i.i.d according to the Type-I extreme value distribution.

Assumptions 1 and 2 are used mostly to make estimation computationally practical, although they are also useful for deriving clean propositions about the model. The extreme value distribution means that this problem fits into the probablistic choice framework of McFadden (1974). It is amenable to estimation using various logit methods; I consider the conditional and mixed logit approaches in the next section. Logit models are well-known to be more practical than alternatives such as multinomial probits for estimating data sets with large sample size and a large number of variables; I have both. Under Assumptions 1 and 2, the probability that a worker chooses occupation j' conditional on human capital H is given by:

$$q(j'|H) = \frac{\left(P^{j'}\right)^{1/(1-\alpha)} \Pi_{s=1}^{S} h_s^{\omega_s^{j'}}}{\sum_{j=1}^{J} \left[\left(P^{j}\right)^{1/(1-\alpha)} \Pi_{s=1}^{S} h_s^{\omega_s^{j}}\right]} \tag{10}$$

The probability that a worker chooses occupation j' is merely the wage that she would earn in occupation j', divided by the sum of the wages she could earn in each possible occupation. By a usual law of large numbers argument, $q(j'|H)$ also represents the fraction of workers with endowment H who choose occupation j'. One convenient result of using the logit framework is that it is straightforward to give the comparative statics results. For this model the key comparative static is how changes in a worker's skill abundance affects her probability of matching in each of the J occupations.

Proposition 2 – ***Abundance-Intensity Matching***
A marginal increase in $\log(h_s)$ makes a worker more likely to work in occupations that are more s-intensive than the expected local alternative and less likely to work in occupations that are less s-intensive than the expected local alternative.

The proposition comes directly from the usual marginal effects equation in a conditional logit model.[3] It is the analogue to a comparative advantage in trade: an increase in s-abundance makes a worker more likely to choose s-intensive occupations. With multiple choices and idiosyncratic preferences, an occupation is s-intensive if its intensity parameter ω_s^j is higher than the probability-weighted local alternative for a given worker.

For a marginal change it is possible to hold prices and wages constant. An important and related question is what would happen to prices and wages if all workers became more s-abundant. Proposition 2 is inherently partial equilibrium, so it offers little guidance to these questions. In the next section, I provide a general equilibrium result.

3.2 Prices and Wages in General Equilibrium

The wages offered to workers who choose two different occupations will in general depend on the prices offered for the output of those occupations, as can be seen by equation (9). Prices are determined in general equilibrium to allocate labor across occupations in a way that is consistent with the CES demand equation of the final goods producer. The primary determinant of the prices is the abundance of different types of skills. We would expect that an abundance of skill type H would lower the prices and wages of occupations in which H-endowed workers have a comparative advantage. If we specialize the economy to the case where all workers have the same human capital vector H, this is easily shown. The relative prices of any two goods in this economy are given by:

$$\frac{P^j}{P^{j'}} = \left[\frac{\Pi_{s=1}^S h_s^{\omega_s^j}}{\Pi_{s=1}^S h_s^{\omega_s^{j'}}}\right]^{-2(1-\alpha)/(\psi(1-\alpha)+1+\alpha)} \tag{11}$$

Proposition 3 follows directly from equations (11) and (9).

Proposition 3 – ***Skill Abundance, Prices, and Wages***
Suppose there are two economies where workers share a common human capital vector; let H be the human capital in the former and H' in the latter. Let workers with human capital H have a comparative advantage in occupation j instead of j', in the sense that $\frac{\Pi_{s=1}^S h_s^{\omega_s^j}}{\Pi_{s=1}^S h_s^{\omega_s^{j'}}} > \frac{\Pi_{s=1}^S (h_s')^{\omega_s^j}}{\Pi_{s=1}^S (h_s')^{\omega_s^{j'}}}$. Then the relative prices and wages of j will be lower in the economy with shared human capital H than in the economy with human capital H'.

Aggregate skill abundance affects wages and prices. Since immigrants have different skills than the average American-born workers, they affect wages and prices.

[3]The exact equation is $\frac{\partial q(j'|H)}{\partial \log(h_s)} = q(j'|H)\left[\omega_s^{j'} - \sum_{j=1}^J \omega_s^j q(j|H)\right]$

4 Empirical Strategy

If there were widely available information on the human capital dimensions of interest, I could use those measures to test the model's sorting predictions, and to estimate the impact of immigrants on wages. But for several measures, such as physical skills or cognitive ability, there is little or no information. Instead, I use the model to estimate the implied human capital endowments of workers. Since workers born in different countries experience different environments, I estimate human capital by country of birth. For measures of human capital where there are data available, I use those data as an exogenous check on my constructed measures.

4.1 Data

The data for this project are taken from two sources. Data on the occupations and characteristics of immigrants come from the 5% sample of the 2000 U.S. Census, drawn from the IPUMS-USA system (Ruggles, Sobek, Alexander, Fitch, Goeken, Hall, King, and Ronnander 2004). The Census asks every respondent to list their country of birth. For privacy reasons, it aggregates this data so that no birthplace with fewer than 10,000 immigrants is reported separately. After aggregation, there are observations for 131 different birthplaces, including the United States.[4] Some of the birthplaces are nonstandard; for instance, there are response categories for Czechoslovakia, the Czech Republic, and Slovakia, since immigrants may have departed before or after the split. I preserve every statistical entity which is separately identified, and refer to them as countries as a shorthand.[5]

The reason for focusing on country of birth is to estimate the results of different environments and sorting processes. To ensure that workers are exposed to their birth-country's environment, I use only workers who immigrate at age 18 or later. Other immigrants have endowments that are plausibly a mixture of birth country and U.S. environments. I also include only those who worked in the previous year and are no older than 65. The resulting sample is quite large, with over 5.7 million workers; there at least 139 workers from every country. Finally, the Census provides information on the occupation of workers based on the Standard Occupation Classification (SOC) system, although they merge some occupations into larger categories to preserve privacy. Overall, the Census version of SOC includes

[4] In another paper, I show that most immigrants come to the United States from their country of birth, with no signs of systematic bias (Schoellman 2008).

[5] There are two exceptions to this policy. First, I merge the United Kingdom together; second, I exclude North Korea, the USSR, and Russia, since it is not possible to identify them separately from other countries. The count of 131 already includes these reductions in sample size.

476 occupations.

Data on the underlying characteristics of occupations are derived from the O*NET database version 12.[6] The O*NET database project is the continuation of occupational characteristic descriptions that used to be provided in the Dictionary of Occupational Titles (DOT), which was last updated in 1991.[7] The database includes information on the worker characteristics, worker requirements, experience requirements, occupational requirements, workforce characteristics, and other occupation-specific information for the occupations in the SOC classification system. The O*NET Database includes information on 812 SOC occupations. I use the provided crosswalk to merge O*NET information into Census occupation codes. When occupations are aggregated I weight them using the employment weights taken from the May 2004 Occupational Employment Statistics Survey from the BLS; earlier surveys did not measure employment for some of the necessary disaggregated statistics.[8] There are 453 matched occupations with all the necessary information.

