Preliminary:
Comments welcome.

Workplace Segregation in the United States: Race, Ethnicity, and Skill*

Judith Hellerstein
Department of Economics and MPRC,
University of Maryland,
and NBER

David Neumark
Public Policy Institute of California,
and NBER

December 2004

*Neumark is also a Research Fellow at IZA, and a Visiting Scholar at the Institute of Business and Economic Research, UC-Berkeley. This is a substantially revised version of an earlier version of this paper by the same name. This research was funded by the Russell Sage Foundation and NIH. We are grateful to Megan Brooks, Joel Elvery, Gigi Foster, and Melissa Powell for outstanding research assistance, to Seth Sanders for helpful discussions, and to seminar participants at the Public Policy Institute of California, the Federal Reserve Board, the NBER Summer Institute, and the Color Lines Conference at Harvard University, August 2003. This paper reports the results of research and analysis undertaken while the authors were research affiliates at the Center for Economic Studies at the U.S. Census Bureau. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. It has been screened to ensure that no confidential information is revealed. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau, the Public Policy Institute of California, or the Russell Sage Foundation.

Workplace Segregation in the United States: Race, Ethnicity, and Skill

Abstract: We study workplace segregation in the United States using a unique matched employer-employee data set that we have created. We present measures of workplace segregation by education and language–as skilled workers may be more complementary with other skilled workers than with unskilled workers–and by race and ethnicity, using simulation methods to measure segregation beyond what would occur randomly as workers are distributed across establishments. We also assess the role of education- and language-related skill differentials in generating workplace segregation by race and ethnicity, as skill is often correlated with race and ethnicity. Finally, we attempt to distinguish between segregation by skill based on general crowding of unskilled poor English speakers into a narrow set of jobs, and segregation based on common language for reasons such as complementarity among workers speaking the same language.

Our results indicate that there is considerable segregation by education and language in the workplace. Racial segregation in the workplace is of the same order of magnitude as education segregation, and segregation between Hispanics and whites is larger yet. Only a tiny portion of racial segregation in the workplace is driven by education differences between blacks and whites, but a substantial fraction of ethnic segregation in the workplace can be attributed to differences in language proficiency.

I. Introduction

Wage differentials by education, race, and ethnicity in the United States have been extensively documented. When it comes to wage differentials by education, the past two decades have generally been marked by increased returns to education, the extent and sources of which have been the subject of much discussion (see, e.g., Katz and Murphy, 1992; Juhn, et al., 1992; Card and DiNardo, 2002; Autor, et al., 2004). As for wage differences by race and ethnicity (as documented in, e.g., Donohue and Heckman, 1991; Cain, 1986; Altonji and Blank, 1999; Welch, 1990; and Ihlanfeldt and Sjoquist, 1990), there has been extensive research trying to uncover their sources. Most researchers agree that skill differences such as education (including its quality) and language account for sizable shares of wage gaps by race and ethnicity (e.g., O'Neill, 1990; Trejo, 1997), with the sharper dispute whether gaps in these and other skills (such as those captured in test scores) fully explain these wage gaps or whether discrimination contributes as well (e.g., Darity and Mason, 1998; Neal and Johnson, 1996).

In contrast to this vast literature on wage differences, much less is known about the extent and sources of segregation in the labor market. There has been speculation that one source of increased wage inequality by education is increased segregation by skill (e.g. Kremer and Maskin, 1996), but there is little evidence of the extent of labor market segregation by education in the first place with which to test this hypothesis. Moreover, while there is widespread agreement that there is labor market segregation by race and ethnicity, and that this segregation accounts–at least in a statistical sense–for a sizable share of wage gaps between white males and other demographic groups (e.g., Carrington and Troske, 1998; Bayard, et al., 1999; King, 1992; Watts, 1995; Higgs, 1977), there has been very little work trying to uncover whether this segregation is due to discrimination or other sources.[1]

---

[1] This segregation may occur along industry and occupation lines, as well as at the more detailed level of the establishment or job cell (occupations within establishments). For example, Bayard, et al. (1999) found that, for men, job cell segregation by race accounts for about half of the black-white wage gap and a larger share of the Hispanic-white wage gap. Carrington and Troske (1997) use data sets much more limited in scope than the one we use here, to examine workplace segregation by race and sex. In general, the paucity of research on workplace segregation is presumably a function of the lack of data linking workers to establishments.

Workplace segregation by education, or by skill more generally, and workplace segregation by race and ethnicity have the potential to be intimately connected. There are numerous models suggesting that employers may segregate workers across workplaces by skill, most likely because of complementarities among workers with more similar skills. Because in U.S. labor markets skill is often correlated with race and ethnicity, an untended effect of profit-maximizing skill segregation in the workplace may be segregation along racial and ethnic lines.[2] Alternatively, race and ethnic segregation in the workplace may be a function of discrimination in the labor market. Perhaps the most convincing evidence of discrimination in employment comes from audit studies of hiring (Cross, et al., 1990; Turner, et al., 1991), although this work does not speak to segregation per se.[3]

This paper has two goals: to use a new matched employer-employee data set to provide the best available measurements of workplace segregation by education, language, race, and ethnicity in the United States; and to present evidence that helps in understanding the sources of this segregation, in particular the role of skill in generating race and ethnic segregation. We pursue these goals using the 1990 Decennial Employer-Employee Database (DEED), a unique data set that we have created. The 1990 DEED is based on matching records in the 1990 Decennial Census of Population to a Census Bureau list of most business establishments in the United States.[4] The matching yields data on multiple workers matched to establishments, providing the means to measure workplace segregation in the United States based on a large, fairly representative data set. In addition, the reliance on the Decennial Census of Population as the source of information on workers creates the capacity to link information on workplace segregation to information on other characteristics of workers. This allows us to examine the

---

[2] On the supply side, labor market networks can also generate workplace segregation; we do not focus on labor market networks in this paper.

[3] Heckman (1998) notes that even if there is hiring discrimination–as audit studies suggest–whether or not a wage differential arises depends on the discriminatory behavior of the marginal rather than the average employer. Black (1995) shows that in a search model discriminatory tastes on the part of some employers can result in a wage gap, even when the discriminatory employers do not hire any minorities.

[4] The 1990 Census of Population is currently the most recent Decennial Census available for matching. The 2000 data have just become available for matching, and we are now beginning the detailed process of constructing a 2000 DEED.

extent of segregation in the workplace by skill, and to examine the impact of skill segregation in generating segregation by race and ethnicity. Thus, the DEED provides unparalleled opportunities to study workplace segregation by race, ethnicity, and skill.

Our empirical analysis proceeds in three steps that exploit these various characteristics of the DEED. First, we present measures of workplace segregation in the United States, focusing on segregation along the lines of education, language, race, and ethnicity.[5] Rather than considering all deviations from proportional representation across establishments as an "outcome" or "behavior" to be explained, we scale our measured segregation to reflect segregation above and beyond that which would occur by chance if workers are distributed randomly across establishments, using Monte Carlo simulations to generate measures of randomly occurring segregation.[6]

Simple calculations of workplace segregation are important in their own right, aside from the questions we consider concerning the sources of workplace segregation. Most research on segregation by race and ethnicity focuses on residential segregation (e.g., Massey and Denton, 1987; and Cutler et al., 1999). But the boundaries used in studying residential segregation are to some extent arbitrary or even endogenously related to characteristics for which one wants to measure segregation; for example, Census tract boundaries are often generated in order to ensure that the tracts are "as homogeneous as possible with respect to population characteristics, economic status, and living conditions."[7] In contrast, workplaces–specifically establishments–are units of observation that are generated by economic forces and in which people clearly do interact in a variety of ways, including work, social activity, labor market networks, etc.[8] Thus, while it is more difficult to study workplace segregation because of data constraints, measuring workplace segregation may be more useful than measuring residential segregation,

---

[5] We focus on Hispanic ethnicity.

[6] This distinction between comparing measured segregation to a no-segregation ideal versus segregation that is generated by randomness is discussed in other work (see, e.g., Cortese, et al., 1976; Winship, 1977; Boisso, et al., 1994; and Carrington and Troske, 1997).

[7] U.S. Census Bureau, www.census.gov/geo/www/GARM/Ch10GARM.pdf (viewed May 10, 2004).

[8] For a discussion of the importance of the workplace as a venue for social interaction between groups see Estlund (2003).

as traditionally defined, for describing the interactions that arise in society between different groups in the population.[9] Of course similar arguments to those about workplaces could be made about other settings, such as schools, religious institutions, etc. (e.g., James and Taeuber, 1985).

Second, our main inquiry probes the relationship between skill segregation on the one hand and racial and ethnic segregation on the other. Numerous models suggest that employers find it useful to group workers of similar skills together. For example, Kremer and Maskin (1996) develop a model in which employers have incentives to segregate workers by skill when workers of different skill levels are not perfect substitutes and different tasks within firms are differentially sensitive to skill.[10] Saint-Paul (2001) generates skill segregation across firms by assuming that there are productivity-related spillovers among workers within an establishment.[11] Cabrales and Calvó-Armengol (2002) show that when workers' utility depends on interpersonal comparisons with nearby workers (such as those in the same firm), segregation by skill results.[12] And, of course, there are potential benefits to employers from grouping together workers who speak the same language.

---

[9] Moreover, industry code, the closest proxy in public-use data to an establishment identifier, is a very crude measure to use to examine segregation. We calculate that racial and ethnic segregation at the three-digit industry level in the DEED is on the order of less than half as large (and sometimes much less) as the establishment-level segregation we document below.

