Children that told about how children lived all over the world. The chapter on the United States was entitled “The Richest Country in the World.” While I was somewhat uncomfortable about having that aspect of my homeland stressed, I had no reasons to question the accuracy of the title.

Things have changed. It is not clear what country is the richest in the world, but one thing is clear: The United States is certainly one of the richest. In 1990 the gross national product (GNP) was about $5.5 trillion [one thousand billion] dollars. The Japanese GNP was $2.9 trillion, West Germany’s $1.5, and $7.0 trillion for all of Western Europe [Bureau of the Census, 1993]. Of course, the absolute comparison does not take into account the size of the countries involved. If you simply divide GNP by the population, Germany and Japan are the richest countries, closely followed by the United States. But this does not take into account the cost of living in each country. A better method is to consider the gross domestic product (GDP) per capita, and then adjust this figure to a comparable currency unit, which takes into account both exchange rates and the cost of goods within each country. The results of this calculation are shown in Figure 1.1. The United States is the leader, by a slight margin.

“Who is first?” is hardly the point. We are clearly one of the richest countries in the world. Perhaps most important, we are by far the largest country with a high level of wealth per person.

What is a bit more disconcerting is that changes in our wealth are not keeping pace with changes in other countries. Figure 1.2 shows
changes in income per capita during the 1980s in several leading countries, expressed as a fraction of U.S. income per capita. On this scale the figures for the United States form a straight, horizontal line. An upward tilting line indicates that a country is gaining in wealth per person, relative to the United States, while a downward tilting line indicates that the United States is gaining on the country. All the lines in the figure tilt upward. Average incomes are increasing faster in other countries than in ours—for both the economic giants, Germany and Japan, and for relatively weak economies, such as Italy and the United Kingdom.

Clearly, the United States lost ground during the 1980s. The recession and slow recovery of the early 1990s, coupled with political upheaval in Eastern Europe associated with the collapse of the Soviet Empire, have changed the picture somewhat. As of 1992 the GDP in the United States grew more rapidly than that of its major competitors (2.1 percent for the United States, compared to .8 for Germany and 1.5 for Japan). However, 1992 was the first time in years that the United States had gained the lead. Only time will tell how permanent the 1992 figures are.

Competitive figures are not important in themselves. There is no
receipt of wealth. Otherwise the poor will resent the wealthy, and the workers will resent the shirkers—not a happy combination for a democracy. Let us look beyond the gross statistics to see how income and productivity are distributed within the country.

Productivity, measured in the value of output per worker per hour, has increased steadily since the end of World War II. However, there is a marked difference in productivity in the manufacturing and service industries. This is shown in Figures 1.3 and 1.4. Figure 1.3 shows the value of products of the goods and service industries since the 1970s. Goods and services contributed approximately equal value to the economy in the 1970s. The relative value of the service sector began to rise in the 1980s. In 1992 the value of services was about 1.4 times the value of goods. Figure 1.4 shows the employment figures, which tell a rather different story. The number of people involved in manufacturing has dropped slightly, while the number of people involved in services has almost doubled. These results are largely due to changes in technology, which permit us to do more things with fewer people. New tech-
nologies have made their impact on both manufacturing and services, but the impact on manufacturing has been far greater.

Since manufacturing jobs generally pay better than service jobs, the result of this process has been that high-wage jobs are being lost, while low-wage jobs are being created. This trend is expected to increase. This is shown in Figure 1.5, which plots the expected growth in employment in selected occupations against the average 1990 income in each group. With some exceptions, the points drop as they move to the right. This indicates that high growth has been concentrated in low-paying jobs.

How did the end of the recession of 1990–1992 affect these trends? Although economic recovery began in late 1992, unemployment remained high. As of January 1994 the unemployment rate was estimated at 6.7%, a figure that would have been considered a recession 30 years earlier. The better-paying industrial jobs were particularly hard hit. Many jobs were permanently lost, either to competition or because employers had retooled so that they could make things with fewer people. In March 1994 the New York Times reported that in 1993, during a recovery from a recession, 615,000 jobs were eliminated. A Times quotation presents the reason in a nutshell: “Manufacturing employment is primarily governed by technology, and new technology requires half the number of people in product assembly every six years” [Laurence
Figure 1.5 The relation between annual compensation and projected changes in the number of jobholders in selected occupations. Source: U.S. Bureau of the Census (1993).


Of course, new jobs were being created, but they were not good jobs. In 1989, 6 percent of the new jobs created were for temporary workers. In 1992, the first postrecession year, 26 percent of the new jobs were for temporaries. In 1993 that fell to "only" 15 percent. Skilled workers in manufacturing, laid off from $15 to $18 per hour jobs, were finding employment in the service industries at $6 to $8 per hour. The anecdotal reports and case studies are mirrored by statistics on earnings. The average weekly earnings in 1992, in constant dollars, were 7 percent below the figure for 1980 and 14 percent below the figure for 1970. This is not just a reflection of the 1990–1992 recession. Annual earnings for employees, again corrected for inflation, decreased in seven of the eight years from 1985 through 1992. We are dealing with a stable, long-term trend.

A small ray of light appeared by the end of 1994. Unemployment had fallen to below 6 percent, and there had been an increase in the number of new, high-paid jobs. Whether or not this represents a reversal of fortune or a minor blip in a long-term trend remains to be seen. Meanwhile, another, and perhaps more disturbing, trend has continued unabated.

Although the general trend has been loss of individual wealth the
losses have hardly been distributed equally. Figure 1.6 shows the percentage of income received by the lowest 20 percent of the population, the next highest 20 percent, and so on. The figure also shows the percentage of the national income received by the top 5 percent. Figures are shown for 1980 and 1991. If the relative income distribution over the population had been unchanged, the black and gray bars would be of equal height. They are not. Every group lost in its share of the national income, except the top 20 percent. Within that group, about half of the increment was due to a rise in the share of the wealthiest 5 percent.

The absolute figures on income illustrate these disparities even more dramatically. In 1980 the median income of families in the bottom 40 percent of the income distribution was [in 1991 dollars] $17,023. In 1991 it was $17,000—virtually unchanged. To gain some feeling for this figure, in 1991 a family of four with an income of $17,405 or less were defined as "working poor" (income of less than 125 percent of the poverty rate); 18.9 percent of all Americans fell into this category. By contrast, the median figure for families in the top 10 percent was $89,465 in 1980 and $102,824 in 1991, an increase of 15 percent (Bureau of Census, 1993, Tables 722, 735). The rich got richer, while the lower middle class and below stayed the same. In absolute terms there has been virtually no movement in the percentage of peo-
Figure 1.7  Trends in median family income, by ethnic status. Source: U.S. Bureau of the Census (1993).

The United States has been plagued by a persistent disparity in income across demographic groups. Figure 1.7 shows the median family income in constant dollars for the nation as a whole, and for black and white families for the period 1970–1989. In spite of major social efforts to improve the situation, the median family income of a black family was 61 percent that of a white family in 1970, 58 percent in 1980, and 58 percent in 1990. For Hispanics the relevant figures are 69 percent, 67 percent, and 63 percent. The poverty level figures reflect the same trends. In 1980, 32.5 percent of all blacks and 25.7 percent of all Hispanics were below the poverty level. In 1990 the figures were 31.9 percent and 28.1 percent.2

The increasing disparity in income extends to both blue-collar and white-collar jobs. Blue-collar workers have been particularly hard hit. In 1972 craft workers received, on the average, 98 percent of the wages of professional and technical employees. In 1987 the figure was 73 percent. Clerical wages dropped from 68 to 54 percent of the professional-technical earnings, and retail sales wages from 65 to 46 percent (Commission on the Skills of the American Workforce, 1990).