The O*NET database contains over 250 different measures for each occupation, an unwieldy amount of information for analytical purposes. I pare this information down into measures of five components of human capital: education, training and experience, cognitive ability, physical skills, and language and communcation skills. To do so, I select a few pieces of information that relate to each of the five dimensions, and then use principal component analysis to aggregate the information into a single measure. For education I focus on requirements for knowledge of subjects taught primarily in high school and college. For training and experience, I use measures of different training requirements and the observed level of experience. For cognitive ability I use measures of ability to reason and think originally. For physical skills I use measures of strength, coordination, and dexterity. For language and communication skills I use frequency of communication in different ways and with different groups. Appendix A provides further details on the data and how the measures are constructed. Tables 7 - 11 provide the comprehensive list of measures used to construct each skill, as well as the highest and lowest scoring occupations as a way to check the reasonableness of the results. All results are scaled to lie between 0 and 1 for convenience. The main results of this paper are reasonably robust to changes in the construction, including using more or less information, using simple averages to aggregate information, excluding some of the dimensions of human capital, and so on.[9] Hence, while

[6] Occupational Information Network (O*NET) and US Department of Labor/Employment and Training Administration (USDOL/ETA) (2007).

[7] U.S. Department of Labor, Employment, and Training Administration (1991).

[8] Bureau of Labor Statistics (2004).

[9] One common alternative would be to include all the variables in a large PCA analysis and extract the first n components. However, there is a problem with interpretability of the results. For instance,

alternative choices are possible and potentially interesting, I view this method of exploring the technological intensity of occupations and skills of workers along these five dimensions as a useful and reasonably robust step for human capital measurement. Before estimating the model, I provide some preliminary evidence that the constructed skill intensity measures and the model are plausible.

4.2 Checks on Intensity Measures

According to Proposition 2, workers who are more s-abundant should choose occupations that are s-intensive. Here, I perform a preliminary, joint test of the model and the measures of skill intensity. The Census provides some proxies for the skill endowments of workers. I test whether workers who are abundant in these proxies for skill endowment choose occupations that are intensive in that skill. I implement this by regressing:

$$\omega_s^j = b_1 + b_2 \tilde{h}_s + e$$

where ω_s^j is the constructed skill intensity of the worker's chosen occupation and $\tilde{h}_s$ is the proxy for skill endowment. I then test whether $b_2 > 0$.

For each of the skills I construct a proxy for abundance. Educational attainment is a straightforward indicator of education and knowledge. Likewise, the Census includes a measure of self-assessed English language proficiency, which I use as a measure of communication skills. The other dimensions are more limited. I use potential experience as a measure of experience and training. The Census also includes three dummy variables on disability status: I use (lack of) vision or hearing disability as a measure of physical skills, and (lack of) difficulty remembering as a measure of cognitive skills. I use the same sample as for the previous section. For the tests other than communication, I use only Americans to avoid complications such as comparing Swedish and Kenyan education; for communication, I use only foreign-born workers.

Table 1 gives the results. With the large sample, every variable is statistically significant. For communication and education, the effect is also large: these are the two best proxy measures.[10] All the coefficients have the right sign except for physical disability. This

a common component from such an analysis relates caring for others, exposure to disease, knowledge of biology and chemistry, and advanced college education requirements. These characteristics describe a pool of occupations (doctor, nurse) and not a set of related, deep technologies. It is not satisfying to say that there are so few Mexican-born doctors because Mexican immigrants lack "the medical factor". Using a human capital perspective makes results more meaningful.

[10] The coefficient for communication survives controlling for birthplace or using only Mexican immigrants (the largest single group), although both changes cut its impact by about half.

Table 1: Check on Measured Skill Intensity

Skill Dimension	Estimated b_2[a]
Education/Knowledge	0.431
Experience/Training	0.0018
Cognitive	0.068
Physical	-0.062
Communication	0.195

[a] For experience and training b_2 is the marginal effect of an additional year of potential experience. For all other variables it is the estimate of the highest category, with the lowest category omitted.

sign may be due to a reverse causality problem: workers with more physically demanding occupations may also be more likely to suffer disabilities from their work.

From these tests I conclude that the constructed measures of skill intensity and the theoretical predictions are reasonable. However, the data limitations for information on the skills of workers is binding. In the next two sections, I use the theory to back out the implied skill endowments.

4.3 Estimation as a Conditional Logit

Denote by i the country of birth of a particular immigrant. The main object of interest here is $f(H|i)$, the conditional distribution of human capital given country of birth. To make progress, I have to make assumptions about the distribution of f. The simplest assumption and estimation is that all workers from a given country have the same human capital endowment, with different occupations for workers from country i arising only because of taste differences. Under this assumption equation (12) simplifies to:

$$q(j'|i) = \frac{\exp\left[\frac{1}{1-\alpha}\log(P^{j'}) + \sum_{s=1}^{S}\omega_s^{j'}\log(\bar{h}_s^i)\right]}{\sum_{j=1}^{J}\exp\left[\frac{1}{1-\alpha}\log(P^{j}) + \sum_{s=1}^{S}\omega_s^{j}\log(\bar{h}_s^i)\right]}$$

This function has the form of the conditional logit as introduced by McFadden (1974), which is straightforward to estimate. As is standard for a conditional logit, it is not possible to estimate a full set of prices and skill endowments because of collinearity. However, I can identify a set of related parameters: $\frac{1}{1-\alpha}\log(P^j) + \sum_{s=1}^{S}\omega_s^j \log(\bar{h}_s^{US})$ for each occupation, and $\log(\bar{h}_s^i) - \log(\bar{h}_s^{US})$ for every country and skill. Note that the second set of estimated parameters is the log of the skill ratio between the average immigrant and the average American, which is exactly the object of interest. It is possible to separately identify each of the relevant parameters by imposing the restrictions of the general equilibrium model, but it does not facilitate interpretation of the empirical results. Hence, I delay doing so until Section 5. Additionally, I restrict $\frac{1}{1-\alpha}\log(P^{452}) + \sum_{s=1}^{S}\omega_s^{452}\log(\bar{h}_s^{US}) = 0$. The choice of numeraire has already pinned down prices, so this normalization pins down the level of A.

The estimates are presented in Table 6, along with their statistical significance and the number of observations for each country. Rather than discussing each of the 650 relative skill endowments separately, I identify broad trends. Given the large sample most estimates are precisely estimated and statistically different from 0 (i.e., different from U.S. endowments). Table 2 gives the net contribution of immigrants to the United States skill distribution, measured as the percentage change in the average of each skill per worker. Taken as a single group, immigrants increase the abundance of cognitive skills and increase the scarcity of communication and experience and training skills. These effects are large: immigrants represent just 8.6% of the sample, but lower the supply of communication skills by 21%.

We may suspect that these numbers mask substantial heterogeneity, particularly for immigrants who did not enter the country through legal channels and for immigrants from countries of different development status. For the former, I face the difficulty that immigrants do not intentionally reveal themselves to be unauthorized. Instead, I define an immigrant as unauthorized if they immigrate from one of the fifteen countries who are estimated to have the highest percentage of immigrants that are unauthorized, as given by Office of Policy and Planning U.S. Immigration and Naturalization Service (2003). It is not a precise estimate, but it works as a useful proxy.[11] 3.3% of immigrants are categorized as unauthorized by this definition. Unauthorized immigrants increase the abundance of physical skills but increase the scarcity of all other skills. Authorized immigrants are particularly abundant in cognitive skills, on the other hand. Both groups lack communication and experience and training.

There is a strong correlation between countries which I have classified as sources of unauthorized immigrants and less developed countries. Figures 4 and 5 plot skills against

[11]The countries are Mexico, El Salvador, Guatemala, Honduras, Dominica, Bolivia, Brazil, Colombia, Ecuador, Venezuela, Liberia, Nigeria, Sierra Leone, Kenya, and Western Samoa.