[10] For example, let the production function be $f(L_1, L_2) = L_1^c L_2^d$, with $d > c$. Assume that there are two types of workers: unskilled workers with labor input equal to one efficiency unit, and skilled workers with efficiency units of $q > 1$. Kremer and Maskin show that for low q, it is optimal for unskilled and skilled workers to work together, but above a certain threshold of q (that is, a certain amount of skill inequality), the equilibrium will reverse, and workers will be sorted across firms according to skill. Hirsch and Macpherson (1999) do not posit a formal model of sorting by skill, but assume that employers tend to hire workers of similar skills, and use this assumption–coupled with an assumption that blacks are on average less skilled than whites in terms of both observed and unobserved (to the researcher) skills–to suggest that the wage penalty associated with working in establishments with a large minority share in the workforce in part reflects lower unobserved skills of workers in such establishments.

[11] For example, positive spillovers may be reflected in each worker's productivity being the product of his productivity and an increasing function of the establishment's average skill level. Negative spillovers may arise because of fixed factors of production. All that is required for segregation in Saint-Paul's model is that over some range of average skill levels of an establishment's workforce there are increasing returns to skill.

[12] These authors also discuss evidence consistent with sorting by skill across employers, including Brown and Medoff (1989) and Davis, et al. (1991).

Because race and ethnicity are correlated with skill (for example, blacks have less education than whites and Hispanics have lower English proficiency), racial and ethnic segregation may not reflect discrimination, but may be generated by segregation along skill lines. We begin by calculating the extent of segregation in the workplace by education. We calculate education segregation measures focusing only on whites, assuming implicitly that segregation by education for whites is made by employers solely for reasons of economic efficiency.

We then measure the extent of segregation between blacks and whites, and calculate how much of this segregation can be explained by differences in educational attainment between blacks and whites. We contrast these results with the extent to which wage differences between blacks and whites in our sample can be explained by education.

We repeat the analysis for the extent of segregation between Hispanics and whites. In considering the impact of skill in generating workplace segregation by Hispanic ethnicity, we measure the extent to which segregation by English language ability can explain Hispanic-white workplace segregation, treating language ability as another important dimension of skill.[13] We also compare these results to those from wage regressions where we measure how much of the Hispanic-white wage gap is driven by English language ability.

Finally, language is associated not only with skill, but also with country of origin, immigrant status, and assimilation. Consequently, if discriminatory forces lead to the segregation of blacks or Hispanics from whites, they can also operate to segregate workers with poor English skills (immigrants, most likely) from other workers, in which case segregation by language would not reflect skill complementarities. We probe this question by exploring segregation among those whose English proficiency is poor, but whose native (and spoken) languages differ.

---

[13] We first documented segregation by language ability and explored its consequences for wages in Hellerstein and Neumark (2003). Because language may reflect things other than skill, there may be additional influences on hiring by language, including customer discrimination or the need for workers to speak the same language as customers, which, coupled with residential patterns, lead to this form of workplace segregation.

Our results point to workplace segregation by education and race, and more so by ethnicity and language (at least for Hispanics). We find, however, that education plays very little role in generating workplace segregation by race. In contrast, segregation by language ability can explain approximately one third of overall Hispanic-white segregation. Finally, the evidence from poor English speakers points to segregation of Hispanics from others, suggesting that the role of language segregation among Hispanics is driven by complementarity in language skills.

II. Data

The analysis in this paper is based on the DEED, which we have created at the Center for Economic Studies at the U.S. Bureau of the Census. The DEED is formed by matching workers to establishments. The workers are drawn from the 1990 Sample Edited Detail File (SEDF), which contains all individual responses to the 1990 Decennial Census of Population one-in-six Long Form. The establishments are drawn from the Standard Statistical Establishment List (SSEL), an administrative database containing information for all business establishments operating in the United States in 1990. Here we provide a brief overview of the construction of the DEED; more details regarding the matching of the data are provided in Hellerstein and Neumark (2003).

Households receiving the 1990 Decennial Census Long Form were asked to report the name and address of the employer in the previous week for each employed member of the household. The file containing this employer name and address information is referred to as the "Write-In" file, and had previously been used only for internal Census Bureau purposes. The Write-In file contains the information written on the questionnaires by Long-Form respondents, but not actually captured in the SEDF. The SSEL is an annually-updated list of all business establishments with one or more employees operating in the United States. The Census Bureau uses the SSEL as a sampling frame for its Economic Censuses and Surveys, and continuously updates the information it contains. The SSEL contains the name and address of each establishment, geographic codes based on its location, its four-digit SIC code, and an identifier that allows the establishment to be linked to other establishments that are part of the

same enterprise, and to other Census Bureau establishment- or firm-level data sets that contain more detailed employer characteristics. We can therefore use employer names and addresses for each worker in the Write-In file to match the Write-In file to the SSEL. Because the name and address information on the Write-In file is also available for virtually all employers in the SSEL, nearly all of the establishments in the SSEL that are classified as "active" by the Census Bureau are available for matching. Finally, because both the Write-In file and the SEDF contain identical sets of unique individual identifiers, we can use these identifiers to link the Write-In file to the SEDF. Thus, this procedure yields a very large data set with workers matched to their establishments, along with all of the information on workers from the SEDF.

Matching workers and establishments is a difficult task, because we would not expect employers' names and addresses to be recorded identically on the two files. To match workers and establishments based on the Write-In file, we use MatchWare–a specialized record linkage program. MatchWare is comprised of two parts: a name and address standardization mechanism (AutoStan); and a matching system (AutoMatch). This software has been used previously to link various Census Bureau data sets (Foster, et al., 1998). Our method to link records using MatchWare involves two basic steps. The first step is to use AutoStan to standardize employer names and addresses across the Write-In file and the SSEL. Standardization of addresses in the establishment and worker files helps to eliminate differences in how data are reported. For example, a worker may indicate that she works on "125 North Main Street," while her employer reports "125 No. Main Str." The standardization software considers a wide variety of different ways that common address and business terms can be written, and converts each to a single standard form.

Once the software standardizes the business names and addresses, each item is parsed into components. To see how this works, consider the case just mentioned above. The software will first standardize both the worker- and employer-provided addresses to something like "125 N Main St." Then AutoStan will dissect the standardized addresses and create new variables from the pieces. For example,

the standardization software produces separate variables for the House Number (125), directional indicator (N), street name (Main), and street type (St). The value of parsing the addresses into multiple pieces is that we can match on various combinations of these components.

We supplemented the AutoStan software by creating an acronym for each company name, and added this variable to the list of matching components. We noticed that workers often included only the initials of the company for which they work (e.g., "ABC Corp"), while the business is more likely to include the official corporate name (e.g., "Albert, Bob, and Charlie Corporation").

The second step of the matching process is to select and implement the matching specifications. The AutoMatch software uses a probabilistic matching algorithm that accounts for missing information, misspellings, and even inaccurate information. This software also permits users to control which matching variables to use, how heavily to weight each matching variable, and how similar two addresses must be in order to constitute a match. AutoMatch is designed to compare match criteria in a succession of "passes" through the data. Each pass is comprised of "Block" and "Match" statements. The Block statements list the variables that must match exactly in that pass in order for a record pair to be linked. In each pass, a worker record from the Write-In file is a candidate for linkage only if the Block variables agree completely with the set of designated Block variables on analogous establishment records in the SSEL. The Match statements contain a set of additional variables from each record to be compared. These variables need not agree completely for records to be linked, but are assigned weights based on their value and reliability.

For example, we might assign "employer name" and "city name" as Block variables, and assign "street name" and "house number" as Match variables. In this case, AutoMatch compares a worker record only to those establishment records with the same employer name and city name. All employer records meeting these criteria are then weighted by whether and how closely they agree with the worker record on the street name and house number Match specifications. The algorithm applies greater weights to items that appear infrequently. So, for example, if there are several establishments on Main St. in a

given town, but only one or two on Mississippi St., then the weight for "street name" for someone who works on Mississippi St. will be greater than the "street name" weight for a comparable Main St. worker. The employer record with the highest weight will be linked to the worker record conditional on the weight being above some chosen minimum. Worker records that cannot be matched to employer records based on the Block and Match criteria are considered residuals and we attempt to match these records on subsequent passes using different criteria.

It is clear that different Block and Match specifications may produce different sets of matches. Matching criteria should be broad enough to cover as many potential matches as possible, but narrow enough to ensure that only high probability matches are linked. Because the AutoMatch algorithm is not exact there is always a range of quality of matches, and we were therefore cautious in accepting linked record pairs. Our general strategy was to impose the most stringent criteria in the earliest passes, and to loosen the criteria in subsequent passes, while always maintaining criteria that erred on the side of avoiding false matches. We did substantial experimentation with different matching algorithms, and visually inspected thousands of matches as a guide to help determine cutoff weights. In total, we ran 16 passes, obtaining most of our matches in the earliest passes. Finally, we engaged in a number of procedures to fine-tune the matching process, involving hand-checking of thousands of matches and subsequent revision of the matching procedures.

