Wages and the university educated: a paradox resolved

Analysis of a new data base to study occupations and employment reveals a surplus of university graduates and a movement of many of these individuals, especially those with lower functional literacy, into high-school-level jobs; only university graduates with literacy skills commensurate with their education have received rising wages

Frederic L. Pryor
and
David Schaffer

Frederic L. Pryor is a professor of economics at Swarthmore College, Swarthmore, PA; David Schaffer is an assistant professor of economics at Haverford College, Haverford, PA.

If the university educated are flooding the job market, so that many must take jobs previously held by those with just a high school diploma, then why are the wages of these university-educated workers rising? This question has been the focus of an intermittent debate on the pages of the *Monthly Labor Review* between Daniel E. Hecker, on the one hand, and John Tyler, Richard J. Murnane, and Frank Levy, on the other.[1] Unfortunately, the participants in the discussion have not asked *which* university graduates have taken the high school jobs, and in their empirical investigations, they have used only a rough and subjective criterion for defining a high school job.

Utilizing much more detailed data on occupations, and taking into account the functional literacy of the workers, which is a critical variable, this article shows that it is primarily those university graduates lacking university-level literacy skills who are taking the high school jobs. Further, it is chiefly the university educated in jobs requiring university-level skills who are obtaining the major wage increases, not those in jobs in which the average level of functional literacy is lower.

Our argument is straightforward. We first summarize briefly the state of the debate about university-educated workers taking high school jobs. Next, we define "high school jobs" and "functional literacy." Then we present data on functional literacy, employment levels, and wages for workers with different levels of education in different types of occupations. These data sets provide the crucial evidence required to resolve the paradox of why an apparent surplus of university graduates is associated with rising real wages of this group. For reasons to be discussed, we focus primarily on prime-age workers, defined as those 25 through 49 years old.

The state of the debate

Hecker opened the debate in his two articles using highly aggregated occupation data to argue that an increasing number of university graduates were taking high school jobs.[2] He also brought into the discussion data from the Recent University Graduates Survey, conducted by the National Center for Education Statistics, indicating that almost 40 percent of the graduates awarded bachelor-of-arts degrees in 1984 and 1986 thought a university degree was not needed to obtain the job they held a year after graduation. Because, as some have pointed out, it takes many graduates several years to find employment suitable to their talents, the relevance of this evidence is not entirely clear. Hecker argued that the data he presented indicated a surplus of university graduates in relation to the number of available jobs, a position consistent with the Department of Labor's oft-stated projection that the U.S. job outlook for the university educated as a group is not as rosy as is commonly believed.[3]

Hecker then faced squarely the following difficult and crucial question: if there is a surplus of university-educated workers, why are the wages of the university educated rising? In answering this question, he proposes and then refutes two possible explanations.

First, many university graduates may have learned little or gained little productivity from their university education and do not qualify for jobs requiring university-level skills. As a result, they finally take jobs in occupations in which most workers have fewer years of education than they. Hecker uses aggregate trends in Scholastic Aptitude Test and Graduate Record Examination scores to argue that there is no evidence of declining average educational levels of university graduates that would encourage a rising number of them to take high school jobs. Moreover, he claims, it is not clear "why employers would place so many of these admittedly less-qualified [university] graduates in jobs that do not require [university]-level skills if they had vacant [university]-level jobs."[4] By way of contrast, we argue in what follows that this explanation is indeed the key to the paradox and that Hecker's rebuttals are not germane to the crucial points.

Second, Hecker's classification of occupations, on which he bases his empirical analysis, may bias his results. Because he uses highly aggregated occupation data, this is a legitimate problem for his discussion. For our highly disaggregated data, however, it is much less of a problem, except insofar as high school jobs are being upgraded so that they can be filled only by university graduates. We deal with this issue later.

Hecker's explanation for the increasing wage gap between high school and university graduates revolves around a restructuring of the economy and the decline of high-wage jobs for blue-collar workers. In this regard, he cites approvingly a conclusion by Lawrence Mishel and Ruy A. Teixeira that the relative return to education increased in the 1980s primarily because of declines in the real wages of the less educated, not because of increasing real wages for the more educated.[5] Later on, we provide contrary evidence on this question.

Tyler, Murnane, and Levy bring additional data, particularly regarding wages, into the argument. They make two major points.

First, recent increases in the proportion of university-educated workers in high school jobs have occurred primarily among older workers, rather than younger ones. The researchers offer evidence that, although male university graduates between 45 and 54 have increasingly accepted high school jobs, this has not been the case for women in the same age cohort or for men and women between 25 and 34. Unfortunately, the three authors use the same highly aggregated data as does Hecker to make this argument. Our data suggest a different picture.

Second, university graduates holding high school jobs receive more than high school wages. A key point to keep in mind is that in many, if not most, occupations, the work can be carried out by people with different levels of formal education. That is, employers face a choice of hiring high school graduates at a given wage or university graduates at a higher wage; presumably, the latter have higher marginal productivities. Further, the wage gap in the same occupations between university graduates and those with just a high school education has risen over the years. So the university educated in high school jobs are not suffering the sad fate of downward occupational mobility suggested by Hecker. John H. Bishop also stresses this phenomenon in his argument that more years of education usually have a payoff, no matter what the job.[6] We, too, offer evidence of same phenomenon.

On a number of issues, the participants of the debate are actually talking past each other. In our empirical discussion, we ask the simple question, Who are the university graduates ending up in high school jobs? We also investigate the relevant wage issues on a more detailed level.

Some crucial definitions

To attack the problems systematically, it is necessary to develop some objective criteria for defining a high school job. It also is important to explore the meaning of "functional literacy."

High school jobs. The participants in the debate about the university-educated taking high school jobs have all used highly aggregated data. Hecker, for instance, defines all work within retail sales, a large proportion of work in administrative support, and various large categories of operator and laborer jobs as "nonuniversity jobs," that is, jobs which do not require a university degree. This, of course, is subjective and does not allow him to take into account either the specific occupations within these broad categories or the actual differences in education of those holding such jobs. Tyler, Murnane, and Levy adopt the same rough statistical procedure, probably to allow easier comparisons with Hecker's results.

We utilize a more detailed and less subjective procedure, justified later, to classify occupations. First, we use a complete set of combined data from the March 1971 and 1972 Current Population Surveys (CPS's) to calculate the average level of education among prime-age (between 25 and 49 years, inclusive) workers in each of 500 detailed occupations.[7] We then employ this information to classify each detailed occupation into 1 of 4 occupational tiers according to level of education. Tier 1 occupations are defined as those in which the average prime-age worker had 10.5 or fewer years of schooling in 1971 and 1972. Tier 4 occupations, at the other extreme, are defined as those in which the average amount of schooling was greater than 14.5 years in 1971 and 1972. Tier 2 occupations consist of those in which the average amount of schooling was more than 10.5, but less than 12, years in the same reference period; and tier 3 occupations are those in which the average amount of schooling was more than 12, but less than 14.5, years. Under this classification system, jobs in tier 1 and tier 2 occupations are definitely non-university-level jobs, in that they are unlikely to require most of the skills taught primarily at the university level. Jobs in tier 3 occupations do not require a university degree and in many cases can be considered high school jobs.

In deciding upon this definition of types of jobs, we chose to exclude workers below age 25 from our calculations for two reasons: first, many young workers may still be attending school, so their current job may have little relation to their eventual job options after completing school; and second, many young workers who have just completed their education end up initially in jobs significantly inferior to those in which they will settle a few years later. We also exclude workers over age 49 so that our results are not influenced by workers in their fifties experiencing either disabilities associated with aging or job losses because of downsizing. In both cases, these individuals cannot obtain work commensurate with their education. Finally, we select 1971 and 1972 as the base years for our definition so that we can compare the U.S. labor market in 4 evenly spaced years that span almost a quarter of a century: 1971, 1979, 1987, and 1995. These years have overall unemployment rates that are roughly similar, so that business cycle conditions do not greatly influence our results.