Table 2: Contribution of Immigrants to Skill Distribution

	Communication	Exp/Train	Cognitive	Physical	Education
All	-21.4%	-10.7%	7.3%	1.3%	1.3%
Authorized Only	-10.9%	-6.7%	12.0%	-3.1%	1.2%
Unauthorized Only	-10.5%	-3.9%	-4.7%	4.4%	0.1%

source country PPP GDP p.c. in 2000, taken from the World Development Indicators (World Bank 2006); income per capita is available for 117 of the 131 countries. The plotted trend lines make a point similar to Table 2. Most countries' immigrants are scarce in communication skills; there is little effect of education; and developed country immigrants have more experience and training but less physical skills. Only for cognitive ability is there a difference. While developing country immigrants in general are abundant in cognitive ability, unauthorized country immigrants are not, likely because immigrants who come through formal U.S. immigration channels are selected for cognitive ability.[12] Otherwise, I do not claim to be separating unauthorized from developing country immigrants.

These estimates are constructed assuming that all workers born in a given country are identical. The next section relaxes this assumption.

4.4 Estimation as a Mixed Logit

A more plausible alternative is to assume that immigrants from a country are drawn from a non-degenerate distribution $f(H|i)$, which is a function of (country-specific) parameters θ^i. Since country of birth but not human capital is observed, the conditional probability $q(j'|i)$ is now:

$$q(j'|i) = \int q(j'|H) f(H|\theta^i) d(H) \tag{12}$$

This equation is a standard mixed logit estimation.[13]

To be more specific, I assume that human capital is lognormally distributed with mean μ^i and diagonal variance-covariance matrix Σ^i. The integral in equation 12 is a five-dimensional normal integral, so it cannot be solved in closed form. Instead, I use Halton

[12] See for example Borjas (1999) for work on ability selection.

[13] Useful information about the issues in estimating a mixed logit can be found in Train (2003) and Hensher and Greene (2001).

draws, a form of subrandom Monte Carlo integration, to speed the numerical computation of the integral. I find that 500 draws are sufficient for integrals to converge.

I view these estimates as robustness checks on the conditional logit estimates. Do more reasonable assumptions on the distribution of skills in the population change the estimated skills of immigrants? The answer is not particularly. For four of the five skills, the μ^i from the mixed logit and the point estimates from the conditional logit line up nearly perfectly; they are correlated 0.99 or better, and the R^2 from regressing one on the other is 0.99 or better. Only for education do the mixed logit results vary in any substantive way from the conditional logit results: the correlation coefficient is just 0.9, and the R^2 of the regression is 0.8. Accounting for heterogeneity seems to matter for the estimates of each country's average education skills. In the aggregate these effects amount to little, as is shown in Table 3, which recalculates the effects of immigration on average skills per worker in the U.S. There are minor differences in a few of the categories, but none of the estimates change susbstantially. From these results I conclude that using conditional logit estimates to conduct counterfactual experiments is a reasonable step.

Table 3: Contribution of Immigrants to Skill Distribution

	Communication	Exp/Train	Cognitive	Physical	Education
All	-20.7%	-10.4%	7.5%	1.3%	-1.0%
Authorized Only	-10.3%	-6.5%	12.3%	-3.1%	-1.1%
Unauthorized Only	-10.4%	-3.9%	-4.7%	4.5%	0.1%

5 Counterfactual Experiments Using Measured Skills

The estimates from the previous section suggest immigrants raise the average level of cognitive ability and lower the average level of communications and experience/training. Further, there is substantial heterogeneity in the bundles of skills offered by immigrants from different countries, particularly along the authorized-unauthorized divide. In this section I conduct two counterfactual experiments simulating the distributions of wages that would occur if all immigrants were removed from the sample, and if only unauthorized immigrants were removed from the sample. As before I note that the second experiment corresponds closely to excluding immigrants from less developed countries, and I do not try to take a stand on exactly which effect the results represent.

Both experiments have certain common features. For convenience, I assume that $\sigma = 1$, so that hours of work is the same across different human capital endowments and countries of origin. Additionally, I assume that the equilibrium R is the same in the initial and counterfactual economies. When R is equal, comparing the two static economies is the same as comparing the steady state of two dynamic economies. Holding R fixed removes the effect of capital differences in the usual growth accounting sense, in effect assuming full adjustment of the capital stock and its distribution across occupations. The experiments also assume that workers with the appropriate skills can adjust fully across occupations. In reality although doctors and lawyers use similar skills, doctors cannot effortlessly become lawyers or vice versa. Hence, the model asks a very long-run question that might be better interpreted as the wages that would have prevailed if no immigrants had entered the country.

I fix the price of a unit of consumption as the numeraire across experiments, so wages are always real wages: they account for the fact that the prices of goods adjust and that the consumption bundle may become relatively more expensive. I use the conditional logit estimates to conduct the experiments. The experiments are easier to conduct with point estimates rather than distributions. Further, the results of the previous section show that the estimated means from the mixed logit are highly correlated with the conditional logit estimates.

The final object of interest is the parameter ψ, which governs the elasticity of substitution across occupational outputs. It also acts with the similarity of occupations in the skill-space to determine the elasticity of substitution across occupations. The best estimates of this elasticity across occupations come from an older literature surveyed by Hamermesh (1993). In particular, studies indicate that the elasticity of substitution across blue and white-collar workers is at least 2.5.[14] Hence, I use 2.5 as a plausible lower bound on the elasticity of substitution. I consider values ranging from 1.25 - that is, half the elasticity of blue and white-collar workers - up to 10. I argue that in the range of reasonable results most American-born workers will see small wage changes, but a few will always see large wage changes.

5.1 Distributional Implications

The distribution of real wage gains from removing immigrants is strongly skewed. No worker loses much; most workers are clustered tightly around a median outcome of losing a small amount; a few occupations see large real wage increases. Figure 2 gives the distribution

[14]Chiswick (1978) estimates an elasticity of 2.5 between professionals and non-professionals; Dougherty (1972) estimates a value of 4.1 across blue and white-collar workers in 8 occupation groups.

of wage gains by occupation from both experiments, with an elasticity of substitution of 2.5. There is a long right tail which is difficult to observe in the figure, but no left tail to speak of. The distributional consequences of removing all immigrants are larger than those of removing only unauthorized immigrants. This outcome is mostly due to the differences in magnitudes; unauthorized immigrants are 38.6% of the total immigrant population.