The final result is an extremely large data set of workers matched to their establishment of employment. The DEED consists of information on 3.3 million workers matched to nearly one million establishments, which account for 27 percent of workers in the SEDF and 19 percent of establishments in the SSEL. Descriptive statistics for the matched workers and establishments, along with comparisons to the full SEDF and SSEL, respectively, are provided in Tables 1 and 2.[14]

[14] For both data sets, we have excluded individuals as follows: with missing wages; who did not work in the year prior to the survey year (1989); who worked in public administration or were self-employed; who were not classified in a state of residence; or who were employed in an industry that was considered "out-of-scope" in the SSEL. ("Out-of-scope" industries do not fall under the purview of Census Bureau surveys. They include many agricultural industries, urban transit, the U.S. Postal Service, private households, schools and universities, labor unions, religious and membership organizations, and

As reported in Table 1, the means of the demographic variables in the DEED are quite close to the means in the SEDF. For example, female workers comprise 46 percent of the SEDF and 47 percent of the DEED. The distribution of workers across races and ethnicities is also relatively similar; in the SEDF, white, Hispanic, and black workers account for 82, 7, and 8 percent of the total, respectively. The comparable figures for the DEED are 86, 5, and 5 percent. Similarly, there is a close parallel between the distributions of workers across education categories in the two data sets. The distributions of workers across industries paint a slightly different picture, as approximately 25 percent of all workers in the SEDF are employed in the manufacturing sector, a figure that is somewhat greater in the DEED (33 percent). Retail workers comprise 20 percent of all workers in the SEDF, and 17 percent in the DEED.

In addition to comparing worker-based means, it is useful to examine the similarities across establishments in the SSEL and the DEED. Table 2 shows descriptive statistics for establishments in each data set. There are 5,237,592 establishments in the SSEL; of these, 972,436 (19 percent) also appear in the DEED. Because only workers who are sent Decennial Census Long Forms are eligible for matching to their employers, it is far more likely that at least one worker in large establishments will be sent a Long Form, and consequently more likely that such establishments are included in the DEED. One can see evidence of the bias toward larger employers by comparing the means across data sets for total employment. (No doubt this also influences the distribution of workers and establishments across industries.) On average, establishments in the SSEL have 18 employees, while the average in the DEED is 53 workers. The distributions of establishments across industries in the DEED relative to the SSEL are similar to those in the worker sample. For example, manufacturing establishments are somewhat over-represented in the DEED, constituting 13 percent of establishments, relative to 6 percent in the SSEL. Overall, analyses reported in Hellerstein and Neumark (2003) indicate that the DEED sample is far more representative than previous detailed matched data sets for the United States.

---

government/public administration. The Census Bureau does not validate the quality of SSEL data for businesses in out-of-scope industries.)

III. Methods

We focus our analysis on one measure of segregation which is based on measures of the percentages of workers in an individual's establishment, or workplace, in different demographic groups. For example, if Hispanics and whites are perfectly segregated, then Hispanics work with 100 percent Hispanics and zero percent whites, and conversely whites work with 100 percent whites and zero percent Hispanics. For a dichotomous classification of workers (e.g., whites and Hispanics), we define two segregation variables: the average percentage of Hispanic workers with which Hispanic workers work, denoted H_H; and the average percentage of Hispanic workers with which white workers work, denoted W_H.[15] The difference between these,

$$CW = H_H - W_H$$

is our measure of observed "co-worker segregation," and measures the extent to which Hispanics are more likely than are whites to work with other Hispanics.[16,17]

There are, of course, many other possible segregation measures, such as the traditional Duncan index (Duncan and Duncan, 1955). We prefer the co-worker segregation measure (CW) to these other measures for two reasons. First, it is sensitive to the proportions of each group in the workforce. For example, if the distribution of Hispanics across establishments remains constant, but the number of Hispanics doubles, CW will rise. In contrast, the Duncan index is invariant to such a proportional change in the representation of any group. Second, the co-worker measure we use is invariant to establishment

---

[15] In the sociological literature, the percentage of Hispanics in a Hispanic's firm is often called the "isolation index" and the percentage of whites in a Hispanic's firm is the "exposure index." Note that these measures are calculated to include the own worker's ethnicity.

[16] To be precise, we exclude an individual's own ethnicity in calculating H_H and W_H. This means that we exclude from our sample workers and establishments where we have only one worker matched to the establishment. In unreported results, we have computed all of the figures in this paper using a segregation measure that includes the individual, and where we can therefore include establishments with only one matched workers. The qualitative results are identical to those reported here.

[17] We could equivalently define the percentages of white workers with which Hispanic or white workers work, H_W and W_W, which would simply be 100 minus these percentages, in which case CW would simply be the opposite sign.

size. The same is not true of the Duncan index or other similar indexes.[18] For these reasons, we prefer the co-worker segregation measure and use it exclusively below.[19]

We first report observed segregation, which is simply the sample estimate of the segregation measure. We denote these measures by appending an 'O' superscript to the segregation measures–i.e., CW^O. One important point that is often overlooked in research on segregation, however, is that some segregation occurs even with random assignment, and we are presumably most interested in the segregation that occurs systematically–i.e., that which is greater than would be expected to result from randomness. In the case of an infinite sample of workers, random allocation across establishments would imply that the co-worker segregation measure as we have defined it would be equal to zero, since, for example, H_H and W_H would be equal to the population Hispanic share. There are two reasons why in our analysis the segregation measure with randomly assigned workers is not necessarily expected to be zero. The first is that some of our segregation measures are calculated conditional on geography and skill. So, for example, when we condition on geography, we calculate the extent of segregation that would be expected if workers were randomly allocated across establishments within a geographic area. If Hispanics and whites are not evenly distributed across geographic borders, random allocation of workers within geography will yield the result that Hispanics are more likely to have Hispanic co-workers than are white workers. Second, although the baseline sample size in our data is large, the actual samples that we use to calculate segregation below are not always large when we condition on geography and skill, or at least not large enough to approximate well this asymptotic result. For that reason, in order to determine how much segregation would occur randomly, we conduct a Monte Carlo simulation of the

[18] For example, consider the case where workers are randomly allocated across establishments by race. Racial segregation calculated with our co-worker segregation measure will be invariant to establishment size. The Duncan index is not, nor are other indexes like it, including the "index of isolation" or the "index of exposure" which typically include the worker's own characteristics in the calculation.

[19] We have computed (but do not report) all the segregation measures in this paper using the Duncan index as well and the results are qualitatively unchanged.

extent of segregation with random allocation of workers (the algorithm is described below). We label this measure "random segregation" and denote it as CW^R.[20]

Following Carrington and Troske (1997), to measure segregation beyond that which would occur randomly, we compute the difference between observed segregation and the mean level of random segregation, and scale the difference by the maximum segregation that can occur. We refer to this as "effective segregation." In terms of the co-worker segregation measure, for $CW^O > CW^R$, the segregation measure is:

$$[\{CW^O - CW^R\}/\{100 - CW^R\}]\times 100 .$$

The denominator, $100 - CW^R$, is the maximum by which observed segregation can exceed random segregation, and so the scaling converts the difference $CW^O - CW^R$ into the share of this maximum possible segregation that is actually observed.[21]

For the Monte Carlo simulations that generate measures of random segregation, we first define the geographic unit–which for most of the analysis is metropolitan areas. We then compute from our data the numbers of workers in each category for which we are doing the simulation–for example, blacks and whites–as well as the number of establishments and the size distribution of establishments (in terms of sampled workers). Within a metropolitan area, we then randomly assign workers to establishments, ensuring that we generate the same size distribution of establishments within a metropolitan area as we have in the sample. One "round" of random assignment is considered a single simulation. We do the simulation 100 times, and compute the random segregation measures as the means over these 100 simulations. It turns out that the random segregation measures are very precise; in all cases the standard deviations were trivially small, and observed segregation was well the outside the 99-percent confidence interval for the random measures.

---

[20] The issue of randomness is more serious when the segregation measure is itself a function of the characteristics of the worker or unit for which it is being calculated, such as for the Duncan index. See Carrington and Troske (1997) for a detailed discussion of random segregation in this situation.

[21] In principle, CW^O can be lower than CW^R. This never happens in our application.

As noted above, most of our analysis focuses on metropolitan areas. We use U.S. Census Bureau measures of metropolitan areas, because these are defined to some extent based on areas within which substantial commuting to work occurs.[22] We first look at segregation within Consolidated Metropolitan Statistical Areas (CMSAs) and Metropolitan Statistical Areas (MSAs) for which there is no CMSA, and then at MSAs and Primary Metropolitan Statistical Areas (PMSAs), which are parts of CMSAs. The restrictions to workers in CMSAs/MSAs or MSAs/PMSAs reduce the sample by about one-third. In addition, the sample is further restricted to the subset of groups considered (e.g., Hispanics and whites only), and metropolitan areas are dropped from the relevant segregation calculation if there is no worker in the DEED from the subset considered (e.g., Hispanics) in that urban area. When we perform the simulations for disaggregated regions and calculate segregation measures, we condition on geographic area of residence (and work) so that we have region-specific segregation measures. We then calculate "conditional" national segregation measures by weighting over the whole sample (that is, summary measures for the extent of workplace segregation, where we condition on the metropolitan area where the worker lives and works).[23] For descriptive purposes we also present some "unconditional" nationwide segregation measures where we do not first condition on metropolitan area, and where in the simulations we randomly assign workers to establishments anywhere in the country.[24] For comparability, when we construct these unconditional segregation measures we use only the workers included in the CMSA/MSA or MSA/PMSA samples.

---

[22] See U.S. Census Bureau, http://www.census.gov/geo/lv4help/cengeoglos.html (viewed July 3, 2003). This is not to say that residential segregation at a level below that of MSAs and PMSAs may not influence workplace segregation. However, an analysis of this question requires somewhat different methods. For example, in conducting the simulations it is not obvious how one should limit the set of establishments within a metropolitan area in which a worker could be employed.

[23] In all cases, a worker must live and work in the same geographical region, or the worker is dropped from the sample.

[24] Not surprisingly, all the simulations we report where we randomly assign workers to establishments anywhere in the country lead to random segregation measures that are zero or virtually indistinguishable from zero.