The disparities in earnings between workers and "top management"
have widened dramatically. Figure 1.8 shows the relative salaries for four different occupations: executives and administrators, professionals (physicians, lawyers, accountants), sales personnel, and mechanics. Figures are shown from 1983 to 1992. If you wanted a raise in pay, it was clearly better to be an administrator than a mechanic. Some recent figures are even more dramatic. From 1989 to 1990 the average income of an American worker, not corrected for inflation, increased a little less than 5 percent. The average increase for corporate executives was 8 percent.\textsuperscript{3} Preliminary figures for the years following the 1990–1991 recession suggest that the situation has gotten even worse. In some cases major industries have asked for wage concessions from employees while paying executives more than a million dollars a year.

Numerous commentators [e.g., Reich, 1991] have maintained that this degree of disparity in wealth is neither appropriate nor healthy for our society. The United States is not approaching the extremes of wealth and poverty seen in some third-world countries, such as India or the oil-producing Arab states. That comparison is not relevant. We are developing greater disparities of wealth than are seen in the industrial nations of Europe and Asia.

Technologies that sharpen the difference between the productivity of highly skilled and less skilled workers redistribute job opportunities in Germany and Japan in exactly the same way that they do in the United States. Advances in robotics offer equal unemployment opportunities! Several European countries are experiencing social problems
American workforce at the turn of the century will not be dominated by workers over 55. In fact, the percentage of workers in this group actually decreases. There will be a substantial increase in the percentage of workers in the 35–54 age range, at the expense of lowered percentages of workers under 35.

This trend is likely to have a major impact on the availability of psychological skills. Older workers are not simply more or less intelligent than younger workers. The old and the young have different cognitive skills. There is a predictable shift toward more knowledge about how things are done, coupled with a reduction in the speed with which new ideas are grasped (Salthouse, 1990). Thus, aging increases the value of a workforce when the workplace is static, but it may decrease the value of the same workforce if the methods and technology of the workplace are changing.

There has been a good deal of discussion about the increased feminization of the workforce. Statistically this will be a fairly small effect. The present workforce is about 45 percent female; this is expected to increase to 47 percent by 2000. What may change, but what is much harder to predict, is the sort of jobs that women hold.

The third trend is a change in ethnic constitution, away from the present dominance of whites. The changes are complex.

Figure 1.10 shows the ethnic distribution of the population in 1990
Figure 1.10 Percentage of workforce and percentage of increment of workforce in various ethnic groups, 1990 and 2005. Source: U.S. Bureau of the Census (1993).

and the projected distribution in 2005. Clearly, there is little change in the overall distribution. The picture is quite different if one looks at the ethnic distribution of increments in subpopulation sizes. This is shown in the right-hand column of Figure 1.10. Although the population as a whole will still be predominantly white, the increment to the population will be split approximately evenly between Asians and non-Hispanic whites and then minority groups. Hispanics will be the most rapidly expanding group. This reflects both a high birthrate, relative to the rest of the population, and anticipated immigration from Central and South America.

Immigration is one of the “wild cards” in the projections. The Office of Technology Assessment (OTA) pointed out that immigration accounted for 22 percent of the growth in workers in the United States in the period 1980–1987 and projected that immigration would have an increased impact on workforce growth in the next 10 years. Immigrants are a mixed group: Some immigrants to the United States contribute at the very top of the workforce, in sciences and the professions. Most immigrants, however, are poorly educated compared with the U.S. workforce, and about one in six does not speak English (Office of Technology Assessment, 1990). Even those who do speak English often have only marginal skills, especially with the written language (Kirsch, Jungeblut, Jenkins, and Kolstad, 1993). Thus, the immediate impact of
Table 1.2  School Enrollees

<table>
<thead>
<tr>
<th></th>
<th>Canada</th>
<th>France</th>
<th>Germany [West]</th>
<th>Italy</th>
<th>Japan</th>
<th>United Kingdom</th>
<th>United States</th>
</tr>
</thead>
<tbody>
<tr>
<td>High school</td>
<td>106</td>
<td>99</td>
<td>104</td>
<td>79</td>
<td>96</td>
<td>84</td>
<td>96</td>
</tr>
<tr>
<td>Beyond high school</td>
<td>70</td>
<td>40</td>
<td>33</td>
<td>31</td>
<td>31</td>
<td>25</td>
<td>72</td>
</tr>
</tbody>
</table>

Source: National Center for Education Statistics (1993), Table 385.

Note: 1990 Enrollment ratios in high school and post-high school programs in selected countries. The enrollment ratio is the number of enrolled students divided by the number of people in the appropriate age group in the population. For post-high school education, the age group used is the number of people in the 20–24 age range.

The latter figure is a bit misleading, however, because the U.S. figure includes enrollees in vocational training courses. In Germany and Japan comparable education is often provided by employers.

Table 1.2 shows enrollment in educational institutions, rather than completion. Figure 1.11 shows the high school dropout rates for the

Figure 1.11  Percentage of dropouts from high school in the population of 16–24-year-olds. Source: National Center for Educational Statistics (1993).
than cognitive skills problems. However, employers in high-technology industries such as IBM and Motorola dissented. These employers expressed more concern over cognitive qualifications (Commission on the Skills of the American Workforce 1990).

What are we to make of these reports? The work ethic problem is virtually impossible to evaluate on the basis of subjective reports. Employers have been complaining for generations that employees do not work hard enough. On the other hand, it would be possible to determine objective indices of workforce ethic, such as absenteeism. Zuboff (1989) has pointed out that this has been a problem since the nineteenth century. We simply do not have adequate data to make a comparison across time.

Although employers seem to denigrate education, and especially the high school diploma, the same employers do pay higher wages for higher education. Figure 1.12 shows the 1991 median income for male and female full-time workers, by educational level. Men and women are shown separately because women generally occupy lower-paid jobs than men do. Income and education are closely associated. The median income of a person without a high school diploma was about half of the median income of a person with a college degree. In spite of all the jokes, it pays to be educated.

Figure 1.12, together with some other statistics in this section,
Figure 1.13  Percentage of students reading at or above indicated reading levels. Source: National Center for Educational Statistics (1993).

Figure 1.13 shows NAEP reading assessments for 17-year-old students between 1975 and 1990. In spite of great concern over "why Johnny can't read," there has obviously been little improvement overall. However, focusing on the overall picture masks some important details. Figure 1.14 shows the percentage of students reading at Carroll's intermediate level, which is probably the minimal level of competence needed in a job beyond entry level. The data are shown separately for whites, blacks, and Hispanics. Two trends are immediately apparent: while blacks and Hispanics still come out of the K–12 system less well prepared than whites, the gap has been reduced dramatically. About 50 percent of the black and Hispanic 17-year-olds were reading at the intermediate level in 1974–1975. Fifteen years later over 70 percent were at that level.

Unfortunately, there has been little change in the percentage of readers at the adept and advanced skill level. In fact, the percentage of readers at the advanced level has declined slightly. It is not clear whether or not this is a problem. As Carroll points out, we do not know what percentage of very good readers we need, because we do not have a good idea of how much essential reading (manuals, directions, legal instructions, and the like) is written at the adept level or above, nor do we know how many people need to understand this material.