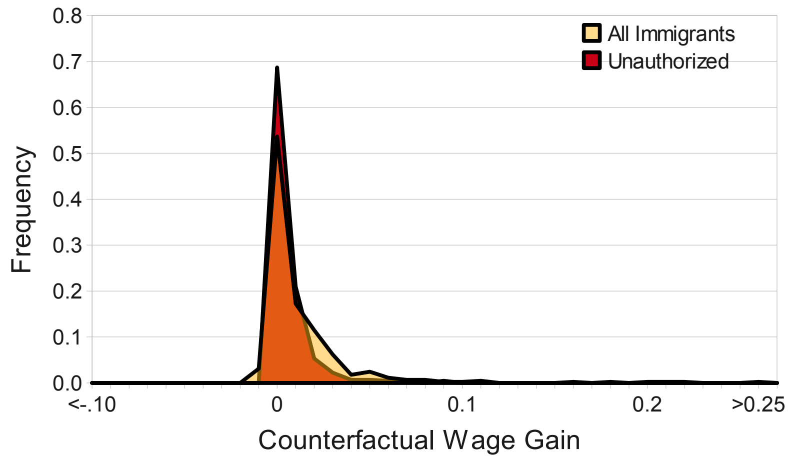

Figure 2: Distribution of Wage Changes, Unweighted

Table 4 gives several measures of the distribution of outcomes for a range of possible ψ, for both experiments. Changing ψ changes the distribution in predictable ways: larger ψ implies strictly smaller wage changes, affecting extreme and median outcomes about equally. In particular, doubling the elasticity of substitution cuts the maximum wage gain by less than half, suggesting that some American-born workers would see large wage gains even for much larger values of ψ. Figure 3 shows the full distribution for the highest and lowest values of ψ. Overall, the distributional effects suggest that immigration has very small positive effects for most workers, but large negative effects for workers in a few occupations or with certain types of skills. In the next section, I study the characteristics of those occupations with higher and lower wages.

5.2 Identifying Which Occupations Gain

Finally, what are the characteristics of occupations that gain and lose from these experiments? Qualitatively, a broad set of occupations stand to lose from the counterfactual removal of immigrants, but particularly those in communications-intensive occupations such as management and services, and those in trade occupations such as locksmiths, riggers,

Table 4: Wage Changes for Different Elasticities

	ψ				
Remove All Immigrants	1.25	2.5	5	7.5	10
Min	-1.65%	-1.19%	-0.76%	-0.56%	-0.44%
Max	139.30%	87.80%	49.96%	34.79%	26.66%
Median	-0.27%	-0.19%	-0.12%	-0.09%	-0.07%
Median Absolute	0.98%	0.71%	0.45%	0.34%	0.27%
Remove Unauthorized Immigrants					
Min	-0.35%	-0.25%	-0.15%	-0.11%	-0.09%
Max	25.50%	17.62%	10.71%	7.53%	5.79%
Median	-0.19%	-0.13%	-0.08%	-0.06%	-0.05%
Median Absolute	0.27%	0.19%	0.12%	0.09%	0.07%

Figure 3: Distribution of Wage Changes, Elastic and Inelastic

and boilermakers. The communications and certification/training requirements of these occupations insulate them from immigrants, so they mostly see the higher price effects. Those who gain most fit into two fairly narrow categories: science professionals and manual laborers. For instance, aerospace engineers, astronomers, dietitians, and medical scientists have the 1st, 3rd, 6th, and 11th largest wage gains; garment pressers, tire builders, production helpers, and shoe machine operators are 2nd, 4th, 7th, and 8th. These occupations are intensive in cognitive ability or physical skills but not communications, and immigrants cause large wage declines. The results for removing only unauthorized immigrants are similar, except that cognitive ability-intensive occupations no longer gain. One striking fact stands out: the fraction of an occupation's labor force that is foreign born is only weakly correlated with large wage gains, because of the potential for reallocation. Hence, some occupations with over a quarter of the work force foreign born still see wage effects of less than 1% for $\psi = 2.5$, including diverse occupations such as taxi drivers, chefs, and economists.

Table 5 is an attempt to make this casual information more systematic. In it, I regress each occupation's wage change from the two experiments on the occupation's skill intensity along the five dimensions, and the fraction of the occupation's labor force that is removed under the given experiment. There are sizeable effects for some of the skill attributes, particularly cognitive ability and communications. Recall that the skill intensity variables are scaled to lie on $[0, 1]$. Then the difference between being the cognitively least and most intensive occupations is a 17.7% wage gain; for communications, the difference is a 15.1% wage decline. The results quickly summarize that immigrants increase the average supply of communications and experience and training and decrease the average supply of communications. Note also that the fraction of the work force that is removed plays no predictive role in determining subsequent wage gains, in a crude OLS sense.

Table 5: Wage Changes for Different Elasticities

	Education	Physical	Cognitive	Exp/Train	Comm	Removed
All Immigrants	1.28%	-3.66%	17.74%	-4.30%	-15.11%	0.32%
	(1.23%)	(0.90%)	(1.74%)	(1.30%)	(1.42%)	(3.18%)
Unauthorized	1.44%	-0.17%	0.54%	-1.59%	-5.11%	1.15%
	(0.36%)	(0.26%)	(0.51%)	(0.38%)	(0.41%)	(1.41%)

Standard errors are in parentheses.

The identities of winners and losers from immigration fits well with recent research, although my results are at a more disaggregated level. Ottaviano and Peri (2006) find that with capital adjustment, immigration should have little wage effects; here, the median workers see only small effects. Peri and Sparber (2008b) find that immigration induces American workers to specialize in interative occupations - similar to my communications-intensive occupations. My findings are subject to the caveat that for some workers and occupations, there are no good substitutes available: the occupations similar to aerospace engineer are also not communications-intensive. They also find interesting results about how new cohorts of immigrants impact the wages of older cohorts, which I do not disentangle. Finally, Peri and Sparber (2008a) and Borjas (2005) both show that high-skill immigration affects the wages and career decisions of high-skilled Americans: again, the theme is that it pushes them to study "soft" subjects in graduate school, which are more writing, language, and communications-intensive. My results add to this previous literature that occupations with formal experience and training requirements are also effective for avoiding competition from immigrants; and that the formal U.S. selection mechanism has resulted in a large net inflow of cognitive abilities.

6 Conclusion

This paper has proposed a simple theory of labor markets where workers vary in their endowment of a vector of skills, and occupations vary in their intensity over the vector of skills. Comparative advantage leads workers to match their endowments to occupations that are appropriately skill-intensive. I use the model to estimate the human capital endowments of workers born in 131 countries over 5 dimensions. Immigrants are net suppliers of cognitive ability and physical skills, but are scarce in experience/training and particularly communications skills. They cause a highly skewed impact to the distribution of wages, reflecting their contributions to the skill distribution. Americans who can substitute to similar occupations that are communications-intensive and mitigate immigrant wage pressure; those who cannot are subject to the large wage declines.

The moderate wage effects estimated here are almost certainly too large. They ignore, for instance, the ability of Americans to export excess goods as predicted in a Heckscher-Ohlin framework - not all aerospace engineering services are consumed in the United States. They also assume that the endowments of Americans are fixed, but as Peri and Sparber (2008a) and Borjas (2005) have shown, Americans change their schooling and human capital accumulation decisions as well. The numbers presented here are probably generous upper bounds.

A skewed distribution of wage impacts naturally suggests political economy stories for government policy with respect to immigration, particularly with respect to unauthorized immigrants and the highly-skilled immigrants in the science occupations. For example, it may help explain why the allocation of H1-B visas is set "low". This question is left for future research.

References

Autor, D. H., F. Levy, and R. J. Murnane (2003): "The Skill Content of Recent Technological Change: An Empirical Exploration," *The Quarterly Journal of Economics*, 118(4), 1279–1333.

Borjas, G. J. (1999): "The Economic Analysis of Immigration," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3A, pp. 1697–1760. Elsevier Science, North-Holland Publishers.

——— (2005): "The Labor-Market Impact of High-Skill Immigration," *The American Economic Review*, 95(2), 56–60.

Bureau of Labor Statistics (2004): "Occupational Employment Statistics," Available online at `http://www.bls.gov/oes/oes_2004_m.htm`.