The segregation measures for each region in the sample have to be aggregated to construct conditional national segregation measures. Because our co-worker observed and random segregation measures are worker based, our national conditional co-worker measure is simply the overall sample mean across all workers, which effectively weights metropolitan areas by the number of workers in them. As a result, for the observed segregation measures the conditional and unconditional measures yield identical results; only the simulations differ. With these conditional national measures of observed and random segregation, we then construct the effective segregation measures, which capture the level of effective segregation to which the average worker is subjected, conditional on the distribution of workers across metropolitan areas.

Finally, in addition to constructing estimates of effective segregation in the workplace along various dimensions, we are interested in comparisons of measures of effective segregation across different samples. Given also that we are sometimes comparing estimates across samples that have some overlap,[25] we assess statistical significance of measures of effective segregation or differences between them using bootstrap methods. These methods are computationally intensive because within each iteration of the bootstrap we have to do the set of simulations needed to construct measures of random segregation, and our samples are often very large. The methods for doing this for a number of cases, and the results are detailed in Appendix A. Briefly, the evidence indicates that our estimates are quite precise, and that the differences between the effective segregation indexes discussed in detail in the next section are generally strongly statistically significant.

IV. Results

Workplace Segregation by Education

The analysis begins with measures of workplace segregation by education for whites. We focus first on whites so as not to confound our measures of segregation by education with segregation that is

[25] For example, we compare effective segregation between Hispanics who speak English poorly and Hispanics who speak English well, to effective segregation between Hispanics who speak English poorly and non-Hispanics who speak English poorly.

driven by other factors, such as race, that are correlated with education. Because it is easiest to characterize segregation with a binary measure of education, we define workers as low education if they have a high school degree or less, and high education if they have at least some college.[26] Table 3 reports results for education segregation, using the full sample of establishments with two or more matched workers. To provide a sense of overall segregation, column (1) provides the various segregation measures at the unconditional national level, looking at all urban areas (CMSAs and MSAs) as a whole. Column (2) presents the conditional national segregation indexes that are constructed by weighting up to the national level each individual CMSA/MSA segregation measure. Column (3) repeats the unconditional national measures, but for the subsample of workers who live and work in the often-smaller MSA/PMSA urban areas, and column (4) provides the conditional MSA/PMSA segregation measures.

In column (1), looking first at the observed co-worker segregation measures, we see extensive segregation. In particular, low education workers on average work in establishments in which 52.3 percent of matched co-workers are also low education. In contrast, high education workers work in establishments with co-workers who are only 32.4 percent low education on average. Below these figures we present the calculations from the simulations. Given that we randomize workers in this sample across the whole United States in conducting this simulation, it is not surprising that the results of the simulation imply that, on average, both low and high educated workers work with co-workers who are 40.4 percent low education–the sample average. That is, for this particular exercise, the random co-worker segregation measure is zero, so that the effective co-worker simulation measure, 19.9, is simply the observed co-worker simulation measure (CW^O). One useful way to interpret this number is that almost 20 percent of the maximum amount of segregation that could arise due to non-random factors is actually observed in the data. While it is not clear to what one should compare these results, these results suggest to us that there is substantial segregation by education.

[26] We further disaggregate workers by education below when we consider how much of segregation by race is attributable to segregation by education.

Column (2) looks at segregation within urban areas defined as CMSAs/MSAs. As noted earlier, observed co-worker segregation is the same within and across urban areas. The random segregation measure is 3.9 (no longer zero for reasons explained above, because worker are reallocated for the simulation only within urban areas); the pattern of random segregation has low education workers working, on average, with co-workers who are 42.8 percent low educated, while for high education workers the corresponding figure is 38.9 percent. As a result, the effective segregation measure in column (2) is 16.6. That is, about 17 percent of the maximum amount of segregation that could arise due to non-random factors is observed in the data.

Columns (3) and (4) of Table 3 repeat this analysis for the urban areas defined by MSAs/PMSAs. The estimates in column (3) are very close to those in column (1). As we would expect, random segregation within the smaller PMSA/MSA definition as reflected in column (4) is slightly higher than in column (2), but overall this has no substantive impact on the estimates or on the qualitative conclusions. In the remainder of the paper, we report only results using urban areas defined by MSAs/PMSAs. Throughout, results for MSAs/CMSAs were not qualitatively different.

We view the results of Table 3 as one benchmark for the extent of segregation that arises from firms efficiently sorting workers across workplaces. With these results in mind, we turn to the extent of segregation by race.

Workplace Segregation by Race

Table 4 reports results for overall black-white segregation, holding the PMSA/MSA constant, and compares it to segregation by education, both for whites and for blacks. In columns (1) and (2) of Table 4, we report the extent of segregation by race (black versus white) in the whole United States for workers in the MSA/PMSA sample; the first column presents results when we simulate random segregation by allocating workers randomly anywhere in the United States, while column (2) presents results where we randomly allocate workers to an establishment within their MSA/PMSA. On average, black workers work with co-workers who are 23.7 percent black, while white workers work with co-

workers who are 5.8 percent black. Based on the sample average of blacks in the population, random allocation across the United States would imply that blacks and whites should each work with co-workers who are 7.1 percent black, so that the overall level of effective segregation as reported in column (1) is 17.8. Because there is some racial segregation across urban areas, when we simulate random segregation within urban areas in column (2), there is some segregation that arises randomly. In particular, random assignment would lead blacks to work in establishments with co-workers who are on average 11.2 percent black, versus an average percent black of 6.8 percent for whites. Based on the comparison between observed and random segregation, the effective segregation measure is 14.1, meaning that just over 14 percent of the maximum amount of racial segregation that could arise due to non-random factors is actually observed in the data.

Column (3) repeats the results from column (4) of Table 3, showing the extent of segregation by education for whites when conditioning on MSA/PMSA, and column (4) reports the extent of segregation by education, this time for the sample of only black workers, when conditioning on MSA/PMSA.. There are more low education blacks in the sample than whites, but the observed and random segregation (CW^O and CW^R) across the two columns are very similar, so that the effective segregation measure for education segregation for blacks is 17.1, similar to the 16.5 for whites. This is at least suggestive that the factors driving skill segregation (as defined here by education) are the same for whites as for blacks, as would be expected if skill segregation is arising due to profit-maximizing behavior.

Comparing column (2) to columns (3) and (4) shows that the extent of segregation by race is very similar to that of segregation by education. Although the overall fraction of black workers is much lower than the fraction of low educated workers in the sample, the observed and random co-worker segregation measures are remarkably similar when comparing racial segregation to education segregation. As a result, the overall extent of racial segregation in the United States is very much of the same order of magnitude as education segregation for whites (14.1 versus 16.5).

Workplace Segregation by Race, Conditional on Education

Next, we measure the extent to which racial segregation in the workplace can be explained by education differences between blacks and whites. We do this by constructing new "conditional" random segregation measures, where we simulate segregation holding the distribution of education fixed across all workplaces. So, for example, if an establishment in the DEED is observed to have three workers with a high school degree, three workers with a high school degree will be randomly allocated to that establishment. We again compute the average (across the simulations) simulated fraction of co-workers who are black for blacks, denoting this B_B^C, and the average (across the simulations) simulated fraction of co-workers who are black for whites, denoting this W_B^C. The difference between these two is denoted CW^C, and we define the extent of "effective conditional segregation" to be $(CW^O-CW^C)/((100-CW^R)*100)$, where CW^R is the measure of random segregation obtained when not conditioning on education. A conditional effective segregation measure of zero would imply that all of the effective segregation between blacks and whites can be attributed to education segregation that is coupled with differences in the education distribution between blacks and whites. Conversely, a conditional effective segregation measure equal to that of the (unconditional) effective segregation measure would imply that none of the effective segregation between blacks and whites can be attributed to education segregation across workplaces. We first do this calculation with the same two-way classification of education used in Table 4, and then expand to four and then six educational categories.

Column (1) of Table 5 reports the results for the two-way education classification. Observed segregation between blacks and whites is unaffected by this conditioning, of course, and so the top part of column (1) of Table 5, which reports the observed segregation between blacks and whites, repeats the results from Table 4, column (2). We report the conditional random segregation measures starting in the middle of the rows of Table 5. On average, random allocation of workers, conditional on randomization within the two education categories and within PMSA results in black workers working, on average, with co-workers who are 11.4 percent black, and white workers working, on average, with co-workers who are

6.8 percent black. These numbers are very close to the (unconditional) simulated numbers reported in Table 4, column (2). As a result, the conditional effective segregation measure is 13.9, very close to the unconditional segregation measure of 14.1. In other words, segregation by the binary education distinction (which we measure to be extensive) can explain only a tiny fraction (1.4 percent) of overall black-white segregation.

We repeat this analysis in column (2) of Table 5, this time conditioning on four education groupings when randomizing workers to workplaces: less than high school; high school degree; some college or associates degree; and bachelors degree or above. The results of the conditional random segregation are very similar to that obtained with two education groupings, so that our conditional effective segregation measure falls only to 13.6. In column (3) we define education by six distinct categories: less than high school; some high school; high school degree; some college; associates degree; and college degree or more. The conditional effective segregation measure remains at 13.6. In each case, then, differences in education between blacks and whites and attendant workplace segregation by education do virtually nothing to explain black-white workplace segregation.

One possible explanation for this finding is that education differences between blacks and whites are too small in this sample to have meaningful consequences for workplace segregation by race. We explore this possibility in Table 6. First, in columns (1) and (2), we report means by race for the six education groups used in the calculation of Table 5, column (3). There clearly are measurable differences in educational attainment between blacks and whites in this sample. Most noticeably, 8.7 percent of whites have a high school diploma whereas 15.1 percent of blacks do. In contrast, 25.4 percent of whites have at least a college degree whereas only 13.5 percent of blacks do.