Table 1 presents data showing the changing share of university-educated workers in high school jobs, under three alternative definitions of a high school job. Panel A lists the percentages of those of various ages with a university degree with jobs in occupations whose average level of education is less than or equal to 12 years. Panel B presents data for the same age groups, but with an education level cutoff of 13 or fewer years, and in Panel C, the cutoff is 14.5 or fewer years. These different calculations allow us to determine the degree to which the definition of a high school job influences the conclusions.

We agree with Hecker, as well as with Tyler, Murnane, and Levy, that the greatest degree of downward occupational mobility occurred in the 1970s. Our conclusions differ, however, in many details from those of the other two studies.

The calculations for all workers parallel similar estimates by Hecker. With his methodology, however, the results show a large increase in the number of university graduates in high school jobs between 1970 and 1980, with very little increase after 1980. We obtain similar results when we use the narrowest definition of a high school job (Panel A). That is, the percentage of university graduates in high school jobs increased most between 1971 and 1979. The increase was less between 1979 and 1987, and there was actually a slight decline between 1987 and 1995, although the 1995 level was still higher than in 1979. If we use either of the two broader definitions, however, the deceleration in the growth of the percentage of university graduates in high school jobs still appears (although the decline does not), but is much weaker, at least until 1987.

It is also striking that, in our analysis, the results for prime-age workers alone are very similar to the results for the entire employed population. This seems to contradict the assertion by Tyler, Murnane, and Levy that there was a differential effect based on sex and age and that the phenomenon was confined to men from 45 to 54, many of whom had difficulty obtaining work commensurate with their education after they were laid off. To get at this issue more directly, we calculated the percentages for the same sex and age subcategories as those reported by Tyler and his collaborators. They found that, between 1979 and 1989, the percentage of university graduates in high school jobs actually *decreased* for all four subcategories, except men aged 45 to 54. By contrast, we find that the percentage *increased* for all four subcategories from 1979 to 1987, and it seems unlikely that the difference in endpoints accounts for the difference in results.

Using the narrowest definition of a high school job (Panel A), we find that the percentage of university graduates in high school jobs improved for all

Table 1. Percentage of workers aged 25 to 54 years with a university education in high school jobs

Panel and year	Total	Prime-age workers (25–49)	Men, 25–34	Women, 25–34	Men, 45–54	Women, 45–54
A:						
1971	7.1	5.7	6.3	3.8	6.3	7.0
1979	9.9	9.7	13.1	8.2	7.9	6.8
1987	10.7	10.4	14.3	9.0	7.7	5.6
1995	10.1	9.6	12.5	6.7	10.1	6.4
B:						
1971	23.7	22.1	21.5	16.0	27.2	24.8
1979	30.6	30.0	33.1	29.0	30.2	23.7
1987	34.5	34.0	37.5	33.2	32.5	26.8
1995	34.8	34.0	36.6	31.8	38.5	26.2
C:						
1971	40.8	40.3	41.2	33.7	44.6	35.8
1979	50.2	49.4	53.9	49.9	48.5	39.9
1987	54.7	54.9	58.1	59.6	51.5	45.0
1995	55.4	55.0	56.9	56.3	56.9	45.2

NOTE: Panel A—jobs in occupations in tiers 1 and 2; average education level, 12 or fewer years in 1971–72. Panel B—jobs in occupations in tiers 1 and 2 and part of tier 3; average education level, 13 or fewer years in 1971–72. Panel C—jobs in occupations in tiers 1, 2, and 3; average education level, 14.5 or fewer years in 1971–72.

SOURCE: Current Population Surveys for March 1971, 1979, 1987, and 1995.

groups but older men from 1987 to 1995. This is similar to the findings of Tyler and colleagues. However, if we use either of the broader definitions, a different scenario appears: the percentages improve for the same three groups, but only by a small amount, and they certainly do not fall below their 1979 values. Such results throw some doubt on the strong assertion of Tyler, Murnane, and Levy that only older men experienced this type of occupational downgrading.

In table 2, we disaggregate the data for prime-age workers to show the number of workers with various levels of education in the occupations classified by educational tier in 1971 and 1995. Reading across the rows under the heads "percentages by rows," we see the percentages of those in occupations of a particular educational tier who have various levels of education. In 1971, for example, 2.9 percent of the workers in jobs in tier 2 occupations (in which the average education was 10.6 to 12.0 years) had a university degree; by 1995, the figure rose to 9.0 percent. Reading down the columns under the heads "percentages by columns," we see the percentage of those with a given education who are pursuing occupations in the various educational tiers. For instance, in 1971, 4.5 percent of all those with a university degree were in tier 2 occupations; in 1995, the corresponding figure was 6.6 percent. Such data confirm the phenomenon discussed by Hecker in a dramatic fashion, although we focus only on prime-age workers, whereas he deals with the entire labor force.

Table 2. Years of education and occupational tiers[1] of prime-age (25 to 49 years) workers, 1971 and 1995

[In percent]

Percentages and tiers	Highest educational attainment				
	Total	High school dropout	High school diploma only	Some university courses	University degree
Percentages by rows, 1971					
Total	100.0	28.5	41.4	13.4	16.7
Tier:					
1	100.0	58.3	36.4	4.5	.8
2	100.0	34.9	51.8	10.5	2.9
3	100.0	11.0	49.4	23.0	16.7
4	100.0	1.4	9.4	11.6	77.6
Percentages by columns, 1971					
Total	100.0	100.0	100.0	100.0	100.0
Tier:					
1	28.5	54.1	23.2	8.9	1.2
2	41.4	32.0	32.7	20.5	4.5
3	13.4	13.3	41.2	59.4	34.6
4	15.7	.6	2.9	11.2	59.7
Percentages by rows, 1995					
Total	100.0	8.9	32.5	28.9	29.6
Tier:					
1	100.0	26.1	49.2	20.4	4.3
2	100.0	12.0	45.9	33.1	9.0
3	100.0	2.4	28.2	37.0	32.4
4	100.0	.5	5.7	13.4	80.3
Percentages by columns, 1995					
Total	100.0	100.0	100.0	100.0	100.0
Tier:					
1	19.9	58.2	30.1	14.1	2.9
2	21.9	29.5	30.9	25.0	6.6
3	41.6	11.5	36.1	53.2	45.5
4	16.6	1.0	2.9	7.7	45.0

[1] Tiers are defined by average education of those in the occupations in 1971. Tier 1: 10.5 or fewer years; tier 2: 10.6 to 12.0 years; tier 3: 12.1 to 14.5 years; tier 4: 14.6 or more years.

NOTE: Percentages by rows show the share of workers with different occupations engaged in occupations with different average degrees of education of practitioners in 1971. Percentages by columns show the percentages of workers with a given education who are engaged in occupations in which the average level of education of practitioners in 1971 varied.

SOURCE: Current Population Surveys, March 1971 and March 1995.

Functional literacy. Functional literacy is the ability to use skills in reading, interpreting documents, and carrying out quantitative calculations in real-life situations. It differs from the skills learned in school because it represents what people can remember and apply in daily living. Indeed, the number of years of schooling a person has explains only about one-third of variations in functional literacy tests.[8] Functional literacy also differs from native intelligence, because, unlike the latter, it can be learned. Of course, greater intelligence allows the acquisition of functional literacy to be accelerated, but other factors also play a role, including the quality of schooling and the attitudes of both the individuals themselves and their classmates (so-called neighborhood effects).[9]

Standard human-capital models use years of education as a proxy variable for what workers can accomplish on the job. Unfortunately, such a proxy does not indicate what a person has remembered or, indeed, what job skills that person really has. More specialized studies of the particular skills needed for particular occupations have often taken into account the performance of a sample of individuals on standardized tests. Unfortunately, on such tests, the sample is frequently unrepresentative of those occupying the jobs in question.