Chiswick, C. (1978): "The Growth of Professional Occupations in U.S. Manufacturing, 1900-73," in *Research in Human Capital and Development*, ed. by I. Sirageldin. JAI Press, Greenwich, Conn.

Dougherty, C. R. S. (1972): "Estimates of Labor Aggregation Functions," *Journal of Political Economy*, 80, 1101–1119.

Gathmann, C., and U. Schonberg (2008): "How General is Human Capital? A Task-Based Approach," Working Paper, Stanford University.

Hamermesh, D. (1993): *Labor Demand*. Princeton University Press.

Hensher, D. A., and W. H. Greene (2001): "The Mixed Logit Model: The State of Practice and Warnings for the Unwary," Mimeo, New York University.

Lazear, E. P. (2003): "Firm-Specific Human Capital: A Skill-Weights Approach," NBER Working Paper No. w9679.

McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Analysis," in *Frontiers in Econometrics*, ed. by P. Zarembka, pp. 105–142. New York: Academic Press.

Occupational Information Network (O*NET) and US Department of Labor/Employment and Training Administration (USDOL/ETA) (2007): "Database 12.0," Available online at `http://www.onetcenter.org/overview.html`.

Office of Policy and Planning U.S. Immigration and Naturalization Service (2003): "Estimates of the unauthorized immigrant population residing in the United States: 1990 to 2000," Available online at `http://www.dhs.gov/xlibrary/assets/statistics/publications/Ill\_Report\_1211.pdf`.

Ottaviano, G. I., and G. Peri (2006): "Rethinking the Effects of Immigration on Wages," Working Paper.

Peri, G., and C. Sparber (2008a): "Highly-Educated Immigrants and Native Occupational Choice," Working Paper.

——— (2008b): "Task Specialization, Comparative Advantages, and the Effects of Immigration on Wages," Working Paper.

Ruggles, S., M. Sobek, T. Alexander, C. A. Fitch, R. Goeken, P. K. Hall, M. King, and C. Ronnander (2004): "Integrated Public Use Microdata Series: Version 3.0 [Machine-readable database]," Minneapolis, MN: Minnesota Population Center [producer and distributor], http://www.ipums.org.

Schoellman, T. (2008): "The Causes and Consequences of Cross-Country Differences in Schooling Attainment," Working Paper, Clemson University.

Spitz-Oener, A. (2006): "Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure," *Journal of Labor Economics*, 24(2), 235–270.

Train, K. (2003): *Discrete Choice Models with Simulation*. Cambridge University Press.

U.S. Department of Labor, Employment, and Training Administration (1991): "Dictionary of Occupational Titles: Revised Fourth Edition," Washington DC: 1991.

World Bank (2006): *World Development Indicators*.

A Measures of Skill Intensity

A.1 Information Used

The O*NET database is built on a content model that divides occupational information into six broad categories: worker characteristics, worker requirements, experience requirements, occupation-specific information, workforce characteristics, and occupational requirements. Within each of these six broad categories information is organized in a hierarchical format similar to the 1-digit, 2-digit, 3-digit format of industry and trade data. For instance, item 1.A.1.a.1 is a 5-digit characteristic of occupations, going from general to specific: Worker Characteristics.Ability.Cognitive Abilities.Verbal Abilities.Oral Comprehension. Throughout, I use the most disaggregated data possible, which can be 3 to 6-digit information.

Data are provided for each category and occupation, and is typically normalized to a 0-7 scale. O*NET provides anchors that represent typical characteristics associated with particular scores. For example, Oral Comprehension is computed on a scale of 0-7. The anchors given are that a score of 2 is equivalent to ability to understand a television commercial; a score of 4 is equivalent to ability to understand a coach's oral instructions for a sport; and a score of 6 is equivalent to ability to understand a lecture on advanced physics. Scores for each occupation-attribute are gathered either from the average score given by occupational analysts or the average score given by survey responses from incumbent workers. For instance, all oral comprehension scores are the average rating of eight analysts, while the mathematics skills score for chief executives is the average of 23 survey responses by actual chief executives.

From the 250+ most disaggregated categories I select those that correspond closely to one of the five skills. I also focus on information that is relatively unique to a specific skill. The reported level anchors are helpful here. For example, I exclude oral comprehension ability because it is not clear from the anchors provided whether it measures a cognitive ability, a communication skill, or a mixture. I use principal component analysis to aggregate the different measures into a single skill intensity for each dimension. I keep only the first component, which accounts for 36-82% of the total variation of the variables. I denote with a * variables that have at least one-third of their variation accounted for by the principal component, indicating that they are well-represented in the resulting skill intensity measure. This criteria produces similar results to the common technique of identifying variables that have factor loadings exceeding a threshold of 35 or 40. For each of the five dimensions, I also identify the three occupations that score as the most skill-intensive, and the three that score as the least skill-intensive. No occupation is repeated on this list, and more generally no cross-intensity correlation exceeds 0.60, implying sufficient variation to identify the skill components separately.

In addition to the baseline intensity measures used in the text, I also try those excluding education (since it seems to have little effect on the results); using simple average instead of PCA to aggregate results; not weighting the underlying occupation characteristics by the size of the workforce; and using a broader or narrower set of factors than is considered here. The qualitative results for these alternatives are similar. Finally, I conduct experiments using the first five factors generated on all the data, without imposing the human capital categorizations. These experiments suggest broader wage distributions with fatter tails, but lack interpretability as mentioned in the text.