Of course, it is still the case that these education differentials between blacks and whites may have no impact on the labor market in general, and therefore have no impact on workplace segregation. In columns (3) and (4) we therefore explore the extent to which education differences between blacks and whites impact another important dimension of the labor market, namely wage differentials. We

report results from wage regressions, regressing the log of the wage for the entire sample used in calculating our black-white segregation measures on a dummy for black, and on controls for education.[27] In column (3) we report results, using the entire sample of workers used in calculating our black-white segregation measures, where we simply regress the log of the wage on a black dummy variable. The estimated coefficient on the black dummy is −0.204, reflecting the commonly found black-white raw wage differential. In column (4) we report results from running the same regression, but adding controls for the six education categories (omitting less than high school). The estimated returns to education are all large and monotonically increasing, and the estimated coefficient on the black dummy falls to −0.126. So in this sample, as in most studies of black-white wage gaps, education differences between blacks and whites can explain a large fraction (here, 38 percent) of the black-white wage gap.[28]

The message from this exercise is that racial differences in educational attainment do have an important impact on racial wage differences here, as in virtually all other studies. In contrast, however, we find that these same education differences have virtually no impact on workplace segregation between blacks and whites. Given that education essentially plays no role in generating what we consider to be the rather substantial amount of racial segregation in the workplace, it is difficult to imagine that unobservable skill differences between blacks and whites could explain a sizeable fraction of workplace segregation by race. The mechanism(s) behind workplace segregation by race therefore appear not to be skill related. Alternative mechanisms such as labor market discrimination, spatial mismatch, or labor market networks are all possibilities worthy of future exploration.

---

[27] For a small fraction of the sample (less than one percent), hourly wages are less than one dollar per hour, rendering the log wage negative. Excluding workers whose measured hourly wage is less than two does not markedly affect either the wage regression results or the measurement of black-white segregation.

[28] We also ran the regressions (not reported) with a quadratic in age and a dummy variable for sex. The impact of education on the black-white wage dummy is robust to including these controls.

Workplace Segregation by Ethnicity

We now turn to an examination of the extent and causes of workplace segregation by Hispanic ethnicity. The baseline estimates for the extent of Hispanic-white segregation are reported in Table 7, and the basic conclusion is that there is extensive workplace segregation by Hispanic ethnicity. The first specific thing to note is that the segregation figures for the unconditional national indexes indicate somewhat more segregation by ethnicity than their counterparts for race as reported in Table 4. For example, in column (1) of Table 7, the average percentage of Hispanics with whom Hispanics work is 39.4 percent, versus a comparable figure of 23.7 percent for blacks. The effective segregation measures are similarly different: 34.9 for Hispanic-white segregation and 17.8 for black-white segregation.

The results are not as starkly different when we condition on metropolitan areas. This occurs because for Hispanics, randomly-generated segregation is quite far from zero, conditional on metropolitan areas. In column (2) of Table 7, for example, the randomly allocated share Hispanic for Hispanic workers is 24.4 percent, compared with a parallel share Hispanic for white workers of 5.6 percent. This difference mainly arises because Hispanics are not as evenly dispersed across metropolitan areas, some of which have few Hispanics. The net result is that, conditional on metropolitan area, the effective co-worker segregation measures are only somewhat higher for Hispanics (19.8) than for blacks (14.1).

In columns (3) and (4) of Table 7, we explore the extent of workplace segregation by English language proficiency for whites and Hispanics separately. As for education, employers may find it efficient to segregate workers by English language proficiency. Indeed, it is possible that the motives for segregation by language are even stronger than for education since workers who cannot communicate with each other impose clearly impose costs on employers relative to the alternative. We divide language proficiency into two categories. The first, "poor English," consists of workers who report speaking English poorly or not at all. The second, "good English," consists of workers who report speaking English well or very well.

In column (3) we report the extent of workplace segregation by language for whites. Less than one half of one percent of the white sample are in the poor English category, yet a worker in this category works, on average, with co-workers for whom 6.9 percent speak English poorly. In contrast, for white workers in the good English category, only 0.4 percent of their co-workers speak English poorly. Random co-worker segregation for this sample, while not zero, is small (0.5). As a result, effective segregation for whites by language proficiency is 6.0. While the scale of this is smaller than for the other effective segregation measures computed thus far, we think it is notable given the very small percentage of poor English speakers among the whites to begin with.

The results on language segregation for Hispanics clearly show that there is extensive workplace segregation by language proficiency. These results are in column (4). Hispanics who speak English poorly or not at all are likely to have Hispanic co-workers among whom, on average, 48.1 percent also speak English poorly or not at all. In stark contrast, Hispanics in the "good English" category are likely to have Hispanic co-workers of whom, on average, only 15.4 percent are in the "poor English" group. The random segregation measures indicate that some segregation arises randomly, conditional on geographic area. Under random allocation Hispanics in the "poor English" category would have 26.8 percent of Hispanic co-workers speaking English poorly or not at all, while workers in the "good English" category would have 21.7 percent of co-workers speaking English poorly or not at all. All together, this implies that the effective segregation measure for language segregation for Hispanics is 29.1, an order of magnitude larger than any other (within MSA/PMSA) segregation measure thus far.

In Table 8, we explore the extent to which the very pronounced language segregation for Hispanics may be driving Hispanic-white workplace segregation, since Hispanics have so much lower English language proficiency, on average, than whites. In the top panel of column (1) we repeat the figures for observed Hispanic-white segregation from Table 7, column (2); as reported earlier, the difference between the co-worker segregation for Hispanics and whites is 34.9. We then report conditional random segregation for Hispanics and whites, conditional on the two language groupings

used in the previous table (in addition to MSA/PMSA). With random allocation within the two language groups, Hispanics on average work with co-workers who are 26.8 percent Hispanic, whereas whites work with co-workers who are 5.5 percent Hispanic. That is, the simulated difference between the co-worker segregation measures is 21.3. Together these numbers lead to an effective segregation measure of 16.7. When we repeat this exercise in column (2), this time randomizing workers within the four language groups about which workers self-report English language proficiency (not at all, poorly, well, very well), the effective segregation measure is 13.5. This figure can be interpreted as saying that of the Hispanic-white unconditional effective segregation measure of 19.8, nearly a third (32 percent = (19.8-13.5)/19.8) can be explained by language segregation.

Finally, in Table 9 we report for comparison purposes the extent to which language proficiency can explain the Hispanic-white wage gap. First, in columns (1) and (2) we report sample means for whites and Hispanics by language proficiency category. As noted above, the fraction of whites who do not speak English at all or who speak it poorly is tiny: 0.0038. In fact, 98.9 percent of whites in the sample report speaking English very well. In contrast, for Hispanics there is much more variation in English language ability. Five percent report not speaking English at all and 14 percent report speaking English poorly, so that 19 percent of the sample of Hispanics falls into the low English speaking group in the binary language categorization we used to compute language segregation. For the remainder of Hispanics in the sample, 18.4 percent report speaking English well and 62.6 report speaking English very well.

Columns (3) and (4) report wage regression results. In column (3), the coefficient on a Hispanic dummy in a log wage regression with no other controls is -0.277. Column (4) reports results from a regression that adds dummies for language proficiency. The coefficients on the language dummies themselves show that the return to language proficiency is monotonic and increasing, and causes the coefficient on the Hispanic dummy to fall to -0.204, a 26 percent drop.[29] Similar results have been found

[29] The result is larger (a 42 percent drop) if we control for a quadratic in age and a sex dummy in the regression, but is very robust to trimming the sample to exclude workers who earn hourly wages

in other work on the Hispanic-white wage gap (and in our previous work with the DEED, in Hellerstein and Neumark, 2003).

In sum, skill differences between Hispanics and whites, at least as defined by language proficiency, explain approximately the same order of magnitude of Hispanic-white workplace segregation as of the Hispanic-white wage gap. This contrasts with the finding that skill differences between blacks and whites, as defined by education differentials, explain virtually none of racial segregation in the workplace.

Understanding Workplace Segregation by Language Proficiency

For Hispanic workers we have documented that substantial workplace segregation is generated by skill differences, at least as defined by language proficiency. One interpretation of this evidence is that employers have good reasons to pursue such segregation, and because language proficiency is correlated with ethnicity, segregation by language arising for non-discriminatory reasons generates segregation by ethnicity. Another possibility, though, is that language is associated with other dimensions along which employers discriminate–such as national origin or socioeconomic factors–and on the basis of which employers crowd workers into a subset of jobs (typically jobs that pay less). It can be difficult to distinguish between these competing hypotheses.[30] In the case of language skills, however, we believe some progress can be made on this question.

In particular, to test whether there are legitimate economic reasons for segregation by language skill, as opposed to simple segregation of those with poor English into a subset of jobs, we can consider employment patterns for workers who speak poor English but who also speak different languages. If

---

computed to be below $2 per hour.

[30] This is potentially true in many contexts, even though it is often ignored. For example, Bertrand and Mullainathan (2004) provide evidence from an audit study that employers are less likely to interview job candidates with "black-sounding" names. This may be because of race discrimination per se, or because of discrimination against workers whose names suggest a certain cultural and socioeconomic upbringing (or the intersection of the two), but the paper has been interpreted as providing evidence of discrimination on the basis of race. (See also Fryer and Levitt, 2003.)