For the purposes of linking functional literacy to labor force status, occupation, and wages, the National Adult Literacy Survey seems most appropriate. This 1992 test was administered carefully to a sample covering the entire adult population and deals with literacy skills used by adults in real-life situations.[10] The questions are more open ended than standard multiple-choice questions, they cover a variety of contexts, and they emphasize carrying out tasks requiring brief written or oral responses. In sum, this competency-based approach focuses on what adults can do with written information.[11]

The test distinguishes three scales of functional literacy. Prose literacy comprises "the knowledge and skills needed to understand and use information from texts including [newspaper] editorials, news stories, poems, and fiction." One question, for example, requires the respondent to summarize the main argument of an op-ed article. Document literacy comprises "the knowledge and skills required to locate and use information contained in materials that include job applications, payroll forms, transportation schedules, maps, tables, and graphs." One question, for instance, asks the respondent to complete an employment application, and another requires the interpretation of a line graph. Finally, quantitative literacy comprises "the knowledge and skills required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed materials, such as to balance a checkbook, complete an order form, or calculate the amount of interest from a loan advertisement." One question asks the respondent to add up the cost of a particular meal, to calculate what change should be returned, and to determine the amount of a 10-percent tip. Although the prose, document, and quantitative scales are different, the scores of the individual respondents along these three scales are highly correlated, with all correlation coefficients at .84 or above. Preliminary regressions linking the scores to different demographic or causal variables yield roughly the same results for each scale. For this reason, we use the *average* of the scores along these three scales as a variable, rather than the scores of the individual scales.

The paradox unmasked

Our argument has two steps: first we examine the functional literacy data by education and occupational tier; then we look at the relevant wage data. In between, we pose the question of whether the literacy results are biased.

Functional literacy. Underlying most of the debate is the assumption that workers with a certain level of education are homogeneous, which is clearly false. Functional literacy is a cognitive skill that plays a key role in the process of matching workers with jobs, as shown in table 3. The table is similar to table 2, except that entries in the cells are the average functional literacy of full-time workers.

Data in the rows in table 3 show clearly that the functional literacy of workers with a given education increases as the occupational tier increases.[12] Thus, those with a university education working in occupations in which most had less than a university education had lower functional literacy than university-educated individuals working in university-level jobs. Such evidence points toward the unfortunate conclusion that the university educated in high school jobs pursued those jobs because they had lower literacy qualifications than other university-educated workers and, moreover, could not obtain jobs in occupations commensurate with their education.

Let us not, however, be too eager to accept such a conclusion, because a certain amount of reverse causation may be present. That is, sometimes functional literacy may be partly the result of on-the-job learning. If this is true, then, to a certain extent, chance factors such as one's initial job after the completion of formal education can play a role in the relation between occupation, education, and functional literacy. Also, functional literacy is in part a result of attitude and motivation, which, too, have an impact on the type of employment chosen.

Bias in the literacy results? If individual occupations are upgrading necessary skills, then the 1971–72 classification of occupations we use in our tables may introduce a bias. We address this issue with two comments.

First, although there is considerable evidence that the overall skill level of the labor force is rising,[13] part of the increase has come about as a result of changes in the number of those pursuing particular occupations. Studies based on compari-

Table 3. Weighted-average functional literacy scores and jobs of the prime working-age (25 to 49 years) population, 1992

Highest level of education	Average level of education in job, 1971–72: Total	Occupational tier[1]: 1	2	3	4
Total	294	262	282	309	335
High school dropout	236	231	246	259	—
High school diploma only	279	267	280	292	297
Some university courses	307	291	301	311	322
University degree	333	—	316	331	340

[1] Tier 1: 10.5 or fewer years; tier 2: 10.6 to 12.0 years; tier 3: 12.1 to 14.5 years; tier 4: 14.6 or more years.

NOTE: The table reports the average scores of the three scales of functional literacy for full-time, prime-age workers. These scores run from 0 to 500, and the original sample came from a 1992 nationwide survey. For this table, the total sample size is 7,520 people, and the standard deviation is about 55. The results for high school dropouts in tier 4 and those with a university degree in tier 1 are not reported because the sample sizes are too small. Entries in the table are calculated from raw data from the National Adult Literacy Study.

sons of skill ratings of particular occupations in various editions of the U.S. Department of Labor's *Dictionary of Occupational Titles* find little net change in the skill requirements, holding the occupational composition constant.[14] More recent studies using other evidence find an upgrading of skills in some parts of the labor force, such as production workers in manufacturing, but not in others, such as clerical workers in manufacturing.[15] The proportion of the latter is, of course, increasing.

Second, whatever bias there may be in the data does not work against our interpretation of table 3. Suppose, for instance, that a number of university- educated workers are fulfilling jobs in 1995 that really do require higher education, even though in 1971 the "real" educational requirements, as well as the average education of the workers, were much lower. Assume, for a moment, that these jobs were properly classified in 1995 and that the university graduates in them had the same functional literacy as those in tier 4 occupations at the time. Then, if the average functional literacy in each cell of the table were recalculated, this would mean that the functional literacy scores of the university educated in university-level jobs would be even higher in comparison to the scores in the other cells. For instance, the university graduate with a functional literacy of 340 who was in a job incorrectly specified as being in a tier 3 occupation would now be classified as having a tier 4 occupation, so that the average functional literacy of tier 3 occupations would now be lower.

Table 4. Hourly earnings and years of education of prime-age workers, 1970–94

Educational tier[1]	Highest educational attainment: Total	High school dropout	High school diploma only	Some university courses	University degree
1970 average hourly wages (1994 prices)					
Total	$13.33	$10.32	$12.33	$14.85	$20.27
1	10.42	9.72	11.22	12.13	—
2	11.94	10.70	12.28	13.64	15.08
3	14.49	11.63	12.74	15.39	20.55
4	19.78	—	16.24	16.96	20.87
1978 average hourly wages (1994 prices)					
Total	14.38	10.64	12.68	14.10	20.36
1	11.55	10.21	12.12	14.05	—
2	12.58	11.03	12.94	13.20	13.41
3	14.90	11.55	12.72	14.37	19.96
4	20.90	—	14.67	15.19	22.15
1986 average hourly wages (1994 prices)					
Total	14.22	9.49	11.86	13.97	20.13
1	10.66	9.07	11.02	12.37	—
2	11.93	9.93	11.93	12.63	13.78
3	15.32	10.29	12.34	14.67	20.24
4	20.42	—	13.41	15.97	21.69
1994 average hourly wages (1994 prices)					
Total	14.80	9.00	11.30	13.18	22.42
1	10.01	8.37	10.42	11.43	—
2	11.33	9.40	10.89	11.77	14.53
3	16.01	11.36	12.19	14.08	21.94
4	23.00	—	14.44	15.26	25.07

[1] Occupations ranked by average education of practitioners in 1971–72. Tier 1: 10.5 or fewer years; tier 2: 10.6 to 12.0 years; tier 3: 12.1 to 14.5 years; tier 4: 14.6 or more years.

NOTE: The wage data represent average annual hourly earnings (total labor income divided by number of hours worked), deflated by the personal consumption price index in the gross domestic product accounts. The appendix describes in more detail the methods used in calculating the entries in this table. The results for high school dropouts in tier 4 and those with a university degree in tier 1 are not reported because the sample sizes are too small.

Some relevant wage data. The wage data presented in table 4 lead to four strong conclusions: (1) The real wages of university-educated workers occupying jobs requiring such education (tier 4 occupations) have increased significantly in the last quarter of a century. (2) The real wages of the university educated in jobs in which the average level of education is 14.5 years or less have remained roughly constant. (3) The real wages of those without a university education have generally declined. (4) The ratio of wages between those with a university degree and those with just a high school degree has increased, by 34 percentage points. As noted by Tyler, Murnane, and Levy, these differentials also widened for various jobs. For instance, between the same 2 years, for tier 2, 3, and 4 jobs, the increases in wage differentials were 11, 19, and 45 percentage points, respectively.

Some conclusions

From the preceding discussion, using data on 500 occupations, we can draw the following four important conclusions:

- Increasingly, university-educated workers are taking jobs in which the average educational level is much lower. In some cases, this may represent a technological upgrading of the occupation; in most cases, however, it appears that other factors are at work in this process of downward occupational mobility.