Mathematical ability is generally considered a necessary cognitive
skill second only to reading. NAEP mathematics scores have been grouped as follows:

1. Basic. Can do arithmetic operations. [NAEP score 250]
2. Moderately complex. Can use fractions and percentages and understands simple geometric concepts. [NAEP score 300]
3. Advanced. Can use mathematical reasoning to solve multistep problems and can use algebra. [NAEP score 350]

Figure 1.15 shows the percentage of 17-year-old students scoring 300 or more on the 1990 NAEP test. The data are shown separately for males and females and for various ethnic groups.

A great deal has been written about male-female differences in mathematical ability. More extreme proponents of a difference sometimes seem to write as if women, generically, "can't do math." The low point in this debate may have been reached in 1992, when the Mattel Corporation marketed a talking Barbie doll that complained that "math class is hard." The facts are more complex. As Figure 1.15 shows, 17-year-old males do very slightly better than females. Some of this difference is because girls tend to take fewer high school mathematics classes than do boys, but that is not the whole story. On the average, males
outscore females at every level of mathematics education. For instance, males who have taken precalculus or calculus courses averaged 347 on the NAEP mathematics test, compared to 341 for females with equivalent training.

The mathematics skills of black and Hispanic students lagged markedly behind those of whites and Asians. The difference is considerably larger than the male-female difference. Like the male-female difference, ethnic differences remain after taking the amount of mathematical training into account. This is consistent with the observation that black and Hispanic students are underrepresented in the more technical fields of higher education, which place a premium on mathematics skills.

The data presented so far have focused on all high school students. Figure 1.16 shows changes over time in the Scholastic Aptitude Test (SAT) scores achieved by college-bound high school seniors. Proficiency dropped from the 1960s until the late 1970s, rebounded slightly, and dropped again. The 1991 and 1992 mathematics scores were the lowest ever.

Figure 1.17 shows ethnic group differences in SAT scores in 1991. The picture is similar to the picture obtained from the NAEP tests.
Figure 1.16  SAT scores of college-bound high school seniors. Source: National Center for Education Statistics (1993).

Figure 1.17  SAT scores, by ethnic groups, 1992. Source: National Center for Education Statistics (1993).
Blacks and Hispanics do markedly worse than whites and Asians. Asians show a marked superiority in mathematics but have lower verbal scores that are intermediate between the scores of whites and other ethnic groups.

Where do we stand internationally? Reading scores cannot be compared because students read different languages. Mathematics scores can be compared, and the United States does not do very well. In 1981 an international comparison of mathematics knowledge in both 13-year-olds and twelfth graders found that the United States ranked last compared with Japan, Canada, France, and the United Kingdom. In 1992 a larger survey of 13-year-olds found that the United States had moved up a bit. We were tied for last place with Spain (Figure 1.18).

Ideally, we should compare U.S. twelfth graders in 1992 with foreign twelfth graders, as was done in 1981. Unfortunately, the data are not available, but it is doubtful that things have changed. The U.S. internal statistics indicate that while we have made progress, we have made little progress in improving the work of our better students. There is no reason to believe that foreign countries have slipped backward.

**CHANGES IN THE NATURE OF WORK**

What skills will be in demand in the coming workplace? This is a difficult question to answer, because the forecast depends upon what as-
Table 1.1  Educational Levels Attained by the Present U.S. Workforce

<table>
<thead>
<tr>
<th></th>
<th>Not High School Graduate</th>
<th>High School Graduate</th>
<th>College [one year or more]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>18.2%</td>
<td>39.6%</td>
<td>42.1%</td>
</tr>
<tr>
<td>Whites</td>
<td>15.8</td>
<td>39.8</td>
<td>44.4</td>
</tr>
<tr>
<td>Blacks</td>
<td>22.7</td>
<td>42.4</td>
<td>34.4</td>
</tr>
<tr>
<td>Hispanics</td>
<td>39.0</td>
<td>33.5</td>
<td>27.4</td>
</tr>
</tbody>
</table>

Source: Office of Technology Assessment [1990], Table 1-3.

Note: Totals are extrapolations based on OTA estimates of ethnic groups in the workforce.

mane standards. A design standard specifies how a product is to be made, while a performance standard specifies what the product must be able to do when it is used. Marshall and Tucker argue that the same distinction applies to education. Requiring that all U.S. citizens attend school until age 16 is a design standard. Requiring that all high-school graduates be able to summarize the argument in a newspaper editorial is a performance standard. Following Marshall and Tucker, we must distinguish between the design standards used to prepare entrants in the workforce [and to maintain their skills once they are in the workforce] and the performance standards that educated workers are expected to meet.

Workforce capabilities are usually expressed in terms of education level. The United States does very well, relative to the rest of the world, in terms of percentages of its citizens who have graduated from high school or have attended college. However, there are major causes for concern because educational qualifications, like wealth, are not spread equally across our population.

Table 1.1 presents the statistics for education levels attained by the present workforce and for the three major ethnic groups [whites, blacks, and Hispanics]. As the table shows, the workforce as a whole is relatively well educated. Although comparable figures are not available worldwide, I suggest that the percentages are quite high compared with other countries. However, there are obvious disparities between ethnic groups.

In order to estimate the future workforce, we need to look at present enrollment in the schools. Table 1.2 compares enrollment rates in high schools and postsecondary education in seven industrialized countries. The United States is one of the leaders insofar as high school enrollment is concerned, and the leader, by a considerable margin over Europe and Japan, in postsecondary [college and university] education.
Table 1.2 School Enrollees

<table>
<thead>
<tr>
<th>Country</th>
<th>Canada</th>
<th>France</th>
<th>Germany [West]</th>
<th>Italy</th>
<th>Japan</th>
<th>United Kingdom</th>
<th>United States</th>
</tr>
</thead>
<tbody>
<tr>
<td>High school</td>
<td>106</td>
<td>99</td>
<td>104</td>
<td>79</td>
<td>96</td>
<td>84</td>
<td>96</td>
</tr>
<tr>
<td>Beyond high school</td>
<td>70</td>
<td>40</td>
<td>33</td>
<td>31</td>
<td>31</td>
<td>25</td>
<td>72</td>
</tr>
</tbody>
</table>

Source: National Center for Education Statistics (1993), Table 385.

Note: 1990 Enrollment ratios in high school and post-high school programs in selected countries. The enrollment ratio is the number of enrolled students divided by the number of people in the appropriate age group in the population. For post-high school education, the age group used is the number of people in the 20–24 age range.

The latter figure is a bit misleading, however, because the U.S. figure includes enrollees in vocational training courses. In Germany and Japan comparable education is often provided by employers.

Table 1.2 shows enrollment in educational institutions, rather than completion. Figure 1.11 shows the high school dropout rates for the

Figure 1.11 Percentage of dropouts from high school in the population of 16–24-year-olds. Source: National Center for Educational Statistics (1993).
Table 1.3  Percentage of College and University Degrees Granted to Members of Various Ethnic Groups, as a Percentage of Total Degrees Granted

<table>
<thead>
<tr>
<th>Degree</th>
<th>White</th>
<th>Black</th>
<th>Hispanic</th>
<th>Asian and Other</th>
<th>Alien</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor’s</td>
<td>83.6</td>
<td>6.0</td>
<td>3.4</td>
<td>4.2</td>
<td>2.7</td>
</tr>
<tr>
<td>Master’s</td>
<td>77.7</td>
<td>4.9</td>
<td>2.6</td>
<td>3.7</td>
<td>11.1</td>
</tr>
<tr>
<td>Doctor’s</td>
<td>65.7</td>
<td>3.1</td>
<td>1.9</td>
<td>4.1</td>
<td>25.2</td>
</tr>
<tr>
<td>M.D., D.D.S., LL.B., or similar</td>
<td>84.4</td>
<td>5.0</td>
<td>3.5</td>
<td>5.7</td>
<td>1.7</td>
</tr>
</tbody>
</table>


Note: Data from 1990–91.