Hispanic poor English speakers (who generally speak Spanish) are not segregated from non-Hispanic poor English speakers (who speak a language other than Spanish), then this would suggest that those with low skills are clustered in the same workplaces for reasons other than efficiency gains from grouping workers who speak the same language; such segregation would be more consistent with simple segregation of "less desirable" workers into a subset of jobs. In contrast, if Hispanic poor English speakers are segregated from those who have poor English skills but speak languages other than Spanish, then segregation by language skills may be arising for reasons of complementarity between workers who speak the same language (or a related economic incentive to segregate workplaces by language).[31]

The results of this analysis are reported in Table 10. Column (1) repeats the calculations from Table 7 for segregation between Hispanic workers with poor English skills and Hispanic workers with good English skills. In contrast, column (2) reports calculations for segregation between Hispanics with poor English skills and non-Hispanics with poor English skills. These figures indicate much more extensive segregation–considerably more than in column (1)–49.5 versus 29.1. Note that in column (2) random segregation is far from zero, much of this resulting from sorting across MSA/PMSAs (and some resulting from randomness due to the small sample). Thus, this evidence suggests that much of the segregation of Hispanics with poor English skills arises because of factors other than the general crowding of low-skilled workers into the same set of low-paying workplaces.

However, some caution is in order in interpreting these estimates. Unlike the case with education differences overall or English language skills among Hispanics, residential segregation between Hispanics and other groups with poor English might be quite strong. If residential segregation by language drives workplace segregation along similar lines, we might expect that random segregation would be higher and effective segregation lower for units of analysis smaller than the MSA/PMSA level. Therefore, to further address whether the segregation by language among poor English speakers is due to

[31] However, the latter finding would not necessarily be decisive, because such segregation by language may be a function of residential segregation and/or hiring networks where workers who speak the same language have access to the same subset of employers. Network relationships can themselves be efficiency enhancing if they make it easier for workers to find jobs or for employers to find workers.

complementarities in the workplace, we explore differences in language segregation by establishment size. After all, in larger establishments there may be considerably more scope for segregating workers within establishments, so that across-establishment segregation is not as critical in achieving language complementarities.[32] Table 11 reports similar calculations to those in Table 10, but for establishments with different minimum total employment cutoffs–roughly the 25th, 50th, and 75th percentiles of the size distribution. As the table shows, segregation of Hispanics by language ability is robust to establishment size,[33] but segregation of Hispanics from non-Hispanics when both groups have poor English skills falls as the minimum establishment size is raised. This pattern suggests that language complementarities do in fact contribute to workplace segregation by language among those who speak poor English.

V. Conclusions

We use a unique data set of employees matched to establishments to study workplace segregation in the United States. We document that there is rather extensive segregation by education for white workers, consistent with models where employers find it efficient to segregate workers by skill. Similarly, among Hispanics we document extensive segregation by language, which is perhaps even stronger evidence that skill complementarities in the workplace. We also document that there is segregation by race in the workplace of the same order of magnitude as education segregation, and segregation by Hispanic ethnicity that is slightly larger.

After documenting these different dimensions of segregation, our analysis focuses on whether racial and ethnic workplace segregation reflects race or ethnicity per se, likely stemming from discrimination, or instead is attributable to skills that differ across race and ethnic groups and along

---

[32] As an anecdotal example, an article in the *New York Times* describes a Texas factory that nearly completely segregates its Hispanic and Vietnamese workers into two different departments in the factory (with the Hispanics working in the lower-paying department). This article also points to the role of language complementarities between workers and supervisors, as one of the company's defenses of this practice is that the supervisor of the higher-paying department speaks Vietnamese but not Spanish (Greenhouse, 2003).

[33] Similarly, segregation by education, for both blacks and whites, was robust across different employment cutoffs.

which employers might find it useful to segregate workers. For racial segregation, we find that virtually none of it is attributable to skill differences, at least as these are manifested in education differences between blacks and whites. In contrast, we show that approximately one third of ethnic segregation in the workplace is attributable to language proficiency.

Finally, in order to further probe the role of skill in generating ethnic (and language) segregation, we ask whether segregation by skill likely arises due to the consignment of less-skilled workers to the same subset of workplaces, perhaps because of discrimination against workers on the basis of numerous characteristics associated with low skills, or whether other factors such as skill-based complementarities lead certain types of workers to work together. Providing evidence inconsistent with the first hypothesis, we find that Hispanics with poor English skills are considerably more segregated from workers with poor English skills who speak other languages than they are from Hispanics with good English skills. It therefore appears that the process by which Hispanic and white workers are sorted into workplaces is not simply one whereby low-skilled workers are relegated to the same set of (low-paying) workplaces.

In addition to finding that there is extensive segregation by skill in the workplace, our results document the reality of racial and ethnic segregation in U.S. workplaces. For blacks, the fact that education differences between blacks and whites explain virtually none of racial workplace segregation means that further research must be conducted to uncover the sources of racial segregation in the workplace, and that this research necessarily must examine explanations that are not skill-based: discrimination, residential segregation, and labor market networks are the most obvious possibilities. While language proficiency can explain a large fraction of ethnic segregation in the workplace, the remainder may be due to other dimensions of skill, or may be generated again by non skill-based explanations.

Appendix A: Statistical Significance

From the point of view of drawing statistical inferences, we need to be able to assess the statistical significance of our effective segregation measures and of differences between them. Given the precision of the random segregation measures as discussed in Section III, the effective segregation measures are also likely relatively precise. To assess this more formally, we explore bootstrapped distributions for the effective segregation measures.

To carry out this procedure, at each iteration of the bootstrap we draw a sample with replacement of the original size of the sample. We sample establishments, not workers. This ensures that we maintain the size distribution of establishments, and in particular that we maintain the restriction that all establishments have at least two matched workers. The bootstrap sample at each iteration is then the workers in these establishments. We then calculate the observed segregation measures, and compute random segregation the same way as described earlier, with 100 Monte Carlo simulations, so that there are 100 iterations within each iteration of the bootstrap. Finally, we collect the information on the empirical distribution of the effective segregation measures.

Given that this procedure is very intensive computationally, we did not carry it out for all of the estimates presented in the paper. Instead, because the estimates in Table 10 are based on the smallest samples, we carried out a detailed analysis for these estimates. We computed the 100 bootstraps for each of the samples in columns (1) and (2) of Table 8. Looking at the results for the sample of Hispanic and non-Hispanic workers who speak English poorly, in column (2), the effective segregation measure was estimated reasonably precisely, with a standard deviation of 1.17. The standard deviation of the effective segregation measure for the larger sample of Hispanics with good English relative to those with poor English (column (1)) were smaller, at 0.55.

Finally, in order to assess whether the differences in estimated effective segregation between samples are statistically significant, we pair each of the 100 bootstraps across the two samples, calculate the difference in the segregation measures across the samples for each bootstrap, and calculate the

standard deviation of the difference in the segregation measures across columns. So while the effective co-worker segregation measures in Table 8 differ across the two samples by 20.6 (29.1 versus 49.7), the standard deviation of the bootstrapped difference in these measures across columns is 1.31, indicating that the observed difference is highly statistically significant. Thus, especially given the far large sample sizes in the other tables in the paper, it seems clear that the differences in effective segregation measures that we obtain in the paper are generally highly statistically significant.

References

Altonji, Joseph G., and Rebecca M. Blank. 1999. "Race and Gender in the Labor Market." In Handbook of Labor Economics, Vol. 3, eds. Ashenfelter and Card (Amsterdam: Elsevier), pp. 3143-259.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2004. "Trends in U.S. Wage Inequality: Re-Assessing the Revisionists." Unpublished manuscript, MIT.

Bayard, Kimberly, Judith Hellerstein, David Neumark, and Kenneth Troske. 1999. "Why Are Racial and Ethnic Wage Gaps Larger for Men than for Women? Exploring the Role of Segregation Using the New Worker-Establishment Characteristics Database." In The Creation and Analysis of Employer-Employee Matched Data, eds. Haltiwanger, Lane, Spletzer, Theeuwes, and Troske (Amsterdam: Elsevier Science B.V.), pp. 175-203.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review*, Vol. 94, No. 4, September, pp. 991-1013.

Becker, Gary S. 1971. The Economics of Discrimination, Second Edition (Chicago: University of Chicago Press).

Boisso, Dale, Kathy Hayes, Joseph Hirschberg, and Jacques Silber. 1994. "Occupational Segregation in the Multidimensional Case." *Journal of Econometrics*, Vol. 61, No. 1, March, pp. 161-71.

Brown, Charles, and James Medoff. 1989. "The Employer Size Wage Effect." *Journal of Political Economy*, Vol. 97, No. 5, October, pp. 1027-59.

Cabrales, Antonio, and Antoni Calvó-Armengol. 2002. "Social Preferences and Skill Segregation." Unpublished paper, Universitat Pompeu Fabra.

Cain, Glen. 1986. "The Economic Analysis of Labor Market Discrimination: A Survey." In Handbook of Labor Economics, Vol. 1, eds. Ashenfelter and Layard (Amsterdam: North-Holland), pp. 693-785.

Card, David, and John E. DiNardo. 2002. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." Journal of Labor Economics, Vol. 20, No. 4, October, pp. 733-83.

Carrington, William J., and Kenneth R. Troske. 1997. "On Measuring Segregation in Samples with Small Units." *Journal of Business & Economic Statistics*, Vol. 15, No. 4, October, pp. 402-9.

Cortese, Charles, F., R. Frank Falk, and Jack K. Cohen. 1976. "Further Considerations on the Methodological Analysis of Segregation Indices." *American Sociological Review*, Vol. 51, No. 4, August, pp. 630-7.