- From 1971 through 1987, a rising share of male and female university-educated workers of all ages took jobs requiring just a high school education. The largest increase occurred in the 1970s and corresponded, at least in the early part of that decade, with the surge of university graduates onto the job market and the declining wage premium of a university degree. Nevertheless, the share of university-educated workers taking high school jobs continued up to 1987, albeit at a decreasing rate. Between 1987 and 1995, by way of contrast, the percentage of university-educated men and women in high school jobs declined slightly among younger workers, while continuing to rise among others, especially older, male, university-educated workers.
- Those university-educated workers experiencing downward occupational mobility have, on the average, considerably lower functional literacy than do other university graduates.
- The considerable increase in wages of the university educated who are pursuing occupations in educational tier 4 reflects a shortage of university-educated workers with the functional literacy that we ordinarily conceive as going with such academic credentials.

In sum, once functional literacy is taken into account, there is no contradiction between a shortage (and rising real wages) of qualified university graduates and an increasing number of university graduates taking high school jobs. □

Footnotes

ACKNOWLEDGMENT: We thank Chad Stone for forcefully raising the paradox we have attempted to answer. We are grateful to Robinson Hollister, Zora Pryor, Elliott Sulcove, and Chris Tilly, as well as the participants in the Tri-College Summer Economics Seminar, for their comments on a previous draft of this article. We also acknowledge the help of Richard Venizky, in working with the National Adult Literacy Survey, and Andrew Kolstad of the National Center for Education Statistics, for providing us with the requisite data. Finally, we are grateful to Maury Gittleman at the Bureau of Labor Statistics and to John Priebe and Lynn Weidman for providing the software and parameter estimates necessary to convert 1970 Census Bureau occupation categories into 1980 Census Bureau occupation categories.

[1] See Daniel E. Hecker, "Reconciling conflicting data on jobs for college graduates," *Monthly Labor Review*, July 1992, pp. 3–12; and the following two articles in the December 1995 issue of the *Review*: Daniel E. Hecker, "College graduates in 'high school' jobs: a commentary," p. 28; and John Tyler, Richard J. Murnane, and Frank Levy, "Are more college graduates really taking 'high school' jobs?" pp. 18–27.

[2] Actually, Hecker uses phrases such as "jobs that traditionally do not require a college degree" (see, for example, "Reconciling conflicting data," p. 3) and the like, definitions that include high school graduates who take jobs traditionally requiring some college, but less than 4 years. We approach the definition of "high school jobs" quite differently. Also, note that we use the term "university," whereas Hecker (as well as Tyler, Murnane, and Levy) uses "college." We take the two to be essentially equivalent, but use the former because of its broader connotation.

[3] John H. Bishop, "Is the Market for College Graduates Heading for a Bust? Demand and Supply Responses to Rising College Wage Premiums," *New England Economic Review* (special issue on earnings inequality) (Boston, Federal Reserve Bank, 1996), reviews a number of Department of Labor projections of job growth in various occupations. After showing that many of these projections diverged considerably from what actually happened, he analyzes various sources of possible errors.

[4] Hecker, "Reconciling conflicting data," p. 8.

[5] Lawrence Mishel and Ruy A. Teixeira, "The Myth of the Coming Labor Shortage," *The American Prospect*, Fall 1991, pp. 98–103.

[6] Bishop, "Is the Market for College Graduates Heading for a Bust?"

[7] Problems arise in matching the 1971 Census Bureau occupation categories used in the March 1971, 1972, and 1979 CPS files to the 1980 and 1990 Census Bureau occupation categories used in the March 1987 and 1995 CPS files, respectively. Our elaborate procedure for conversion is based on work done at the Census Bureau by Clifford C. Clogg, Donald B. Rubin, Nathaniel Schenker, Bradley Schultz, and Lynn Weidman, "Multiple Imputation of Industry and Occupation Codes in Census Public-use Samples Using Bayesian Logistic Regression," *Journal of the American Statistical Association*, March 1991, pp. 68–78; and Lynn Weidman, "Final Report—Industry and Occupation Imputation," report series number 3, SRD/89/03 (U.S. Department of Commerce, Bureau of the Census, April 1993), and is described in the appendix.

[8] This result is the adjusted coefficient of determination (R^2), with the average score of the prose, document, and quantitative tests as the dependent variable and years of schooling as the independent variable. Note that we plotted linear, logarithmic-linear, and logarithmic specifications of the relationship.

[9] A special issue of the *New England Economic Review* (May/June 1996) is devoted to empirical investigation of these effects.

[10] Andrew Kolstad of the National Center for Education Statistics supplied the raw data. He is not, however, responsible for the use that we have made of these data. A similar type of test for functional literacy has been employed in several previous studies; see, for instance, Irwin S. Kirsch and Ann Jungeblut, *Literacy Profiles of America's Young Adults*, report 16-L-02 (Princeton, NJ, Educational Testing Service, 1986); and Irwin S. Kirsch, Ann Jungeblut, and Anne Campbell, *Beyond the School Doors: The Literacy Needs of Job Seekers Serviced by the U.S. Department of Labor* (Washington, and Princeton, NJ, U. S. Department of Labor and Educational Testing Service, 1992). A variant of this type of test has also been used in an international comparison of functional literacy by the Organization for Economic Cooperation and Development, although its scores are not directly comparable with those of the National Adult Literacy Survey. (See *Literacy, Economy and Society* (Paris, Organization for Economic Cooperation and Development, 1995).)

[11] In the discussion that follows, we draw upon Anne Campbell, Irwin S. Kirsch, and Andrew Kolstad, *Assessing Literacy* (Washington, U.S. Department of Education, National Center for Educational Statistics, 1992); and Irwin S. Kirsch, Ann Jungeblut, Lynn Jenkins, and Andrew Kolstad, *Adult Literacy in America: A First Look at the Results of the National Adult Literacy Survey* (Washington, National Center for Educational Statistics, 1993). The National Adult Literacy Survey speaks only of "literacy," but we have added the modifier "functional" to distinguish this test from the old-fashioned literacy tests that focused only on the *ability* to read, rather than the *comprehension of what* is read.

[12] The conclusions set forth in this section can be established more exactly by means of a regression analysis in which the dependent variable is the literacy score of the individual and the independent variables include not just the usual demographic and personal background variables, but also the average literacy scores of all who are in the same occupation as the person being sampled. The calculated regression statistic of this variable is positive and

statistically significant at the .05 level. It indicates that, for every additional average year of education of the people in that occupation, a given individual has a functional literacy score on all scales that is about 4 units higher.

[13] See, for example, Frederic L. Pryor, *Economic Evolution and Structure: The Impact of Complexity on the U.S. Economic System* (New York, Cambridge University Press, 1996), tables 3–1 and A–7.

[14] See, for instance, Kenneth I. Spenner's two articles, "Deciphering Prometheus: Temporal Change in Work Content," *American Sociological Review*, December 1983, pp. 824–37; and "Technological Change, Skill Requirements, and Education: The Case for Uncertainty," in Richard M. Cyert and David C. Mowery, eds., *The Impact of Technological Change on Employment and Economic Growth* (Cambridge, MA, Ballinger Publishing Company, 1988), pp. 131–84.

[15] See, for example, two articles by Peter Cappelli: "Are Skilling Requirements Rising? Evidence from Production and Clerical Jobs," *Industrial and Labor Relations Review*, April 1993, pp. 515–30; and "Technological and Skill Requirements: Implications for Establishing Wage Structures," *New England Economic Review*, special issue on earnings and inequality, May/June, 1996, pp. 139–54.

APPENDIX: Statistical considerations

Imputing 1994 Census Bureau occupation codes for the March 1971 and 1979 CPS samples. The research results presented in this article require a single, consistent set of detailed occupation categories for all 4 years examined (1971, 1979, 1987, and 1995). Unfortunately, such data are not available, as the detailed (that is, "three-digit" or "500-category") occupation categories used in the March 1971 and 1979 CPS samples are quite different from those used in the March 1987 and 1995 samples. This problem has hampered previous research on the relation between occupation and other variables.