United States from 1972 to 1991. Overall the dropout rate has decreased from about 13 percent to just under 10 percent. As is well known, dropout rates are higher for the larger minority groups than they are for whites and Asian Americans. The dropout rate for black students is declining, but the dropout rate for Hispanics remains high. This is particularly discouraging because Hispanics are the most rapidly growing ethnic group in the United States.

Higher education figures also show a marked discrepancy between ethnic groups. Table 1.3 shows the distribution of degrees to members of different ethnic groups in 1989–1990. Among our largest minorities, blacks and Hispanics, half as many degrees were granted as would be expected on the basis of their percentage of the population. The discrepancy grows greater the higher the degree and the more technical the field. In 1990–1991, 4,164 Ph.D. degrees were granted in the physical sciences, 103 to blacks and Hispanics combined and 1,326 to nonresident aliens.

Direct Indicators of Competence

As educators frequently say, and students occasionally hear, the point of education is not to get a degree, it is to learn something. How cognitively proficient is our highly educated workforce? There are three ways to assess workers’ cognitive capabilities. One is to ask employers whether they think that their workforce performs adequately. Another way to proceed is to determine whether employers pay higher wages to more highly educated employees. Finally, cognitive competence could be measured directly, by testing people in the educational or workforce systems. Each of these methods has its advantages and drawbacks. Interestingly, each produces a somewhat different picture.
| 1 | Occupations with Most Employees                          | 2 | Occupations with Most New Jobs                           | 3 | Occupations with Highest Growth Rates                   |
|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
| Food service workers             | Food service workers             | Systems analysts                 |
| Retail sales clerks              | Retail sales clerks              | Computer programmers             |
| Personnel related jobs           | Chefs and cooks                  | Service sales representatives    |
| Secretaries                      | Registered nurses                | Health service workers           |
| Chefs and cooks                  | Janitors                         | Security guards                  |
| Janitors                         | General managers and executives   | Computer operators               |
| Truckdrivers                     | Cashiers                         | Electrical/electronics            |
| General managers and executives   | Truckdrivers                     | repairs and service workers      |
| General clerks                   | General clerks                   | Real estate agents               |
| Cashiers                         | General engineers                | Registered nurses                |
|                                  |                                  | Receptionists                    |
| 4 | Occupations with Most Job Losses                        | 5 | Occupations with Highest Loss Rate                      |
|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
| Farmers                          | Textile machine operators        | Farmers                          |
| Typists and word processors      | Farmers                          | Data entry key operators         |
| Clothing workers                 | Typists and word processors      | Machinists                       |
| Moving equipment operators       | Data entry key operators         | Typists and word processors      |
| Textile machine operators        | Moving equipment operators       | Clothing workers                 |
| Data entry key operators         | Textile machine operators        | Metalworkers                     |
| Metalworkers                     | Textile machine operators        | College teachers                 |
| Machinists                       | Textile machine operators        | College faculty                  |
| College teachers                 | Textile machine operators        | Quality control inspectors       |
| Quality control inspectors       | Textile machine operators        | Textile machine operators        |
Table 1.5 Growth Rates for the Workforce, Manufacturing, and Selected Service Occupations (1980–1992)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Workforce</td>
<td>90,406</td>
<td>108,437</td>
<td>1.20</td>
</tr>
<tr>
<td>Production of goods and services</td>
<td>20,285</td>
<td>18,190</td>
<td>.90</td>
</tr>
<tr>
<td>Financial and realty</td>
<td>5,160</td>
<td>6,672</td>
<td>1.29</td>
</tr>
<tr>
<td>Computer and data processing</td>
<td>304</td>
<td>817</td>
<td>2.69</td>
</tr>
<tr>
<td>Legal</td>
<td>498</td>
<td>906</td>
<td>1.82</td>
</tr>
<tr>
<td>Advertising</td>
<td>153</td>
<td>232</td>
<td>1.52</td>
</tr>
<tr>
<td>Education</td>
<td>1,138</td>
<td>1,716</td>
<td>1.51</td>
</tr>
<tr>
<td>Hotel industry</td>
<td>1,076</td>
<td>1,597</td>
<td>1.48</td>
</tr>
<tr>
<td>Personal services</td>
<td>818</td>
<td>1,107</td>
<td>1.35</td>
</tr>
</tbody>
</table>


If anything, these figures understate the extent of the shift. Manufacturing itself has changed, so that the people counted as being in industrial plants may be symbol analysts themselves.

If anything, Table 1.5 understates the dominance of Reich’s symbol analyst. Within the production industries, symbol processing has become increasingly important. The Industrial Revolution moved people away from providing the physical energy needed to transform objects, but they still had to provide the perceptions that guided the machinery. Now they provide abstract information that a smart machine uses to guide its own cutting and mixing tools. A single example, Toshiba, the Japanese electronics manufacturer, employs 4,000 in its plant at Ome, Japan; 500 are blue-collar workers, and 3,000 are engineers and scientists [Zuckerman, 1991]. All 4,000 of these workers can be counted as being “in manufacturing,” but very few of them are involved in the physical manipulation of products.

**Customizing the Product**

The Toshiba example provides a bridge to a final comment about changes in the nature of work.

When Henry Ford first began producing a production-line car, the Model T, he is supposed to have said, “The customer can have any color he wants, so long as it’s black.” This apocryphal quotation very
the task of assembling the computer. The arrows from the positions in the mental space to the line represent the equation used to map from positions in the space to positions on the line. The location that the arrow points to on the line is the predicted score for a person. The dots on the line represent the scores actually obtained. The difference between the predicted and the obtained scores represents the error term of Equation 2.1.

The horizontal line on the right represents a hypothetical “minimal level of predicted score.” People whose arrows fell above the line would be considered qualified as assemblers, though their actual performance might fall below it. People whose arrows fell below the line would be considered unqualified, though their actual performance, if they had been hired, might have been above the line.

Prediction works to the extent that the predicted scores (the arrows) fall close to the actual scores (the dots on the vertical line). The extent of the agreement is indicated by the correlation coefficient \( r \) between predicted and obtained scores. This is called the validity coefficient.

Validity coefficients (like all correlation coefficients) vary between \(-1\) and \(+1\). Since a negative correlation can always be converted into a positive one by multiplying the predicted score by \(-1\), we can consider only positive values. If the validity coefficient is one, the obtained scores will be equal to the predicted scores. In Figure 2.1 the arrows would point exactly to the dots on the right. If the validity coefficient is zero, the dots might be anywhere, no matter where the corresponding arrow is. Validity coefficients between zero and one indicate that the
Figure 2.2 Hypothetical data generated from a population with test-performance correlation of .5; sample correlation .43.

will not be perfect. If it were, all the points in Figure 2.2 would lie on the regression line, which they obviously do not. However, the predicted score would be more accurate, on the average, than a guess that everyone would achieve the mean performance score.

Extending the example, all points above the regression line represent individuals whose performance scores were higher than predicted from their test scores, while points below the regression line represent individuals whose scores were lower than predicted.

The next step is to explain how values were chosen for $A$ and $b$ in Equation 2.2, and how this relates to the correlation coefficient. If prediction were perfect ($r = 1$), the data points for all individuals would lie on the regression line. In this case $E_i$ would be zero for every case, $i$. We want to choose values $A$ and $b$ that bring us as close to this situation as possible.