Cross, Harry, Genevieve Kenney, Jane Mell, and Wendy Zimmerman. 1990. Employer Hiring Practices: Differential Treatment of Hispanic and Anglo Job Seekers (Washington, DC: Urban Institute Press).

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 1999. "The Rise and Decline of the American Ghetto." *Journal of Political Economy*, Vol 107, No. 3, June, pp. 455-506.

Darity, William A., Jr., and Patrick L. Mason. 1998. "Evidence on Discrimination in Employment: Codes of Color, Codes of Gender." *Journal of Economic Perspectives*, Vol. 12, No. 2, Spring, pp. 63-92.

Davis, Steve J., John Haltiwanger, Lawrence F. Katz, and Robert Topel. 1991. "Wage Dispersion Between and Within U.S. Manufacturing Plants, 1963-1986." *Brookings Papers on Economic Activity: Microeconomics*, Vol. 1, pp. 115-200.

Donohue, John J., and James Heckman. 1991. "Continuous Versus Episodic Change: The Impact of Civil Rights Policy on the Economic Status of Blacks." *Journal of Economic Literature*, Vol. 29, No. 4, December, pp. 1603-43.

Duncan, Otis D., and Beverly Duncan. 1955. "A Methodological Analysis of Segregation Indices." *American Sociological Review*, Vol. 20, No. 2, April, pp. 210-7.

Estlund, Cynthia. 2003. Working Together: How Workplace Bonds Strengthen a Diverse Democracy (New York: Oxford University Press).

Foster, Lucia, John Haltiwanger, and C.J. Krizan. 1998. "Aggregate Productivity Growth: Lessons from Microeconomic Evidence." NBER Working Paper No. 6803.

Fryer, Roland G., and Steven D. Levitt. 2003. "The Causes and Consequences of Distinctively Black Names." NBER Working Paper No. 9938.

Greenhouse, Steven. 2003. "At a Factory in Houston, Hispanics Fight to Work in Coveted Department." *New York Times*, February 9, p. 14.

Heckman, James J. 1998. "Detecting Discrimination." *Journal of Economic Perspectives*, Vol. 12, No. 2, Spring, pp. 101-16.

Hellerstein, Judith, and David Neumark. 2003. "Ethnicity, Language, and Workplace Segregation: Evidence from a New Matched Employer-Employee Data Set." *Annales d'Economie et de Statistique*, Vol. 71-72, July-December, pp. 19-78.

Higgs, Robert. 1977. "Firm-Specific Evidence on Racial Wage Differentials and Workforce Segregation." *American Economic Review*, Vol. 67, No. 2, March, pp. 236-45.

Hirsch, Barry T., and David A. Macpherson. 2003. "Wages, Sorting on Skill, and the Racial Composition of Jobs." IZA Discussion Paper No. 741.

Ihlanfeldt, Keith, and David Sjoquist. 1990. "Job Accessibility and Racial Differences in Youth Employment Rates." *American Economic Review*, Vol. 80, No. 1, March, pp. 267-76.

James, Daniel R., and Karl E. Taeuber. 1985. "Measures of Segregation." In Sociological Methodology, ed. Brandon Tuma (San Francisco: Jossey-Bass), pp. 1-32.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy*, Vol. 101, No. 3, June, pp. 410-42.

Katz, Lawrence F., and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *Quarterly Journal of Economics*, Vol. 107, No. 1, February, pp. 35-78.

King, Mary C. 1992. "Occupational Segregation by Race and Sex, 1940-1988." *Monthly Labor Review*, April, pp. 30-7.

Kremer, Michael, and Eric Maskin. 1996. "Wage Inequality and Segregation by Skill." National Bureau of Economic Research Working Paper No. 5718.

Massey, Douglas, and Nancy Denton. 1987. "Trends in the Residential Segregation of Blacks, Hispanics, and Asians: 1970-1980." *American Sociological Review*, Vol. 52, No. 6, December, pp. 802-25.

Neal, Derek A., and William R. Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy*, Vol. 104, No. 5, October, pp. 869-95.

O'Neill, June. 1990. "The Role of Human Capital in Earnings Differences between Black and White Men." *Journal of Economic Perspectives*, Vol. 4, No. 4, Fall, pp. 25-45.

Rivera, Elaine. 2003. "Area Bosses Try to Bridge Language Gaps." *Washington Post*, May 6, p. B1.

Saint-Paul, Gilles. 2001. "On the Distribution of Income and Worker Assignment under Intrafirm Spillovers, with an Application to Ideas and Networks." *Journal of Political Economy*, Vol. 109, No. 1, February, pp. 1-37.

Turner, Margery Austin, Michael Fix, and Raymond J. Struyk. 1991. Opportunities Denied, Opportunities Diminished: Racial Discrimination in Hiring (Washington, DC: Urban Institute Press).

U.S. Census Bureau. "Census Geographic Glossary." http://www.census.gov/geo/lv4help/cengeoglos.html (viewed July 3, 2003).

U.S. Census Bureau, "Census Tracts and Block Numbering Areas." http://www.census.gov/geo/www/GARM/Ch10GARM.pdf (viewed May 10, 2004).

Watts, Martin J. 1995. "Trends in Occupational Segregation by Race and Gender in the U.S.A., 1983-92: A Multidimensional Approach." *Review of Radical Political Economics*, Vol. 27, No. 4, Fall, pp. 1-36.

Welch, Finis. 1990. "The Employment of Black Men." *Journal of Labor Economics*, Vol. 8, No. 2, April, pp. S26-S75.

Winship, Christopher. 1977. "A Revaluation of Indexes of Residential Segregation." *Social Forces*, Vol. 55, No. 4, June, pp. 1058-66.

Table 1: Means of Worker Characteristics

	SEDF (1)	Full DEED (2)
Age	37.08	37.51
	(12.78)	(12.23)
Female	0.46	0.47
Married	0.60	0.65
White	0.82	0.86
Hispanic	0.07	0.05
Black	0.08	0.05
Full-time	0.77	0.83
Number of kids (if female)	1.57	1.53
	(1.62)	(1.55)
High school diploma	0.34	0.33
Some college	0.30	0.32
B.A.	0.13	0.16
Advanced degree	0.05	0.05
Ln(hourly wage)	2.21	2.30
	(0.70)	(0.65)
Hourly wage	12.10	12.89
	(82.19)	(37.07)
Hours worked in 1989	39.51	40.42
	(11.44)	(10.37)
Weeks worked in 1989	46.67	48.21
	(11.05)	(9.35)
Earnings in 1989	22,576	25,581
	(26,760)	(29,475)
Industry:		
Mining	0.01	0.01
Construction	0.07	0.04
Manufacturing	0.25	0.34
Transportation	0.08	0.05
Wholesale	0.05	0.07
Retail	0.20	0.17
FIRE	0.08	0.08
Services	0.26	0.24
Observations	12,143,183	3,291,213

Standard deviations of continuous variables are reported in parentheses.

Table 2: Means for Establishments

	SSEL	Full DEED
Total employment	17.57	52.68
	(253.75)	(577.39)
Establishment size:		
1 - 25	0.88	0.65
26 - 50	0.06	0.15
51 - 100	0.03	0.10
101 +	0.03	0.10
Industry:		
Mining	0.00	0.01
Construction	0.09	0.07
Manufacturing	0.06	0.13
Transportation	0.04	0.05
Wholesale	0.08	0.11
Retail	0.25	0.24
FIRE	0.09	0.10
Services	0.28	0.26
In MSA	0.81	0.82
Census Region:		
North East	0.06	0.06
Mid Atlantic	0.16	0.15
East North Central	0.16	0.20
West North Central	0.07	0.08
South Atlantic	0.18	0.16
East South Central	0.05	0.05
West South Central	0.10	0.10
Mountain	0.06	0.05
Pacific	0.16	0.15
Payroll ($1000)	397	1,358
	(5,064)	(10,329)
Payroll/total employment	21.02	24.24
	(1,385.12)	(111.79)
Share of employees matched	–	0.17
Multi-unit establishment	0.23	0.42
Observations	5,237,592	972,436

Standard deviations of continuous variables are reported in parentheses. 55 establishments in the DEED sample do not have valid county data from the SSEL. For these 55, the workers reported place of work was used to determine MSA status.

Table 3: Segregation by Education for Whites

	Establishment education composition:			
	U.S., CMSA/MSA sample	Within CMSA/MSA	U.S., MSA/PMSA, sample	Within MSA/PMSA
	%Low ed	%Low ed	%Low ed	%Low ed
	(1)	(2)	(3)	(4)
Co-worker segregation				
Observed segregation				
Low education workers (L_L^O)	52.3	52.3	53	53
High Education workers (H_L^O)	32.4	32.4	33.1	33.1
Difference (CW^O)	19.9	19.9	19.9	19.9
Random segregation				
Low education workers (L_L^O)	40.4	42.8	41.3	43.7
High Education workers (H_L^O)	40.4	38.9	41.3	39.6
Difference (CW^R)	0	3.9	0	4.1
Effective segregation, $[\{CW^O - CW^R\}/\{100 - CW^R\}]\times 100$	**19.9**	**16.7**	**19.9**	**16.5**
Number of workers	**1,611,445**	**1,611,445**	**1,500,322**	**1,500,322**
Number of establishments	**287,786**	**287,786**	**273,084**	**273,084**

Low education is defined as high school degree or less. High education is defined as more than high school. Calculations are for establishments with two or more matched workers. For the CMSA/MSA (MSA/PMSA) sample of workers, the median number of workers matched to an establishment is 8 (8), and the median share of the workforce matched 7.9 (7.7) percent. For the sample of establishments, the median number of matched workers is 3 (3), and the median share of the workforce matched is 8.9 (8.7) percent. While the median numbers of workers matched are low, this arises because there are many small establishments in the data; the shares of the workforce matched range from 8.0 to 9.1 percent, relative to a hypothetical maximum of 16.7 percent, given that only 1/6 of workers receive the Census long form. All medians are reported as "fuzzy medians" to comply with confidentiality restrictions; but they are extremely close to actual medians.