Before every decennial census in the United States, the Census Bureau revises its definitions of detailed occupation and industry categories. Most of the time, these revisions are minor. In preparing for the 1980 census of population, however, the Census Bureau developed a new occupation classification system that was significantly different from the systems used in the previous two censuses. As a result, many 1980 detailed occupation categories are not comparable to earlier categories with similar names. For example, of the people classified as "accountants" according to the 1960 and 1970 systems, some would be classified as "accountants and auditors" under the 1980 system, while others would be classified as "financial managers," "other financial officers," "inspectors and compliance officers, except construction," and "bookkeepers, accounting, and auditing clerks." A detailed description of the logic behind the changes is given in a 1989 Census Bureau publication.[1]

This dramatic change in classification systems has made it difficult to compare post-1980 occupation data with previous occupation data. In particular, the March CPS's used the 1960 definitions from March 1964 through March 1970, the 1970 definitions from March 1971 through March 1982, the 1980 definitions from March 1983 through March 1991, and a slight revision of the 1980 definitions from March 1992 through March 1995. Thus, it is possible to compare occupation data within the range from 1964 through 1982 and within the range from 1983 through 1995, but not across the two ranges.

Responding to this problem in the late 1980s, the Census Bureau took a subsample of 127,125 persons in the experienced civilian labor force from the 1970 Census of Population and "double coded" each individual. In other words, the Bureau went back to the original descriptions of occupations recorded by the surveyors in 1970 and determined in which 1980 occupation category these persons would be classified. Because each person in the subsample had already been classified according to the 1970 system, the Census Bureau now had two occupation classifications.

Then, a group of Federal Government economists and statisticians used this double-coded sample to develop a separate statistical model for each 1970 occupation.[2] These researchers followed a nested logit approach to determine the probability of a person being classified into various 1980 occupation categories, given that person's 1970 occupation category, as well as his or her sex, race, age, education level, class of job, industry of job, average hours worked per week, average weeks worked per year, and yearly earnings. Instead of reporting one set of parameters for each model, with standard errors attached, they reported *five* sets of parameters for each model, randomly generated from the asymptotic normal posterior distribution of the estimated parameters for each model. These five sets of estimated parameters may be used for multiple imputation of 1980 occupation codes for any individual in any comprehensive set of data utilizing 1970 occupation codes.

For this article, we only divided occupations into four broad categories (tiers). However, the empirical use of these categories required detailed knowledge about the average level of education in each detailed occupation, as well as knowledge of the correct detailed occupation to which to assign each working individual. In order to obtain such knowledge, we needed to convert the detailed occupation category for each individual in the March 1971 and 1979 CPS samples to the newer categories used in the March 1987 and 1995 CPS samples.

We acquired the necessary software and the tens of thousands of related parameter estimates from Lynn Weidman and John Priebe of the Census Bureau. After one small change in the software, we applied the estimates to the March 1971 CPS sample to impute five 1980 occupation codes for each employed person in our sample. We also made a few minor manipulations to update each of these 1980 codes to the revised 1980 occupation codes used in the March 1994 CPS.[3] All further calculations done with the March 1971 CPS data were made five times, once for each set of imputed occupation codes, and then averaged.

The one change we made in the software was to modify the parameters used to calculate each individual's within-occupation yearly earnings quartile. Because the estimated parameters used in the imputation program were based on data from the 1970 Census, the earnings data actually refer to calendar year 1969. The March 1971 CPS sample, however, is based on each person's earnings in 1970. Between 1969 and 1970, there was some price inflation, some real wage growth, and, possibly, some relative wage changes, all of which should ideally be reflected in slightly changed earnings quartiles for 1970. To account for the first two factors, we used data from the U.S. National Income and Product Accounts on the wage and salary component of national income and on total civilian employment, to calculate average wage and salary earnings per employed person in both 1969 and 1970. We then shifted up each of the within-occupation earnings quartiles by exactly the percent change in this number. We made no attempt to adjust for any relative wage changes that may have occurred between 1969 and 1970.

Finally, we prepared to apply the software to the March 1979 CPS sample, which is based on 1978 earnings data. However, the above method for adjusting earnings quartiles did not seem appropriate because, between 1969 and 1978, there were many relative wage changes, which would make a simple scaling up of the 1969 earnings quartiles inadequate for capturing the position of each individual in the 1978 earnings distribution. Therefore, we decided to estimate the actual within-occupation earnings quartiles for 1978. In some detailed occupations, though, there were only a few observations, so the earnings quartiles could not be determined with much accuracy. For this reason, we pooled the numbers we obtained with those from the following year's March CPS and the previous year's March CPS.

In other words, to estimate the within-occupation earnings quartiles to be used for imputation with the March 1979 sample, we first pooled all the observations from the March 1978, 1979, and 1980 samples, thus tripling our sample size. Just before pooling, we adjusted the March 1978 earnings numbers upward (and the 1980 numbers downward) by the percent difference in wage and salary earnings per employed person between 1977 and 1978 (and between 1978 and 1979). This pooled data set gave us enough observations to estimate the earnings quartiles within most detailed occupations with a high degree of certainty. These estimates were then used on each individual in the March 1979 sample to calculate his or her within-occupation yearly earnings quartile.

After adjusting the earnings quartiles, we were left with one additional problem related to the education variable before we could impute new occupations to the March 1979 sample. When the Census Bureau estimated the original nested logit equations, anywhere from zero to five education dummies were included in the set of equations for each occupation. The choice of how many education dummies to include, and what levels of education to associate with each dummy variable, was apparently based on trial and error, as well as some statistical analysis. Because the average level of education in the population, and within many detailed occupations, increased dramatically between 1970 and 1979, it is unlikely that the original education categories, based on 1970 values, were ideal for making imputations with the March 1979 data. As the original values were not based on anything as objective as quartiles, however, it was difficult to know the best way to update these categories.

Accordingly, we decided to make an assumption similar to the one made about yearly earnings: that it was an individual's *relative* education (rather than the absolute amount of education) within his or her 1970 Census Bureau occupation category that influenced in which 1980 Census Bureau occupation category the individual belonged. With earnings, however, the same relative scale (quartiles) had been used for all occupations, whereas with education, a different and unknown relative scale had been used for each occupation. Thus, the first estimates we had to make were the relative education categories (in percentile terms) implicitly used in the original software. To do this, we used a pooled sample of the March 1971 and 1972 data to estimate, for each detailed occupation, the percentile points of the education distribution at which the original education categories changed. For example, in 1970 Census Bureau detailed occupation category 1 ("accountants"), the original imputation program used four education categories (and therefore three dummy variables): 0 to 12 years of education, 13 to 15 years of education, 16 years of education, and 17 or more years of education. We used the pooled March 1971 and 1972 data to determine that 28 percent of the observations in occupation 1 would fall into category 1, 27 percent into category 2, 34 percent into category 3, and 11 percent into category 4.

Then, to prepare for imputing from the March 1979 data, we first determined the education distribution within occupation 1 in the pooled March 1978, 1979, and 1980 sample and then calculated at what level of education the 28th percentile occurred, at what level the 55th percentile occurred, and at what level the 89th percentile occurred. With occupation 1, the 28th percentile occurred at 13 years of education, the 55th at 15 years of education, and the 89th at 17 years of education. Thus, the new education categories used for occupation 1 in the March 1979 imputation were 0 to 13 years of education, 14 to 15 years of education, 16 to 17 years of education, and 18 or more years of education.

At this point, we were finally able to apply the Census Bureau software to the March 1979 sample, with modified earnings quartile and education category variables, to impute five 1980 occupation codes for each employed person in our sample. We then performed a few minor manipulations to update each of these 1980 codes to the revised 1980 occupation codes used in the March 1995 CPS. All further calculations done with the March 1979 CPS data were made five times, once for each set of imputed occupation codes, and then averaged.

We made two sets of imputations for March 1971 and two sets for March 1979, in each case one using the "retrospective" variables in the CPS and the other using the "current" variables. The retrospective variables include information on every person's occupation and industry, total earnings, typical number of hours per week worked, and number of weeks worked, during the previous year. The current variables include information on every person's current occupation and industry and how many hours per week the person currently works. There are no current variables analogous to the retrospective variables of total earnings and total weeks worked.