$E_i$ is a value associated with an individual, just as $X_i$ and $Y_i$ are. Therefore, we can consider the error term, $e$, as a variable with its own standard deviation, $s(e)$ and variance, $s^2(e)$. In regression analysis values of $A$ and $b$ are chosen so that the expected value of $e$ is zero ($E[e] = 0$) and the variance of $e$ is minimized. If prediction were perfect, all points $E_i$ would be zero, so both $E[e]$ and $s^2[e]$ would be zero. This ideal is seldom achieved. For instance, the regression line in Figure 2.2 was chosen to minimize $s^2[e]$, but there clearly are nonzero error terms.

Now consider the case in which there is no relationship between
test and performance measures, that is, \( r = 0 \). The regression line will be flat, with \( b = 0 \) and \( A \) equal to the mean value of the performance score. (This amounts to predicting that everyone will have average performance, regardless of their test score.) The variance of the error term will be equal to the variance in the score being predicted, \( s^2(e) = s^2(y) \).

In virtually all practical situations the correlation coefficient has a value intermediate between zero and one. Therefore, it is important to develop some intuitions about what different intermediate values mean. As one would expect, there is a precise algebraic definition. The square of the validity coefficient is the percentage of variance in the criterion (performance) score that is associated with the predictor score.\(^6\) Unfortunately, the algebraic definition is not intuitive and can even be misleading. A better way of understanding the meaning of validity coefficients is to look at a concrete example.

Figure 2.2 shows some hypothetical data in which the correlation between the predictor and criterion score is .43. This is well within the range of the population validity coefficients obtained in industrial situations. Figure 2.3 shows a distortion of the same data, in which the validity coefficient is .22. In each case the regression line was calculated so that the average error in prediction is zero. However, the variance of the error term is smaller in Figure 2.2 than in Figure 2.3. More importantly, note how the regression lines have changed. In Figure 2.2,
forming at or above the minimum acceptable level. A regression analysis is conducted on the reduced set of data.

2. If the goal is to evaluate a test that is already being used for hiring decisions, data will be available only for those individuals who had test scores above the preliminary cut score (since they had to be hired) and whose performance levels are satisfactory (since the employer has continued to employ them). The regression analysis is conducted on these employees.

Figure 2.4 illustrates what happens in each case. Suppose that we define minimum levels of acceptable performance \( Y_{\text{min}} \) and a minimum cut score \( X_{\text{min}} \). These scores divide the space of scores into four regions. Region A contains individuals whose test scores are above the cut level and whose performance level is satisfactory. Region B contains individuals whose test scores are adequate, but whose performance is not. Region C contains individuals whose test scores and performance levels are inadequate. Region D contains individuals whose test scores are below the cut point, but whose performance is adequate.

In a case of perfect prediction \( r = 1 \), all scores would either be in region A (correct acceptances) or region C (appropriate rejections). Regions B and D fill up only to the extent that prediction is imperfect.

In the ideal design regression analysis is applied to data points in all four regions. In reduced design \( (1) \), where all current employees are
Figures 2.5, 2.6, and 2.7 present an example of the $s = c$ case. These figures were constructed by randomly sampling points for a hypothetical population in which the predictor-criterion correlation was .5. These points were then plotted. The vertical and horizontal lines in the panels correspond to the $X_s$ and $Y_s$ points for selection ratios of .2, .5, and .8. The efficiency of selection can be determined empirically, by comparing the ratio of the number of points above the horizontal line (qualified) and to the right of the vertical line (selected) to the number of points to the right of the vertical line (selected, regardless of qualification). The benefit of using the predictor can then be calculated using Equation 2.9. The results are shown in Table 2.1. Benefits increase as the selection ratio decreases, even though the correlation between the predictor and criterion measures is unchanged. Testing is most useful when there are many applicants for a small number of jobs.

Let us pause to consider what this means in the big picture. Virtually every analysis of economic recovery following the 1990–1991 recession has stressed the loss of permanent jobs. Many observers feel that, due to technological changes, this will be a permanent condition. If these analyses are correct, the future will be characterized by a higher ratio of applicants to jobs than the present is; selection ratios will be lower; and, therefore, employers will have more of an incentive to use screening tests than they do at present.
Figure 2.6  The example of Figure 2.5 repeated with selection and qualification ratios of .5.

Figure 2.7  Efficiency of selection with $s = q = .8$. 
Figure 2.8  Selection efficiency as a function of the validity coefficient \( r \) and selection ratio. A qualification coefficient of .3 is assumed.

\[
E^*(q,0,s) = q.  \tag{2.11b}
\]

\( E(q = s,r) \) can be thought of as a special case of Equation (2.10) in which \( q = s \), which is equivalent to saying that the selector wants to select all available qualified applicants and no one else. Clearly, this is a rather unusual employment situation.

The mathematics for solving the necessary equations are well known (Taylor and Russell 1939). A typical case is shown in Figure 2.8. The figure shows the probability of selecting a qualified person for a job that 30 percent of the population can do, given selection tests with validities ranging from .2 to .5. These figures are well within the values typically found. The efficiency of random selection (validity of zero) and the efficiency of perfect selection (validity of one) are also shown. Actual selection efficiencies have to fall between these extremes. The figure makes it clear that at low selection ratios using tests with modest validity coefficients can make a major change in selection efficiency.

Accuracy of selection is not the only thing to be considered. In order to calculate the economic benefit of a selection program, we also have to consider how much it costs to hire (and then replace) a person who is not qualified.

Let \( C \) be the cost to the selector of making a mistake, such as hiring
10 persons well qualified for a particular occupation. If industry has to fill 15 jobs, then 5 of the people hired are going to be “not well qualified,” even if screening is perfect (r = 1). This is shown by the r = 1 line in Figure 2.8.

The efficiency of selection can be greatly increased, regardless of the selection ratio or validity of the predictor, by increasing the fraction of qualified people in the population. Figure 2.9 makes this point dramatically. The figure can be looked at as a mathematical illustration of the arguments that governments make when they urge industry to relocate to their jurisdiction to take advantages of a well-educated workforce. We will return to this point in later chapters, when we consider the economic benefits of training and general education.

In many realistic situations people do not fall sharply into “qualified” and “not qualified” categories. Instead, we are concerned with the mean qualification of the individuals hired. Not surprisingly, this depends on how successful we are in identifying the best people among our applicants. Of course, this is what a predictor test is supposed to do.

The expected value of the performance of accepted candidates can be calculated if the predictor validity coefficient and selection ratio are known (Zeidner and Johnson, 1989, pp. 2–9; see also Cronbach and Gleser, 1965). Figure 2.10 shows the expected performance level of
Figure 2.10 Expected performance of selected applicants as a function of selection ratio and validity coefficient (r). If no testing were used, the expected performance would be 100.

Individuals selected using screening tests with validities of .3 or .5. The values shown in the figure should be compared to 100, the expected performance level using random selection (in arbitrary units). At low selection ratios marked improvement in performance can be obtained by using tests with validity coefficients in the .3 range.