Table 4: Black-White Segregation and Education Segregation by Race

	Establishment racial composition:			Establishment education composition:	
	Black-white segregation in U.S. (MSA/PMSA sample)	Black-white segregation within MSA/PMSA		Education segregation for whites	Education segregation for blacks
	%Black	%Black		%Low ed	%Low ed
	(1)	(2)		(3)	(4)
Co-worker segregation					
Observed segregation					
Black workers (B_B^O)	23.7	23.7	Low education workers (L_L^O)	53.0	58.9
White workers (W_B^O)	5.8	5.8	High Education workers (H_L^O)	33.1	41.0
Difference (CW^O)	17.8	17.8	Difference (CW^O)	19.9	17.9
Random segregation					
Black workers (B_B^S)	7.1	11.2	Low education workers (L_L^O)	43.7	51.6
White workers (W_B^S)	7.1	6.8	High education workers (H_L^O)	39.6	48.3
Difference (CW^R)	0	4.4	Difference (CW^R)	4.1	3.3
Effective segregation, $[\{CW^O - CW^R\}/\{100 - CW^R\}]\times 100$	**17.8**	**14.1**		**16.5**	**15.1**
Number of workers	1,618,876	1,618,876		1,500,322	83,401
Number of establishments	285,988	285,988		273,084	19,062

In columns (3) and (4), low education is defined as high school or less; high education is more than high school.

Table 5: Black-White Segregation Conditional on Education

	Black-white segregation conditional on 2 education groups	Black-white segregation conditional on 4 education groups	Black-white segregation conditional on 6 education groups
	%Black	%Black	%Black
	(1)	(2)	(3)
Co-worker segregation			
Observed segregation			
Black workers (B_B^O)	23.7	23.7	23.7
White workers (W_B^O)	5.8	5.8	5.8
Difference (CW^O)	17.8	17.8	17.8
Conditional random segregation			
Black workers (B_B^C)	11.4	11.6	11.6
White workers (W_B^C)	6.8	6.8	6.8
Difference (CW^C)	4.6	4.8	4.8
Effective conditional segregation, $[\{CW^O - CW^C\}/\{100 - CW^R\}]\times 100$	**13.9**	**13.6**	**13.6**
Number of workers	1,618,876	1,618,876	1,618,876
Number of establishments	285,988	285,988	285,988

In column (1), the two education groups are: high school or less; more than high school. In column (2), the four education groups are: less than high school; high school degree; some college or associates degree; bachelors degree or higher. The groups in column (3) are: less than high school; some high school; high school degree; some college; associates degree; college degree or more.

Table 6: The Distribution of Education by Race and the Impact of Education on Black-White Wage Gaps

	Sample means		Regression results	
	Whites	Blacks		
	(1)	(2)	(3)	(4)
Black	0	1	-0.204 (0.002)	-0.126 (0.002)
Less than high school	0.016	0.029		–
Some high school	0.087	0.151		-0.126 (0.004)
High school degree	0.309	0.319		0.090 (0.004)
Some college	0.242	0.282		0.194 (0.004)
Associates degree	0.091	0.083		0.309 (0.004)
College degree or above	0.254	0.135		0.638 (0.004)
Number of observations	1,503,640	115,236	1,618,876	1,618,876

The dependent variable in the regressions reported in columns (3) and (4) is the log of the hourly wage. There is a constant in the regressions; therefore, less than high school is omitted from the regression in column (4).

Table 7: Hispanic-White Segregation and Language Segregation by Ethnicity

	Establishment ethnic composition:			Establishment language composition:	
	Hispanic-white segregation in U.S. (MSA/PMSA sample)	Hispanic-white segregation within MSA/PMSA		Language segregation for whites	Language segregation for Hispanics
	%Hispanic	%Hispanic		%Poor English	%Poor English
	(1)	(2)		(3)	(4)
Co-worker segregation					
Observed segregation					
Hispanic workers (H_H^O)	39.4	39.4	Poor English workers (P_P^O)	6.9	48.1
White workers (W_H^O)	4.5	4.5	Good English workers (G_P^O)	0.4	15.4
Difference (CW^O)	34.9	34.9	Difference (CW^O)	6.5	32.7
Random segregation					
Hispanic workers (H_H^O)	6.9	24.4	Poor English workers (P_P^O)	0.9	26.8
White workers (W_H^O)	6.9	5.6	Good English workers (G_P^O)	0.4	21.7
Difference (CW^R)	0	18.8	Difference (CW^R)	0.5	5.1
Effective segregation, $[\{CW^O - CW^R\}/\{100 - CW^R\}]\times 100$	**34.9**	**19.8**		**6.0**	**29.1**
Number of workers	1,625,953	1,625,953		1,491,434	81,595
Number of establishments	293,989	293,989		271,101	21,933

Results in columns (3) and (4) are derived within MSA/PMSA; poor English is defined as speaking English poorly or not at all; good English is speaking English well or very well.

Table 8: Hispanic-White Segregation Conditional on Language

	Hispanic-white segregation conditional on 2 language groups	Hispanic-white segregation conditional on 4 language groups
	%Hispanic	%Hispanic
	(1)	(2)
Co-worker segregation		
Observed segregation		
Hispanic workers (H_H^O)	39.4	39.4
White workers (W_H^O)	4.5	4.5
Difference (CW^O)	34.9	34.9
Conditional random segregation		
Hispanic workers (H_H^O)	26.8	29.2
White workers (W_H^O)	5.5	5.3
Difference (CW^C)	21.3	23.9
Effective conditional segregation, $[\{CW^O - CW^C\}/\{100 - CW^R\}]\times 100$	16.7	13.5
Number of workers	1,625,953	1,625,953
Number of establishments	293,989	293,989

In column (1), the two language groups are: speak English poorly or not at all; speak English well or very well. In column (2), the four language groups are: speak English not at all; speak English poorly; speak English well; speak English very well.

Table 9: The Distribution of Language by Ethnicity and the Impact of Language on Hispanic-White Wage Gaps

	Sample means		Regression results	
	Whites	Hispanics		
	(1)	(2)	(3)	(4)
Hispanic	0	1	-0.277 (0.002)	-0.204 (0.002)
Speak English "not at all"	0.0002	0.05		–
Speak English poorly	0.0036	0.14		0.210 (0.009)
Speak English well	0.0072	0.184		0.396 (0.009)
Speak English very well	0.989	0.626		0.471 (0.009)
Number of observations	1,513,277	112,676	1,625,953	1,625,953

The dependent variable in the regressions reported in columns (3) and (4) is the log of the hourly wage. There is a constant in the regressions; the category speak English not at all is omitted from the regression in column (4).

Table 10: Language Segregation, Within MSA/PMSA

	Establishment ethnic and skill composition:		
Hispanic workers, poor English- Hispanic workers, good English		Hispanic workers, poor English- non-Hispanic workers, poor English	
	%Hispanic, poor English		%Hispanic, poor English
	(1)		(2)
Co-worker segregation			
Observed segregation			
Hispanic workers, poor English	48.1	Hispanic workers, poor English	90.0
Hispanic workers, good English	15.4	Non-Hispanic workers, poor English	26.0
Difference	32.7		64.0
Random segregation			
Hispanic workers, poor English	26.8	Hispanic workers, poor English	80.2
Hispanic workers, good English	21.7	Non-Hispanic workers, poor English	51.5
Difference	5.1		28.7
Effective segregation, $\{CW^O - CW^R\}/\{100 - CW^R\}]\times 100$	**29.1**		**49.5**
Number of workers	81,595		19,926
Number of establishments	21,933		6,393

Calculations are for establishments with two or more matched workers.

Table 11: Language Segregation, Within MSA/PMSA, Sensitivity to Establishment Size

	Establishment ethnic and skill composition:						
Hispanic workers, poor English- Hispanic workers, good English				Hispanic workers, poor English- non-Hispanic workers, poor English			
	%Hispanic, poor English				%Hispanic, poor English		
	Employment > 10	Employment > 60	Employment > 170		Employment > 10	Employment > 60	Employment > 170
	(1)	(1')	(1")		(2)	(2')	(2")
Co-worker segregation							
Observed segregation							
Hispanic workers, poor English	47.9	45.7	42.1	Hispanic workers, poor English	89.8	88.3	85.6
Hispanic workers, good English	15.4	14.7	12.7	Non-Hispanic workers, poor English	27.0	29.4	32.0
Difference	32.5	31.0	29.4		62.8	58.9	53.6
Random segregation							
Hispanic workers, poor English	26.9	25.2	21.6	Hispanic workers, poor English	80.5	79.9	77.7
Hispanic workers, good English	21.7	20.3	17.1	Non-Hispanic workers, poor English	51.5	50.7	49.3
Difference	5.2	4.9	4.5		29.0	29.2	28.4
Effective segregation, $\{CW^O - CW^R\}/\{100 - CW^R\}]\times 100$	**28.8**	**27.4**	**26.1**		**47.6**	**42.0**	**35.2**
Number of workers	78,602	61,347	40,297		19,241	14,770	8,878
Number of establishments	20,566	13,734	7,056		6,077	4,223	2,187

Calculations are for establishments with two or more matched workers. The employment cutoffs chosen are approximately the 25th, 50th, and 75th percentiles of the establishment size distribution.