In calculating the wage variables for table 3, we used each individual's retrospective variables. For this first set of imputations, the analysis was done exactly as described above. However, in calculating the employment status and occupation categories for tables 1 and 2, we used the current variables. Because the imputation program requires data on total yearly earnings and on total weeks worked per year, and because there are no "current" values relating to these two variables, we had to use retrospective values for them. In most cases, this posed no additional difficulties. However, in a small number of cases, a person's current job was in a different 1970 occupation category from his or her primary job the previous year. In these instances, we used the nested logit model appropriate for the person's current occupation, but used the weeks worked and earnings data from the individual's primary job the previous year. Clearly, this is the best option available, but it does introduce an extra amount of uncertainty into the validity of the imputations.

Finally, although the text of this article focuses on the differences between the employment status and occupation categories in March 1971 and those in March 1995, we made the same full set of calculations for both March 1979 and March 1987. These results are presented in table A–1, which has the same format as table 2 in the text.

Estimating hourly wage data for the March 1971, 1979, 1987, and 1995 CPS samples. The March CPS sample includes information about each respondent's current employment status, as well as his or her employment status during the previous calendar year. The sample also includes information about the respondent's total earnings during the previous calendar year. However, only in recent years, and only for one-fourth of the sample, are data on current earnings collected. Therefore, any hourly wage rate calculations have to be based on the "retrospective" earnings and employment data referring to the previous calendar year.

Our initial step for estimating hourly wage data from the March 1979, 1987, and 1995 CPS data sets was straightforward. The more recent years of the survey, dating back to March 1976, contain data on every person's earnings, number of weeks worked, and usual hours per week worked, during the previous year. Dividing annual earnings by the product of weeks worked and usual hours worked yields an estimate of the effective hourly wage rate each person earned the previous year.

However, for the March 1971 CPS, there was insufficient information available to calculate the hourly wage rate in this way. For all the March data sets prior to March 1976, the data on weeks worked and usual hours worked per week during the previous year are much less precise than corresponding data after March 1976. Although the interviewees were asked the exact number of weeks they worked and the usual hours they worked per week, the Census Bureau surveyors recorded their answers only by indicating what range the numbers fell into. For weeks worked the previous year, the ranges were 1–13 weeks, 14–26 weeks, 27–39 weeks, 40–47 weeks, 48–49 weeks, and 50–52 weeks. For usual hours worked per week, the ranges were 1–34 and 35 or more. Fortunately, annual earnings were recorded as an actual number; however, the problem was how to estimate total hours worked the previous year from just the two coded responses.

Some researchers have simply used the midpoints of the indicated ranges as their estimates of these numbers. This may be a reasonable approach for the data on weeks worked, but it is much less satisfactory for the data on usual hours worked per week, for which there are only two ranges, with the top one having no midpoint. To get around this problem, we used a more complex approach, similar in spirit, but not in all of the details, to that used by Chinhui Juhn, Kevin M. Murphy, and Robert H. Topel.[4] Our basic approach was to employ the complete data from the more recent March CPS data sets to estimate an econometric model for inferring the number of weeks worked and usual hours worked from the set of variables available on each individual in the March 1971 sample. An outline of our technique follows.

The first step was to choose the data set to use for estimating the model. Juhn, Murphy, and Topel used the maximum number of observations available to them by pooling all of the individual observations from the March 1976 through March 1990 CPS samples. This gave them more than 1 million observations for their regressions. However, their choice raises the question of whether the underlying behavior, as well as the precise definitions of economic variables, used in March 1990 are consistent with those used in March 1971. In particular, because the definitions of most detailed occupations were radically changed after March 1982, it is problematical to utilize occupation categories as one of the estimating variables. Perhaps for this reason, Juhn, Murphy, and Topel did not use occupation categories in their model. Our calculations, based just on the data for March 1976 through March 1982, suggest that occupation category is one of the most important predictors of weeks worked and usual hours worked per week. Therefore, we pooled only the individual observations from the March 1976 through March 1982 CPS samples. This still gave us several hundred thousand observations to work with, as well as more consistent definitions and less likelihood of behavioral changes.

The second step was to separate the data into subgroups based on the most important characteristics. We divided our data into 48 such groups. Each person was assigned to one of the groups based on which of the six weeks-worked categories, which of the two usual-hours-worked-per-week categories, and, for the week before the March survey, which of the four employment status categories (employed and at work the previous week, employed but not at work the previous week, unemployed, or not in the labor force) the respondent was classified into. For the estimation of usual hours worked, we pooled the data across the six weeks-worked categories

Table A-1. Years of education and occupational tiers[1] of prime-age (25 to 49 years) workers, 1979 and 1987

[In percent]

Percentages and tiers	Highest educational attainment				
	Total	High school dropout	High school diploma only	Some university courses	University degree
Percentage by rows, 1979					
Total	100.0	17.1	39.7	19.0	24.2
Tier:					
1	100.0	41.4	44.7	10.8	3.0
2	100.0	22.5	52.3	18.7	6.6
3	100.0	5.8	41.7	27.4	25.2
4	100.0	.7	6.1	9.9	83.2
Percentages by columns, 1979					
Total	100.0	100.0	100.0	100.0	100.0
Tier:					
1	22.5	54.4	25.3	12.9	2.8
2	24.4	32.0	32.1	24.0	6.6
3	38.4	13.0	40.3	55.5	40.0
4	14.7	.6	2.3	7.7	50.6
Percentages by rows, 1987					
Total	100.0	11.2	39.7	21.4	27.7
Tier:					
1	100.0	30.4	52.6	12.9	4.1
2	100.0	15.0	54.9	21.4	8.8
3	100.0	3.6	37.2	29.3	30.0
4	100.0	.4	6.9	11.2	81.5
Percentages by columns, 1987					
Total	100.0	100.0	100.0	100.0	100.0
Tier:					
1	20.6	55.8	27.2	12.5	3.0
2	22.8	30.4	31.5	22.8	7.2
3	41.3	13.1	38.7	56.7	44.7
4	15.3	.6	2.7	8.0	45.1

[1] Tiers are defined by average education of those in the occupations in 1979. Tier 1: 10.5 or fewer years; tier 2: 10.6 to 12.0 years; tier 3: 12.1 to 14.5 years; tier 4: 14.6 or more years.

NOTE: Percentages by rows show the share of workers with different occupations engaged in occupations with different average degrees of education of practitioners in 1979. Percentages by columns show the percentages of workers with a given education who are engaged in occupations in which the average level of education of practitioners in 1979 varied.

SOURCE: Current Population Surveys, March 1979 and March 1987.

and therefore had only eight separate models, or subcategories, to estimate.

In the two largest of these eight subcategories, consisting of people employed and at work the previous week, we had data on the number of hours worked that week. Because such data correlate well with usual hours worked per week the previous year, we followed Juhn, Murphy, and Topel's strategy of using the other variables to estimate the gap between usual hours worked per week the previous year and hours worked the previous week, rather than estimating the former directly. For those individuals who usually worked 35 or more hours per week the previous year, we regressed the logarithm of this gap on 50 dummy variables, for weeks worked last year (5 dummy variables), years of education (5), real earnings in 1976 dollars (7), occupation category (11), industry category (10), class of job (2), sex/race category (3), marital status (3), and potential years of experience (4), an estimate derived from the person's age and number of years of education. After dropping the industry, sex/race, and marital status categories (due to lack of statistical significance), we ended up with an *R*-squared of about .48.

For those individuals who usually worked 1 to 34 hours per week the previous year, we followed a similar procedure, except that we used the gap instead of the logarithm of the gap as our dependent variable. After dropping the industry, class-of-job, and marital status categories from this regression, we ended up with an *R*-squared of about .51. In applying these two sets of results to individuals in the March 1971 data, we used as our estimate of usual hours worked per week the previous year the number of hours worked in the previous week plus the likely gap between the two, given the person's characteristics. Because hours worked the previous week is already a good estimator of usual hours worked per week the previous year, and because we were able to explain half of the remaining difference as well, the overall fit between the two models was very good.