Figure 2.10 describes a large number of realistic personnel selection situations. For instance, the figure can be thought of as the function that describes the relationship between the quality of work that would be done by college students who were either selected randomly or selected by psychometric tests of varying validity. Testing would be of most benefit for institutions that could afford to set low selection ratios, such as the prestigious private universities. Alternatively, Figure 2.10 can be thought of as the differences in productivity that could be expected from workers selected with or without a psychometric screening test. Testing would be of most benefit to employers who offered desirable jobs, as these employers could afford to set low selection ratios. Such employers could substantially increase productivity if they combined testing with inducements to apply. The testing would identify the better applicants, while the inducements would lower the selection ratio by increasing the number of applicants.
Table 2.1  Data Taken from the Simulated Data Shown in Figures 2.5–2.7

<table>
<thead>
<tr>
<th>Selection Ratio</th>
<th>Efficiency Ratio $E(r,s)$</th>
<th>Benefit $B(r,s)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.2</td>
<td>.50</td>
<td>.30</td>
</tr>
<tr>
<td>.5</td>
<td>.62</td>
<td>.12</td>
</tr>
<tr>
<td>.8</td>
<td>.84</td>
<td>.04</td>
</tr>
</tbody>
</table>

Note: Efficiency ratios and benefits were obtained from selection using the predictor test with a population validity coefficient of .42 and varying the selection ratio.

The figures and Table 2.1 are actually understatements of the normal benefits of selection, because they embody an overly rigid definition of "benefits." The selection case has been presented for the situation in which the selector tries to select exactly the best applicants: the top 10 for 10 positions, the top 20 for 20 positions, and so on. In most selection situations, the goal is to select qualified applicants, but not necessarily the best qualified. Also, in many cases the fraction of applicants who are qualified in the population is greater than the fraction who can be selected. An illustrative case was brought to my attention while I was writing this book: A major forest product company opened a sawmill that would employ about 100 people. Over 500 applied. It is likely that about half of them could have performed well had they been hired. The employer wanted to use tests to make sure that the majority of people hired would, in fact, be qualified. However, the employer did not feel a commitment to hiring exactly the best of the applicants.

This suggests an expansion of Equation 2.5, using an additional parameter. Let $q$ be the fraction of applicants who are qualified in the applicant population, and for simplicity, consider only the $c = s$ case. Then

$$E^*(q,r,s) = \text{probability that a selected person will be qualified.}$$

In the case of perfect selection,

$$E^*(q,1,s) = \begin{cases} 1 \text{ if } s \leq q, \\ q/s \text{ otherwise.} \end{cases}$$

This follows because the selector will first select the qualified, but will be forced to go to the unqualified if there are more positions to fill than there are qualified applicants. In the case of random selection,
Figure 3.1  The correlation between the appropriate ASVAB composite and different components of performance as a soldier. Source: McHenry et al. (1990), Table 7. The correlations shown are estimates of the population correlation.

many of the soldiers took the ASVAB two years or more before their performance was evaluated. It is clearly a useful personnel screening device for jobs whose performance is primarily determined by cognitive variables.

On the other hand, the ASVAB did not do nearly as well in predicting the personality and motivational aspects of job performance.

One final statistic from Project A should be noted. Recall that three broad classes of measures were available: knowledge assessed by a written test, hands-on performance observations, and ratings by supervisors and peers. The correlation between cognitive abilities and the written scores was .64; the comparable correlation between cognitive abilities and ratings was .16 (McHenry et al. 1990, Table 9). This is a striking example of a widely held conjecture: Cognitive tests will predict objective measures of on-the-job performance better than they will predict ratings of that performance.5

It has been frequently found that cognitive tests are more accurate in predicting performance in training and education than they are in predicting workplace performance after training. This fact has sometimes been used to argue that, given training, differences due to initial cognitive ability are unimportant. The Project A findings challenge this view somewhat, since there were substantial correlations between test
performance and objectively evaluated job performance. An overall summary of the total Department of Defense effort provided even more striking evidence that cognitive skills do continue to make a difference, even after years on the job.

Figure 3.2 shows a summary of the relation between tested performance, ASVAB scores (grouped into categories that are used in recruitment), and time on the job [Wigdor and Green, 1991]. Three points are noteworthy: (1) People with higher level scores do better on the job. Years of experience do not eliminate this difference. On the average, enlisted men and women in ASVAB categories I and II did better during their first year on the job than those in category IV with more than three-years' experience. (2) On-the-job learning does take place; tested performance improves with experience within every ASVAB category. (3) The cognitive skills and experience variables interact. The greatest improvement over time was shown by category IV personnel during the first year on the job. Category I and II personnel reached their stable performance levels more quickly.

How generalizable are the results of these studies to the civilian workforce? There are two obvious drawbacks: They were restricted to military jobs and they examined a workforce that is, on the average, substantially younger than the general civilian workforce.

The first drawback is probably not a major one. Of the more than 4,000 individuals studied in project A, slightly fewer than 2,000 were
Figure 3.3  An illustration of a matrix problem. Which of the boxes below the line should be used to fill the box in the lower right, above the line?

standard of tests to identify fluid intelligence, since it evaluates the examinee's ability to deal with what are almost certain to be novel problems.

There have been several analyses of the cognitive behaviors required to solve Raven matrix problems (Carpenter, Just, and Shell, 1990; Hunt, 1974). The gist of these analyses is that in order to solve matrix problems the examinee must be able to spot progressive changes in the patterns both across the rows and down the columns of the test figures and must be able to infer how these changes combine to produce the figure in the lower right. Thus, the test seems to evaluate the ability to notice regularities underlying change and the ability to combine different trends. Both tasks place a formidable information-processing burden on short-term memory.

In the United States Raven tests are primarily research instruments. In Europe they have been used for personnel evaluation, especially in the military. From 1950 through the 1980s several European countries routinely gave Raven tests to all military inductees. As a result we have a very good idea of the distribution of scores in different populations. Some of this information is relevant to projections of future workplace
ministered by the Educational Testing Service (ETS), a nonprofit corporation that specializes in test development and administration. The SAT is administered under the aegis of the College Board, a consortium of educational institutions, for whom ETS acts as a contractor. While it has competitors, the SAT clearly dominates its market.

The SAT is a paper-and-pencil, multiple-choice test that requires about two and a half hours to administer. (Actual times have fluctuated over the years, as the test forms have been changed slightly.) It is broken down into a number of subtests dealing with specific types of verbal comprehension or mathematical facility. Sample questions from different subtests are shown in Table 3.1.

Scores on the verbal and mathematical tests are summed to produce two composite scores: one for verbal aptitude [SAT-V] and one for mathematical aptitude [SAT-M]. These scores are used on the quite reasonable assumption that different types of colleges [e.g., a liberal arts school and an engineering institute] place different emphases on language skills and mathematical skills. When a prediction is made based on "the SAT score," this usually refers to a prediction made using the combined SAT-V and SAT-M scores.

Although the SAT is based on less than three hours' behavior, it is a surprisingly accurate evaluation of educational potential. The estimated population correlation between SAT scores and first-year grade

<table>
<thead>
<tr>
<th>Table 3.1 Sample SAT Subtest Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Verbal</strong></td>
</tr>
<tr>
<td><strong>Antonyms:</strong> Given a word choose the opposite in meaning.</td>
</tr>
<tr>
<td>Example: Good</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>(a) sour</td>
</tr>
<tr>
<td>(b) bad</td>
</tr>
<tr>
<td>(c) red</td>
</tr>
<tr>
<td>(d) hot</td>
</tr>
<tr>
<td>(e) ugly</td>
</tr>
</tbody>
</table>

| **Analogy:** Select the lettered pair that represents a relationship similar to the relationship between the word pair in capitals. |
| Example: YAWN: BOREDOM                  |
|                                        |
| (a) dream:sleep                        |
| (b) anger:madness                      |
| (c) smile:amusement                    |
| (d) face:expression                    |
| (e) impatience:rebellion               |
Table 3.1  (Continued)

Sentence completion: Choose the word or set of words that best fills in the blanks in the sentence.
Example: Although its publicity has been _____, the film itself is intelligent, well acted, handsomely produced, and altogether _____.
   [a] tasteless . . . respectable
   [b] extensive . . . moderate
   [c] sophisticated . . . amateur
   [d] risqué . . . crude
   [e] perfect . . . spectacular

Reading comprehension: Examinees read a paragraph about a contemporary topic (e.g., use of National Parks) and answer multiple-choice questions about its meaning.