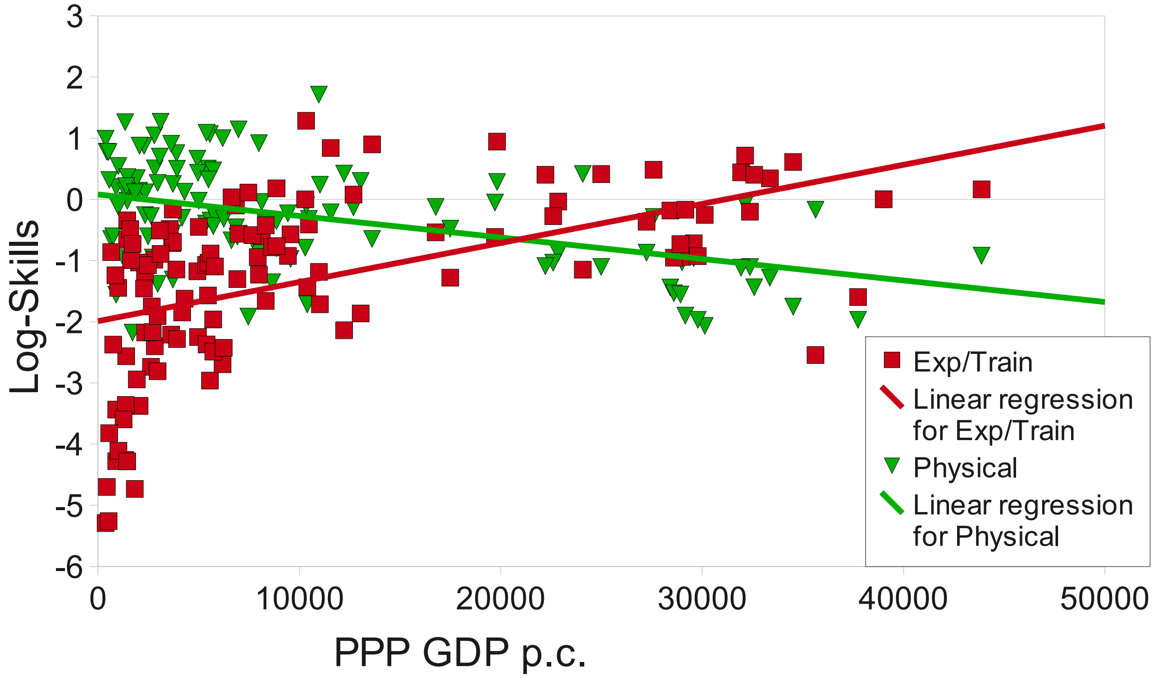

Figure 4: Skills - GDP Relationship

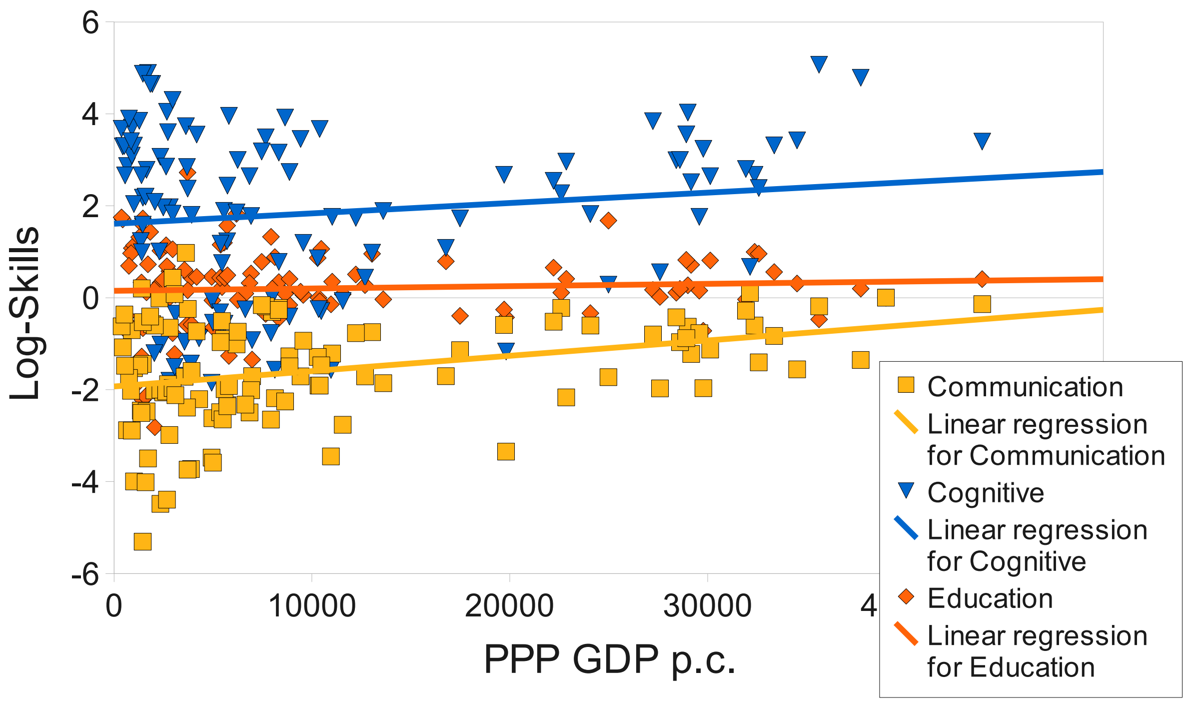

Figure 5: Skills - GDP Relationship

Table 6: Estimated Human Capital, Conditional Logit

Country	Obs	Communication	Exp/Train	Cognitive	Physical	Education
United States	5285011	0	0	0	0	0
Puerto Rico	12676	-2.00 *	-1.10 *	-0.27 *	0.19 *	0.69 *
Canada	10894	-0.60 *	-0.20 *	2.67 *	-1.11 *	1.00 *
Bermuda	165	1.05 *	0.90 *	-1.66 *	-0.49 *	-0.20 *
Cape Verde	447	-4.49 *	-1.22 *	-1.02 *	-0.26 *	0.32 *
Mexico	136866	-3.45 *	-1.18 *	-1.57 *	1.71 *	-0.14 *
Belize/British Honduras	685	-0.51 *	-1.57 *	0.75 *	0.50 *	0.23 *
Costa Rica	1250	-2.19 *	-0.60 *	-1.57 *	-0.04 *	0.88 *
El Salvador	14825	-3.48 *	-1.18 *	-1.85 *	0.66 *	0.45 *
Guatemala	8707	-3.73 *	-1.15 *	-1.43 *	0.75 *	0.44 *
Honduras	5238	-2.99 *	-0.98 *	-1.81 *	1.05 *	0.58 *
Nicaragua	3384	-2.02 *	-1.03 *	-1.21 *	0.14 *	0.17 *
Panama	1787	-0.22 *	-1.66 *	0.78 *	-0.39 *	0.34 *
Cuba	10009	-1.12 *	-0.28 *	-0.77 *	0.35 *	-0.55 *
Dominican Republic	9399	-2.62 *	-2.25 *	-0.05 *	0.44 *	-0.06 *
Haiti	7832	-2.46 *	-4.26 *	1.23 *	1.26 *	1.63 *

Continued on Next Page

Table 6: Estimated Human Capital, Conditional Logit

Country	Obs	Communication	Exp/Train	Cognitive	Physical	Education
Jamaica	9882	-0.90 *	-2.96 *	1.89 *	1.07 *	1.20 *
Antigua-Barbuda	355	-0.75 *	-1.86 *	0.98 *	0.30 *	0.96 *
Bahamas	336	-0.25 *	-1.69 *	0.80 *	-0.04 *	0.92 *
Barbados	933	-1.15 *	-2.22 *	0.48 *	0.37 *	1.64 *
Dominica	312	-2.34 *	-1.96 *	1.23 *	0.47 *	1.57 *
Grenada	538	-1.02 *	-2.69 *	1.85 *	1.00 *	1.82 *
St. Kitts-Nevis	224	-1.22 *	-1.71 *	1.76 *	0.24 *	0.35 *
St. Lucia	259	-0.31 *	-1.23 *	0.02	0.92 *	0.20 *
St. Vincent	369	-0.97 *	-2.37 *	1.18 *	1.09 *	1.15 *
Trinidad & Tobago	3542	-0.77 *	-2.14 *	1.74 *	0.42 *	0.51 *
Argentina	2173	-1.29 *	0.00	0.86 *	-0.79 *	0.86 *
Bolivia	994	-1.72 *	-0.48 *	-0.96 *	-0.69 *	0.60 *
Brazil	4329	-2.65 *	-0.94 *	-0.77 *	-0.83 *	1.33 *
Chile	1442	-1.47 *	-0.41 *	-0.29 *	-0.32 *	1.06 *
Colombia	8987	-2.49 *	-1.07 *	-0.32 *	-0.39 *	0.43 *
Ecuador	4964	-2.65 *	-1.03 *	-0.62 *	0.32 *	-0.57 *
Guyana/British Guiana	3838	-0.65 *	-2.40 *	1.97 *	0.50 *	0.01
Paraguay	215	-3.74 *	-0.16 *	-1.70 *	-1.31 *	2.72 *
Peru	5495	-1.99 *	-0.88 *	-0.56 *	-0.35 *	0.43 *
Uruguay	484	-1.28 *	0.18 *	-0.41 *	-0.36 *	-0.16 *
Venezuela	1645	-0.94 *	-0.57 *	1.19 *	-0.97 *	0.06 *
Denmark	511	-0.28 *	0.44 *	2.80 *	-1.12 *	-0.04 *
Finland	364	-0.80 *	-0.37 *	3.84 *	-0.87 *	0.17 *
Norway	386	-0.14 *	0.16 *	3.39 *	-0.92 *	0.40 *
Sweden	879	-0.63 *	-0.94 *	4.03 *	-1.03 *	0.27 *
United Kingdom	11346	-0.43 *	-0.18 *	2.99 *	-1.44 *	0.11 *
Northern Ireland	2783	0.11 *	0.72 *	0.67 *	-0.07 *	0.17 *
Belgium	399	-1.12 *	-0.25 *	2.64 *	-2.07 *	0.81 *
France	2477	-1.22 *	-0.17 *	2.51 *	-1.90 *	0.71 *
Netherlands	1159	-0.83 *	0.34 *	3.31 *	-1.28 *	0.56 *
Switzerland	723	-1.56 *	0.61 *	3.42 *	-1.75 *	0.31 *
Albania	660	-2.21 *	-1.62 *	-0.88 *	0.13 *	-0.73 *
Greece	2231	-0.61 *	-1.15 *	1.82 *	0.42 *	-0.34 *
Macedonia	329	-1.71 *	-0.56 *	-0.92 *	1.14 *	-1.35 *
Italy	4182	-1.97 *	0.48 *	0.55 *	-0.29 *	0.02 *
Portugal	2638	-3.35 *	0.94 *	-1.17 *	0.29 *	-0.43 *
Azores	296	-3.40 *	0.41 *	-1.47 *	0.85 *	-0.79 *
Spain	1453	-1.73 *	0.41 *	0.29 *	-1.10 *	1.68 *
Austria	533	-1.41 *	0.40 *	2.39 *	-1.44 *	0.96 *