For the other six, much smaller, subcategories, the individuals worked the previous year, but not in the week before the survey. We could not, therefore, use the same approach as just set forth. Instead, we regressed either the level of usual hours worked per week the previous year or the logarithm of this level on the same set of 50 dummy variables described above. We then dropped those variables which had relatively little statistical significance and reestimated the model. For the three subcategories in which the individuals usually worked 35 or more hours per week the previous year, we used the logarithm of usual hours worked as the dependent variable. For the three subcategories in which the individuals usually worked 1 to 34 hours per week the previous year, we used the actual number of usual hours worked as the dependent variable. The *R*-squared values for the six regressions ranged from .12 to .26.

For the estimation of weeks worked the previous year, we pooled the data across the two usual-hours-worked-per-week categories and the four employment status categories. We had, therefore, only six separate models to estimate. For each of the categories, we regressed the logarithm of weeks worked the previous year on 1 continuous variable and 48 dummy variables. The continuous variable was the logarithm of real earnings in 1976 dollars, and the dummy variables were for full-time/part-time status the previous year (1 dummy variable), years of education (5), real earnings in 1976 dollars (2), occupation category (11), industry category (10), class of job (2), sex/race category (3), marital status (3), potential years of experience (4), and real nonlabor income in 1976 dollars (7). After dropping insignificant variables, we found that the *R*-squared value for the 1–13 weeks-worked category was .40, but the *R*-squared values for the other cases ranged only from .04 to .08. Thus, for all but the first category, these results were not much better than just using the means of the subcategories from the March 1976 to March 1982 CPS samples as the weeks-worked estimate within each range.

After estimating the missing values in the March 1971 sample for weeks worked the previous year and usual hours worked per week that year, we calculated the average hourly wage rate by the same method we used on the March 1979, 1987, and 1995 samples: we divided annual earnings by the product of (estimated) weeks worked and (estimated) usual hours worked, to get an estimate of the effective hourly wage each person earned in the previous year.

Following the wage calculations, we still had to face the well-known "top-coding problem" for all 4 years. In the March 1971 and 1979 samples, any person who reported his or her annual earnings as greater than or equal to $50,000 was listed as having annual earnings of exactly $50,000. Thus, it is impossible to know whether such a person earned exactly $50,000 or perhaps much more than that in the previous year. Similarly, in the March 1987 and 1995 samples, any person who reported his or her annual earnings as greater than or equal to $100,000 was listed as having annual earnings of exactly $100,000.

The impact of this top coding is most severe for the March 1995 CPS and is least severe for the March 1971 survey. In the March 1995 data series (based on earnings in 1994), 1.6 percent of all persons who were employed during at least part of 1994 had their annual earnings top coded. However, this number is significantly larger for some subpopulations. For example, 5.4 percent of persons with a university degree or higher and 9.3 percent of white males with a university degree or higher were top coded. Similarly, 16 of the approximately 500 detailed occupations had more than 10 percent of their sample top coded. In the March 1971 data series (based on earnings in 1970), annual earnings are top coded at $50,000. Based on the personal consumption expenditures component of the gross domestic product price deflator, this would be equivalent to $178,135 in 1994 dollars. Therefore, only 0.2 percent of all persons who were employed during at least part of 1970 had their annual earnings top coded. The figure was 0.7 percent for persons with a university degree or higher and 1.1 percent for white males with a university degree or higher. Only one detailed occupation, physicians, had more than 6 percent of its sample top coded.

There are several different ways to deal with the "top-coding problem." The simplest and most common method is to ignore it and just assume that everyone reported as earning $50,000 in March 1971 and 1979, or as earning $100,000 in March 1987 and 1995, actually earned that amount. This clearly biases downward any estimate of the mean earnings level over the entire population. It also biases downward any estimates of occupation-specific mean earnings levels, with the degree of bias depending on how many top-coded observations occur in each occupation. For example, the estimate of the mean earnings of physicians might be strongly biased downward, while the estimate of the mean earnings of janitors might not be biased at all (if no janitors earn above the top code).

A second method for dealing with top coding is to find an independent estimate—perhaps from data from the Internal Revenue Service—of the average earnings of people who are earning above the top code and then apply this number to everyone in the sample who is top coded. If the other data source is reliable, this eliminates the bias in calculating the mean earnings level over the entire population. However, it still leaves a bias in calculating the variance in the mean earnings level over the population. Also, and more importantly for our work, it changes the bias in unpredictable ways when one calculates occupation-specific mean earnings levels.

A third way to deal with the top-coding problem is to estimate statistically, from the CPS data series itself, the missing or truncated right-side tail of the earnings distribution for the entire population. A modified maximum-likelihood method for doing this has been worked out and tested by Sandra A. West.[5] Then, the mean earnings level within this estimated tail can be calculated and assigned to everyone in the sample who is top coded. This has many of the same advantages and disadvantages as the second method outlined above.

We have adopted a fourth method, namely, estimating statistically, from the CPS data series itself, the missing or truncated right-side tail of the earnings distribution within each detailed occupation. This can be done using the modified maximum-likelihood method of West applied to each occupation separately. Then, the mean earnings level within each of these estimated tails can be calculated and assigned to everyone in that occupation who is top coded. The technique eliminates the bias in calculating the mean earnings level over the entire population. It also reduces, but does not eliminate, the bias in calculating the variance in the mean earnings level over the population. Finally, and most importantly for our work, it eliminates the bias in calculating occupation-specific mean earnings levels.

In order to implement this strategy, we had to deal with two problems. West's research suggests the following three-step strategy for estimating the right-side tail of an earnings distribution: (1) calculate the mean value of earnings over the entire (nontruncated) distribution; (2) assume that the distribution, from the mean point on, is a simple Pareto distribution; and (3) use the data from the mean point on to estimate the single critical parameter of the simple Pareto distribution. Our first problem had to do with the first step: we were unable to calculate the mean value of earnings over the entire distribution because we had only the truncated data. West correctly points out that if just the top 2 percent or 3 percent of the sample is truncated, then the so-called Windsorized mean can be substituted for the actual mean. The Windsorized mean is just a standard mean calculated over all the observations, using the truncated or top-coded values of earnings wherever they occur. No attempt is made to adjust the top-coded values upward.

Unfortunately, in some of our detailed occupation samples, more than 5 percent of the observations were top coded. In these cases, the Windsorized mean becomes a less acceptable approximation of the true mean. We dealt with this by using information on the median value of earnings. As long as the truncation is less than 50 percent of the entire sample, the calculated median is unaffected by the degree of truncation. Also, it is clear that, with a typical skewed-right earnings distribution, the median value of earnings is less than the mean value and is usually equal to about 85 percent to 95 percent of the mean value. Thus, if the calculated Windsorized mean is below or slightly above the median, some multiple (something between 100/85 and 100/95) of the median may be a better predictor of the true mean than is the Windsorized mean.

To lean on the side of caution, we used, as our estimate of the true mean level of earnings in each detailed occupation, the maximum of (1) the Windsorized mean level of earnings in that occupation, (2) 100/95 times the median level of earnings in that occupation, and (3) the median value of earnings in the entire population. This last term was included as a final "defense" against perverse results, as the modified maximum-likelihood method fails much more dramatically if the starting point of the distribution is too far to the left rather than too far to the right. After estimating the mean earnings level for each occupation in this way, we estimated the shape of the missing part of the tail using West's modified maximum-likelihood method. Finally, from the estimated tail, we calculated the estimated mean value of the top-coded cases within each occupation and imputed this value to each person with top-coded earnings within that occupation.