Mathematical

Regular mathematics: Simple algebraic and geometric problems are presented in multiple-choice form.
Example: If \( 2y = 3 \), then \( 3|2y|^2 = \)
   [a] 27/4
   [b] 18
   [c] 81/4
   [d] 27
   [e] 81

Data sufficiency: A mathematical relationship is presented, followed by two assertions about variables in the relationship. The examinee determines whether or not the relationship can be shown to be true on the basis of one, either, both, or neither of the assertions.
Example: Is \( a + b = a^2 \) ? Assertions: \(|1| b = 0 \)
   \(|2| a = 10 \).

Quantitative comparisons: Two quantities, A and B, are defined. The examinee indicates whether A is greater than B, B is greater than A, the two are equal, or a relation cannot be determined.
Example: A: The least positive integer divisible by 2, 3, and 4.
   B: 24.

Source: Descriptions and examples paraphrased from illustrations given in Marcé et al. (1990).
Table 3.2  Armed Services Vocational Aptitude Battery
Subtests

<table>
<thead>
<tr>
<th>General science</th>
<th>Paragraph comprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arithmetic reasoning</td>
<td>Numerical operations</td>
</tr>
<tr>
<td>Coding speed</td>
<td>Word knowledge</td>
</tr>
<tr>
<td>Auto shop information</td>
<td>Mathematical knowledge</td>
</tr>
<tr>
<td>Electronics information</td>
<td></td>
</tr>
</tbody>
</table>

evaluate knowledge about mechanics, automobiles, and electronics, which is required in a number of military positions. ASVAB questions are generally easier than comparable questions on the SAT because the SAT is intended to be maximally sensitive at identifying ability in, generally speaking, the “upper half” of the graduating high school population. The ASVAB is intended to evaluate abilities in the “middle half,” that is, the 25th to 75th quartiles of ability, in that population.

In addition to being a screening tool, the ASVAB is used in personnel assignment. Every armed services recruit begins his or her career in basic training (“boot camp” in the navy and marines), where the enlistee becomes generally familiar with the rules and customs of military life. Following basic training, an enlistee is assigned to an entry-level military occupational specialty (MOS). These specialties cover a wide range of occupations. During the 1980s the army alone had 276 entry-level MOSs. Some MOSs, such as infantryman, are specific to the military. Others, such as administrative specialist (clerk), are analogous to civilian occupations.

The subtests of the ASVAB are used to compute several composite scores, somewhat like the SAT-V and SAT-M scores. Each composite is based on a different weighted sum of subtest scores. The Armed Forces Qualification Test (AFQT), which stresses verbal skills and mathematical reasoning but does not count the different specialized knowledge tests or coding speed, is a summary composite used to decide whether or not to enlist a candidate. Other composites are used to predict performance in selected groups of MOSs. For instance, the electronics composite is the sum of scores on the arithmetic reasoning, mechanical knowledge, electrical information, and general science subtests (Johnson, Zeidner, and Schalarios 1990).

During the 1980s the armed services conducted an extensive study aimed at both evaluating the ASVAB and determining whether or not other predictors of the performance of enlisted personnel could be found (Wigdor & Green, 1991). The study was probably the largest industrial assessment study ever done. Its results are relevant to our interests.
Table 3.3  Military Occupational Specialties Selected for Intensive Study
in U.S. Army Project A

Specifically military specialties
Infantryman
Cannon crew member
Armor (tank) crew member

Specialties with civilian analogs
Single channel radio operator
Motor transport operator
Medical specialist (technician)
Vehicle mechanic
Administrative specialist
Military policeman

Source: Campbell (1990), Table 1.
Note: Classification into military and military/civilian specialties is my own.

While all four armed services participated, the largest single study
was the army's Project A (Campbell, 1990), which was continued into
the 1990s. It involved testing and evaluation of the performance of over
10,000 soldiers and will be described in some detail. We will be con-
cerned here with results bearing directly on the validity of the ASVAB.

Nine MOSs selected for intensive study are listed in Table 3.3. The
sample in this portion of Project A consisted of slightly more than 4,000
soldiers, with from 300 to 600 in each MOS. All the soldiers had taken
the ASVAB when they were enlisted. (As part of the study, the soldiers
took a number of further tests that will not concern us, as they are not
typical of widely used industrial personnel tests.)

Four different techniques were used to evaluate performance (Camp-
bell, McHenry, and Wise, 1990). Written personnel records were exam-
iąned. Supervisors and peers were asked to rate the soldiers' performance
on a variety of tasks associated with either their technical MOS—that
is, their job narrowly defined—or more generally their performance as
soldiers. This distinction would apply to virtually any job. Conceptu-
ally, at least, one can distinguish between an individual's technical
competence and his or her performance as a member of a working
group. Soldiers were also given written tests to determine their knowl-
edge of their specialty, as defined by army manuals. Finally, and
perhaps most interestingly, soldiers were asked to perform certain
hands-on tasks that had been selected as important for their MOS.
Their performance was evaluated by trained raters from the Project A
research group. For instance, a vehicle operator was evaluated for the
Table 3.4  Job Factors Identified in Project A

<table>
<thead>
<tr>
<th>Core technical skill:</th>
<th>Job performance as defined by the MOS.</th>
</tr>
</thead>
<tbody>
<tr>
<td>General soldiering:</td>
<td>Ability to perform tasks generally required of soldiers.</td>
</tr>
<tr>
<td>Effort and leadership:</td>
<td>Extent to which soldiers attempted to improve themselves and the extent to which they assumed leadership roles.</td>
</tr>
<tr>
<td>Personal discipline:</td>
<td>Interpersonal reliability in social and organizational settings.</td>
</tr>
<tr>
<td>Fitness and bearing:</td>
<td>Extent to which soldiers were physically fit and conformed to the army's expectation of a good soldier's appearance.</td>
</tr>
</tbody>
</table>

Note: The factors are further defined by Campbell, McHenry, and Wise (1990).

ability to drive without riding the clutch. Thus, Project A, and the other service studies, went well beyond the usual practice of relying on supervisor ratings and personnel records for criterion measurement.

The standard psychometric model (see Chapter 2), in which abilities are described in terms of underlying factors, was applied to job performance (Campbell, McHenry, and Wise, 1990). Five factors underlying job performance were evaluated. The factors, together with my own gloss of the authors' longer description, are shown in Table 3.4. Except for "fitness and bearing," which seems to be unique to the military, the job factors apply to both military and civilian jobs.

The five factors fall into three separate groups: the two technical proficiency factors, the personal discipline and fitness and bearing factors, and, between these two groups, the effort and leadership factor. The best indicators of the technical knowledge factors were the hands-on tests and the written tests of job knowledge. The best indicators of the personality and appearance aspects of job performance were leader and peer ratings.

The Project A findings show, quite clearly, that a person's performance on the job depends on technical skill and "personality" variables, including attitude and motivation. Are either (or both) of these dimensions of job performance predictable from psychometric measures of mental competence?