Continued on Next Page

Table 6: Estimated Human Capital, Conditional Logit

Country	Obs	Communication	Exp/Train	Cognitive	Physical	Education
Bulgaria	721	-2.01 *	-1.30 *	1.78 *	-0.61 *	0.53 *
Czechoslovakia	506	-2.00 *	0.32 *	1.22 *	-0.46 *	0.23 *
Slovakia	260	-1.71 *	0.08 *	0.43 *	-0.13 *	0.27 *
Czech Republic	414	-1.71 *	-0.54 *	1.08 *	-0.12 *	0.79 *
Germany	9144	-0.77 *	-0.72 *	1.76 *	-0.96 *	0.15 *
Hungary	1016	-1.86 *	0.90 *	1.89 *	-0.65 *	-0.04 *
Poland	7841	-2.76 *	0.84 *	-0.07 *	-0.21 *	0.07 *
Romania	2264	-2.50 *	-0.10 *	2.64 *	-0.46 *	0.32 *
Yugoslavia	1230	-2.17 *	0.21 *	-0.23 *	-0.33 *	-0.40 *
Croatia	642	-1.91 *	1.28 *	-0.25 *	-0.10 *	0.00 *
Serbia	173	-2.33 *	0.03 *	-0.25 *	-0.67 *	0.09 *
Bosnia	1846	-3.59 *	-0.45 *	-0.54 *	-0.02 *	-0.65 *
Kosovo	150	-2.04 *	-1.16 *	-1.91 *	0.24 *	-0.67 *
Latvia	204	-2.26 *	-0.78 *	3.92 *	-1.35 *	0.09 *
Lithuania	262	-1.71 *	-0.92 *	3.45 *	-0.23 *	0.13 *
Byelorussia	593	-1.92 *	-1.10 *	3.96 *	-0.27 *	-1.27 *
Moldovia	313	-1.44 *	-0.34 *	2.18 *	-0.02 *	-0.66 *
Ukraine	3915	-2.39 *	-0.72 *	2.84 *	-0.55 *	-0.59 *
Armenia	821	-0.01 *	-1.45 *	1.01 *	0.88 *	-0.59 *
Azerbaijan	220	-2.06 *	-1.07 *	1.96 *	-0.60 *	0.39 *
Georgia	163	-2.00 *	-2.17 *	3.07 *	0.12 *	1.02 *
Uzbekistan	299	-2.47 *	-0.99 *	2.78 *	-0.82 *	0.12 *
China	19090	-4.39 *	-1.75 *	4.05 *	-1.81 *	0.68 *
Hong Kong	3327	-1.97 *	-0.93 *	3.24 *	-1.97 *	-0.72 *
Taiwan	6439	-1.83 *	-0.23 *	4.81 *	-2.28 *	-0.39 *
Japan	5764	-0.97 *	-0.95 *	3.00 *	-1.53 *	0.19 *
South Korea	2000	-1.14 *	-1.28 *	1.72 *	-0.48 *	-0.40 *
Cambodia	1911	-4.00 *	-1.45 *	2.04 *	-0.12 *	-1.94 *
Indonesia	1150	-1.88 *	-2.17 *	3.60 *	-0.95 *	0.07 *
Laos	2531	-5.31 *	-0.66 *	1.59 *	0.13 *	-2.22 *
Malaysia	1048	-1.91 *	-1.43 *	3.67 *	-1.72 *	-0.08 *
Philippines	29294	-2.04 *	-2.74 *	2.85 *	-0.27 *	1.14 *
Singapore	393	-1.35 *	-1.60 *	4.79 *	-1.97 *	0.20 *
Thailand	2355	-2.36 *	-2.49 *	2.44 *	-0.44 *	0.49 *
Vietnam	17344	-4.01 *	-0.49 *	2.20 *	-0.72 *	-2.15 *
Afghanistan	631	0.39 *	-3.32 *	3.18 *	0.96 *	-1.23 *
India	23130	-3.50 *	-0.73 *	4.89 *	-2.18 *	0.73 *
Bangladesh	1681	-0.71 *	-4.28 *	3.70 *	0.20 *	-0.64 *
Burma (Myanmar)	718	-2.88 *	-0.86 *	2.87 *	-0.65 *	-0.39 *