The second problem we ran into in implementing our strategy was that we had too few observations on some of our detailed occupations to generate reliable results. To handle this problem, we estimated the mean value of earnings in the top-coded cases at two different levels: once at the most detailed occupation level (that is, the three-digit or 500-category level) and once at a slightly more aggregated level (the two-digit or 50-category level). For every detailed occupation for which there were at least 300 observations, we used the estimated value for that occupation. For every detailed occupation for which there were fewer than 300 observations, we used a weighted average of the value calculated for that detailed occupation and the value calculated for the two-digit occupation of which the detailed occupation was a component. The weight given to the estimate from the detailed occupation was equal to the number of observations divided by 300. In addition, if one of the two estimates was greater than $200,000 in March 1971 and 1979, or $400,000 in March 1987 and 1995, then only the remaining estimate was used. Finally, if both of the estimates were greater than $200,000 or $400,000 for the respective samples, then the estimated mean value of the top-coded cases was set at $200,000 or $400,000, respectively. This last case occurred between zero and three times over the 4 years examined.

Footnotes to the appendix

[1] *The Relationship Between the 1970 and 1980 Industry and Occupation Classification Systems*, Technical Paper 59 (U.S. Department of Commerce, Bureau of the Census, 1989).

[2] See Clifford C. Clogg, Donald B. Rubin, Nathaniel Schenker, Bradley Schultz, and Lynn Weidman, "Multiple Imputation of Industry and Occupation Codes in Census Public-use Samples Using Bayesian Logistic Regression," *Journal of the American Statistical Association*, March 1991, pp. 68–78; and Lynn Weidman, "Final Report—Industry and Occupation Imputation," report series number 3, SRD/89/03 (U.S. Department of Commerce, Bureau of the Census, April 1993).

[3] See *Current Population Survey, March 1994: Technical Documentation* (U.S. Department of Commerce, Bureau of the Census, 1994).

[4] Chinhui Juhn, Kevin M. Murphy, and Robert H. Topel, "Why Has the Natural Rate of Unemployment Increased Over Time?" *Brookings Papers on Economic Activity*, no. 2, 1991, pp. 75–126.

[5] Sandra A. West, "Estimation of the Mean from Censored Income Data," *Proceedings of the 1986 Annual Meeting of the American Statistical Association, Vol. 2: Survey Research Methods* (Washington, American Statistical Association, 1986), pp. 665–70; and "Measures of Central Tendency for Censored Earnings Data from the Current Population Survey," *Proceedings of the 1987 Annual Meeting of the American Statistical Association, Vol. 4: Business and Economic Statistics Section* (Washington, American Statistical Association, 1987), pp. 751–56.

Table 2. Years of education and occupational tiers[1] of prime-age workers, 1971 and 1995

[In percent]

Percentages and tiers	Highest educational attainment				
	Total	High school dropout	High school diploma only	Some college	College degree
Percentages by rows, 1971					
Total	100.0	28.5	41.4	13.4	16.7
Tier:					
1	100.0	58.3	36.4	4.5	.8
2	100.0	34.9	51.8	10.5	2.9
3	100.0	11.0	49.4	23.0	16.7
4	100.0	1.4	9.4	11.6	77.6
Percentages by columns, 1971					
Total	100.0	100.0	100.0	100.0	100.0
Tier:					
1	28.5	54.1	23.2	8.9	1.2
2	41.4	32.0	32.7	20.5	4.5
3	13.4	13.3	41.2	59.4	34.6
4	15.7	.6	2.9	11.2	59.7
Percentages by rows, 1995					
Total	100.0	8.9	32.5	28.9	29.6
Tier:					
1	100.0	26.1	49.2	20.4	4.3
2	100.0	12.0	45.9	33.1	9.0
3	100.0	2.4	28.2	37.0	32.4
4	100.0	.5	5.7	13.4	80.3
Percentages by columns, 1995					
Total	100.0	100.0	100.0	100.0	100.0
Tier:					
1	19.9	58.2	30.1	14.1	2.9
2	21.9	29.5	30.9	25.0	6.6
3	41.6	11.5	36.1	53.2	45.5
4	16.6	1.0	2.9	7.7	45.0

[1] Tiers are defined by average education of those in the occupations in 1971. Tier 1: 10.5 or fewer years; tier 2: 10.6 to 12.0 years; tier 3: 12.1 to 14.5 years; tier 4: 14.6 or more years.

NOTE: Percentages by rows show the share of workers with different occupations engaged in occupations with different average degrees of education of practitioners in 1971. Percentages by columns show the percentages of workers with a given education who are engaged in occupations in which the average level of education of practitioners in 1971 varied.

SOURCE: Current Population Surveys, March 1971 and March 1995.

Table 4. Hourly earnings and years of education of prime-age workers, 1970–94

Educational tier[1]	Highest educational attainment				
	Total	High school dropout	High school diploma only	Some college	College degree
1970 average hourly wages (1994 prices)					
Total	$13.33	$10.32	$12.33	$14.85	$20.27
1	10.42	9.72	11.22	12.13	—
2	11.94	10.70	12.28	13.64	15.08
3	14.49	11.63	12.74	15.39	20.55
4	19.78	—	16.24	16.96	20.87
1978 average hourly wages (1994 prices)					
Total	14.38	10.64	12.68	14.10	20.36
1	11.55	10.21	12.12	14.05	—
2	12.58	11.03	12.94	13.20	13.41
3	14.90	11.55	12.72	14.37	19.96
4	20.90	—	14.67	15.1	22.15
1986 average hourly wages (1994 prices)					
Total	14.22	9.49	11.86	13.97	20.13
1	10.66	9.07	11.02	12.37	—
2	11.93	9.93	11.93	12.63	13.78
3	15.32	10.29	12.34	14.67	20.24
4	20.42	—	13.41	15.97	21.69
1994 average hourly wages (1994 prices)					
Total	14.80	9.00	11.30	13.18	22.42
1	10.01	8.37	10.42	11.43	—
2	11.33	9.40	10.89	11.77	14.53
3	16.01	11.36	12.19	14.08	21.94
4	23.00	—	14.44	15.26	25.07

[1] Occupations ranked by average education of practitioners in 1971–72. Tier 1: 10.5 or fewer years; tier 2: 10.6 to 12.0 years; tier 3: 12.1 to 14.5 years; tier 4: 14.6 or more years.

NOTE: The wage data represent average annual hourly earnings (total labor income divided by number of hours worked), deflated by the personal consumption price index in the gross domestic product accounts. The appendix describes in more detail the methods used in calculating this table.

Table 1. Years of education and occupational tiers[1] of prime-age workers, 1979 and 1987

[In percent]

Percentages and tiers	Highest educational attainment				
	Total	High school dropout	High school diploma only	Some college	College degree
Percentage by rows, 1979					
Total	100.0	17.1	39.7	19.0	24.2
Tier:					
1	100.0	41.4	44.7	10.8	3.0
2	100.0	22.5	52.3	18.7	6.6
3	100.0	5.8	41.7	27.4	25.2
4	100.0	.7	6.1	9.9	83.2
Percentages by columns, 1979					
Total	100.0	100.0	100.0	100.0	100.0
Tier:					
1	22.5	54.4	25.3	12.9	2.8
2	24.4	32.0	32.1	24.0	6.6
3	38.4	13.0	40.3	55.5	40.0
4	14.7	.6	2.3	7.7	50.6
Percentages by rows, 1987					
Total	100.0	11.2	39.7	21.4	27.7
Tier:					
1	100.0	30.4	52.6	12.9	4.1
2	100.0	1 5.0	54.9	21.4	8.8
3	100.0	3.6	37.2	29.3	30.0
4	100.0	.4	6.9	11.2	81.5
Percentages by columns, 1987					
Total	100.0	100.0	100.0	100.0	100.0
Tier:					
1	20.6	55.8	27.2	12.5	3.0
2	22.8	30.4	31.5	22.8	7.2
3	41.3	13.1	38.7	56.7	44.7
4	15.3	.6	2.7	8.0	45.1

[1] Tiers are defined by average education of those in the occupations in 1979. Tier 1: 10.5 or fewer years; tier 2: 10.6 to 12.0 years; tier 3: 12.1 to 14.5 years; tier 4: 14.6 or more years.

NOTE: Percentages by rows show the share of workers with different occupations engaged in occupations with different average degrees of education of practitioners in 1979. Percentages by columns show the percentages of workers with a given education who are engaged in occupations in which the average level of education of practitioners in 1979 varied.

SOURCE: Current Population Surveys, March 1979 and March 1987.