Continued on Next Page

Table 6: Estimated Human Capital, Conditional Logit

Country	Obs	Communication	Exp/Train	Cognitive	Physical	Education
Pakistan	4114	-0.57 *	-2.94 *	4.65 *	0.35 *	-0.60 *
Sri Lanka (Ceylon)	642	-1.96 *	-1.91 *	4.30 *	-1.38 *	1.05 *
Iran	5388	-0.19 *	-0.58 *	3.49 *	-0.65 *	-0.35 *
Nepal	258	-2.89 *	-3.43 *	3.09 *	-1.56 *	1.08 *
Cyprus	185	-2.17 *	-0.03 *	2.96 *	-0.87 *	0.41 *
Iraq	1381	-1.46 *	-0.87 *	1.67 *	0.32 *	-0.76 *
Israel/Palestine	1861	-0.22 *	-0.27 *	2.27 *	-1.03 *	0.11 *
Jordan	787	0.98 *	-2.21 *	3.73 *	0.91 *	-1.72 *
Kuwait	248	-0.19 *	-2.54 *	5.07 *	-0.17 *	-0.47 *
Lebanon	1860	-0.26 *	-0.43 *	3.16 *	-0.42 *	-0.44 *
Saudi Arabia	139	-0.59 *	-0.61 *	2.67 *	-0.05 *	-0.26 *
Syria	941	-0.24 *	-0.69 *	2.37 *	0.25 *	0.16 *
Turkey	1456	-1.49 *	-0.76 *	2.73 *	-0.80 *	0.41 *
Yemen Arab Republic (North)	253	-0.61 *	-3.37 *	2.09 *	0.88 *	-2.82 *
Algeria	240	-0.74 *	-2.43 *	2.99 *	-0.32 *	-0.05 *
Egypt/United Arab Republic	2232	-0.73 *	-1.85 *	3.55 *	-0.30 *	0.45 *
Morocco	803	0.43 *	-2.81 *	1.83 *	0.27 *	-0.77 *
Sudan	324	-2.51 *	-3.36 *	2.67 *	0.13 *	0.33 *
Ghana	1625	-1.53 *	-4.11 *	3.31 *	0.55 *	1.14 *
Liberia	784	-1.08 *	-4.70 *	3.29 *	0.78 *	1.70 *
Nigeria	3317	-0.53 *	-4.28 *	4.87 *	0.36 *	1.72 *
Senegal	212	0.20 *	-2.56 *	0.99 *	0.22 *	-1.28 *
Sierra Leone	507	-0.63 *	-5.29 *	3.68 *	0.99 *	1.75 *
Ethiopia	1463	-0.37 *	-5.26 *	3.30 *	0.78 *	-0.68 *
Kenya	863	-1.48 *	-3.59 *	3.85 *	-0.96 *	1.32 *
Somalia	452	-2.45 *	-3.84 *	1.58 *	0.72 *	-0.62 *
Tanzania	268	-2.03 *	-1.24 *	3.41 *	-1.26 *	0.97 *
Uganda	306	-1.74 *	-2.37 *	3.90 *	-0.60 *	0.69 *
Zimbabwe	247	-1.10 *	-1.28 *	3.75 *	-0.82 *	0.78 *
Eritrea	372	-1.48 *	-3.82 *	2.66 *	0.32 *	-0.43 *
Cameroon	283	-0.41 *	-4.73 *	4.65 *	0.11 *	1.43 *
South Africa (Union of)	1308	-0.16 *	0.11 *	3.18 *	-1.92 *	0.78 *
Australia	1227	-0.88 *	-0.73 *	3.55 *	-1.56 *	0.82 *
New Zealand	560	-0.52 *	0.40 *	2.54 *	-1.09 *	0.65 *
Fiji	593	-1.59 *	-2.28 *	1.80 *	0.50 *	-0.58 *
Tonga	288	-2.12 *	-0.89 *	-0.35 *	1.27 *	0.07 *
Western Samoa	254	0.08 *	-0.51 *	-1.50 *	0.70 *	-1.22 *

Note: All values are estimates of the difference in log-skills between that country and the United States. A * denotes significance at the 95% level. Obs is the number of observations in the 5% sample of the 2000 U.S. Census meeting the sample criteria for that country.

Table 7: Dimensions of Human Capital: Education and Knowledge

Measure[a]	Intensity Ranking[b]
Knowledge Category	Most Intensive
Engineering and Technology	1. Physicians and Surgeons
Design	2. Miscellaneous Social Scientists
Mathematics	3. Psychologists
Physics	
Chemistry	Least Intensive
Biology*	1. Food and Tobacco Machine Operator/Tender
Psychology*	2. Taxi Driver and Chauffeur
Sociology*	3. Desktop Publishers
Geography	
Medicine and Dentistry*	
Therapy and Counseling*	
Foreign Language*	
Fine Arts	
History and Archaelogy*	
Philosophy and Theology*	
Law and Government*	
Other Category	
Required Education Level*	

[a] Name of measure in O*NET system. An asterisk indicates that the first principal component captures at least 1/3 of the variation in the measure.

[b] Three occupations that score highest and lowest for skill intensity.

Table 8: Dimensions of Human Capital: Training and Experience

Measure[a]	Intensity Ranking[b]
Training and Experience Required	Most Intensive
On-the-Job Training*	1. Elevator Installers and Repairers
Required Work Experience*	2. Ship Engineers
On-Site/In-Plant Training*	3. Podiatrists
General Preparation	
	Least Intensive
Observed Job Experience	1. Ushers, Lobby Attendants, and Ticker Takers
< 1 Year*	2. Telemarketers
1-5 Years*	3. Dishwashers
6-9 Years	
10+ Years*	

[a] Name of measure in O*NET system. An asterisk indicates that the first principal component captures at least 1/3 of the variation in the measure.

[b] Three occupations that score highest and lowest for skill intensity.

Table 9: Dimensions of Human Capital: Cognitive Abilities

Measure[a]	Skill Intensity[b]
Worker Abilities	Most Intensive
Fluency of Ideas*	1. Aerospace Engineers
Originality*	2. Astronomers and Physicists
Problem Sensitivity*	3. Mechanical Engineers
Deductive Reasoning*	
Inductive Reasoning*	Least Intensive
Information Ordering*	1. Miscellaneous Construction Equipment Operators
Category Flexibility*	2. Laborers and Freight/Stock/Materials Movers, Hand
	3. Grinding Tool Setters/Operators/Tenders

[a] Name of measure in O*NET system. An asterisk indicates that the first principal component captures at least 1/3 of the variation in the measure.

[b] Three occupations that score highest and lowest for skill intensity.

Table 10: Dimensions of Human Capital: Physical Abilities

Measure[a]	Intensity Ranking[b]
Ability	Most Intensive
Arm-Hand Steadiness*	1. Fire Fighters
Manual Dexterity*	2. Electricians
Finger Dexterity*	3. Emergency Medical Technicians and Paramedics
Control Precision*	
Multilimb Coordination*	Least Intensive
Response Orientation*	1. Public Relations Specialist
Rate Control*	2. Actuaries
Reaction Time*	3. Loan Counselors and Officers
Wrist-Finger Speed*	
Speed of Limb Movement*	
Static Strength Ability*	
Explosive Strength	
Dynamic Strength*	
Trunk Strength*	
Stamina*	
Extent Flexibility*	
Dynamic Flexibility	
Gross Body Coordination*	
Gross Body Equilibrium*	
Near Vision	
Far Vision	
Visual Color Discrimination*	
Night Vision*	
Peripheral Vision*	
Depth Perception*	
Glare Sensitivity*	
Hearing Sensitivity*	
Auditory Attention*	

[a] Name of measure in O*NET system. An asterisk indicates that the first principal component captures at least 1/3 of the variation in the measure.

[b] Three occupations that score highest and lowest for skill intensity.

Table 11: Dimensions of Human Capital: Language and Communication

Measure[a]	Intensity Ranking[b]
Frequency of Communication by Method	Most Intensive
Public Speaking*	1. Gaming Managers
Telephone*	2. Postmasters and Mail Superintendents
Letters and Memos*	3. Public Relations Specialists
Face-to-Face Discussions*	
	Least Intensive
Frequency of Communication by Type	1. Pressers, Textile, Garment, and Related Materials
Contact with Others*	2. Tire Builders
Work with Group or Team*	3. Shoe Machine Operators and Tenders
Deal with External Customers*	

[a] Name of measure in O*NET system. An asterisk indicates that the first principal component captures at least 1/3 of the variation in the measure.

[b] Three occupations that score highest and lowest for skill intensity.