Figure 3.1 shows the ASVAB validity coefficients for each of the five job factors. Technical aspects of job performance were predicted quite well. The population validity correlation estimates between score and skill are in excess of .60, which is impressive, when one considers that
Table 3.5  General Aptitude Test Battery Subtests

<table>
<thead>
<tr>
<th>Paper-and-Pencil Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>{1} Name comparison. Compare names to see if they are identical.</td>
</tr>
<tr>
<td>{2} Computation: Simple arithmetic. The test is timed.</td>
</tr>
<tr>
<td>{3} Three-dimensional space: The examinee is shown a diagram of a flat “piece of metal” to be folded along marked lines and must decide what the object will look like when folded.</td>
</tr>
<tr>
<td>{4} Vocabulary: Identify words with the same or opposite meanings.</td>
</tr>
<tr>
<td>{5} Tool matching: The examinee is shown a “probe” picture of a common tool and four comparison pictures and must decide which of the four comparisons exactly matches the probe.</td>
</tr>
<tr>
<td>{6} Arithmetic reasoning; Simple arithmetic word problems.</td>
</tr>
<tr>
<td>{7} Form matching: The examinee must pick out two identical figures in a group of abstract geometric figures.</td>
</tr>
<tr>
<td>{8} Mark making: The examinee must make three pencil marks in a small box. The score is the number of boxes marked in a minute.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Manipulation-of-Objects Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>{9} Place: Move pegs from one place to another in a peg board. The score is the number of pegs moved in 15 seconds. Three attempts are allowed.</td>
</tr>
<tr>
<td>{10} Turn: Turn over as many of the pegs as possible in 30 seconds.</td>
</tr>
<tr>
<td>{11} Assemble: Assemble rivet and washer units for 90 seconds.</td>
</tr>
<tr>
<td>{12} Disassemble: The examinee disassembles as many of the units assembled in test {11} as he/she can within 60 seconds.</td>
</tr>
</tbody>
</table>

both the GATB and various scoring methods for it. The NAS report contains an excellent review of the GATB studies and in addition makes a number of points about the use of tests in general [Hartigan & Wigdor, 1989].

Although the NAS report generally endorsed the GATB, further attacks were made on the test, both by minority group advocates who argued for within-group scoring and by others who argued that within-group scoring amounted to an illegal quota system. In 1990 the DOL ceased using the test for employment referrals. This was essentially a political decision.

Even though the sponsor has ceased using the GATB, the large body of research done on the test is relevant for our purposes. A discussion of this research both furthers our understanding of test use in industry and illustrates some of the problems that arise in evaluating tests that are not targeted to a particular, easily identified group of individuals, such as college freshmen or military personnel.

The GATB consists of the 12 subtests shown in Table 3.5. The prob-
### Table 3.6 Wechsler Adult Intelligence Scale-Revised Verbal and Performance Subtests

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal</td>
<td></td>
</tr>
<tr>
<td>Information:</td>
<td>Test of general knowledge, such as names of presidents.</td>
</tr>
<tr>
<td>Comprehension:</td>
<td>Answer simple questions about daily life and about hypothetical problems.</td>
</tr>
<tr>
<td>Arithmetic:</td>
<td>Simple arithmetical word problems.</td>
</tr>
<tr>
<td>Similarities:</td>
<td>Identify similarity between concepts, e.g., between &quot;horse&quot; and &quot;cow.&quot;</td>
</tr>
<tr>
<td>Digit span:</td>
<td>Examinee must repeat a sequence of spoken digits.</td>
</tr>
<tr>
<td>Vocabulary:</td>
<td>Define words.</td>
</tr>
<tr>
<td>Performance</td>
<td></td>
</tr>
<tr>
<td>Digit-symbol test:</td>
<td>The examinee must copy symbols that are arbitrarily associated with numbers.</td>
</tr>
<tr>
<td>Picture-completion:</td>
<td>The examinee must complete a picture with a missing detail, such as a horse without a tail.</td>
</tr>
<tr>
<td>Block design:</td>
<td>The examinee must arrange colored, odd-shaped blocks into a specified pattern.</td>
</tr>
<tr>
<td>Picture arrangement:</td>
<td>The examinee is given a set of &quot;mixed up&quot; pictures, rather like a comic strip that has been cut into frames, and must rearrange the frames into a coherent story.</td>
</tr>
<tr>
<td>Object assembly:</td>
<td>The examinee must put together objects from their constituent parts.</td>
</tr>
</tbody>
</table>

Problem-solving methods in an unfamiliar situation—and crystallized intelligence—the ability to utilize acquired knowledge in an appropriate way. The Raven Progressive Matrix Tests (Raven and Court, 1960) are intended to evaluate fluid intelligence. The Raven test is actually a sequence of tests of progressively greater difficulty. All the tests, however, use the same format, which has come to be called a matrix question. Figure 3.3 displays a fairly easy, made-up example to illustrate the principle. Actual items vary widely in difficulty. Many psychometricians regard the Raven Progressive Matrix Test as a gold
Figure 4.1  Changes over time in the fraction of the workforce over 35. Source: U.S. Bureau of the Census (1993).

One of the most frequently repeated findings in the literature is that age affects verbal and nonverbal intelligence in different ways. A typical finding is shown in Figure 4.3, which displays the mean scores achieved on the WAIS-R by different age groups (Wechsler, 1981). There are only small differences between 25–30-year-olds and 50–60-year-olds in verbal intelligence scores, but there are substantial differences in performance scores.

A number of psychologists prefer to distinguish between crystallized (Gc) and fluid (Gf) intelligence rather than verbal and nonverbal intelligence. (See Chapter 2.) They interpret the verbal-nonverbal distinction shown in Figure 4.3 as basically a Gc-Gf distinction (Horn, 1985).

These views can be amalgamated by distinguishing three different types of intelligence. The first type can be called either crystallized (Gc) or verbal intelligence. The reason for not distinguishing between the two is that, in practice, most tests that are used to predict success in the workplace require skills that fit the definition of either crystallized or verbal intelligence. The second type, somewhat related to the first, is an ability to deal with abstract reasoning problems. This is probably the closest to what we mean by fluid intelligence. I shall use the terms “reasoning” and “fluid intelligence” more or less interchangeably. The third type is the ability to deal with visual-spatial reasoning. In the workplace visual-spatial reasoning seems to be impor-
time. Their conclusions were carefully worded, and it does some injustice to the original authors to present a summary. Nevertheless, I shall try. The most accurate “one liner” describing their conclusions is that the black American population mean on general intelligence tests was, as of 1970, about one standard deviation below the white American mean. Jensen (1980) reached a similar conclusion. Subsequently, Humphreys (1988) reviewed a number of results on scholastic achievement tests, largely conducted in the 1970s and early 1980s. He agreed with Loehlin, Lindzey, and Spuhler and with Jensen, and noted that the Hispanic American mean score appears to be about .8 standard deviations below the white mean.

Figure 4.4 provides a concrete example. The figure shows SAT scores from 1978 to 1992. Scores are shown separately for whites, blacks, and Mexican Americans. The figure also shows the score 200 points below the white mean, which is the score that would be achieved by a group whose mean score on each subtest was approximately one standard deviation below the white group’s score. The black score approximates the white 200 score, while the Mexican American score lies somewhat above it. The group differences are remarkably consistent, even though during this period considerable efforts were expended to improve educational opportunities for minority students.

Differences in SAT scores do not automatically generalize to differences in cognitive skills throughout the workforce. People who take
Figure 4.5 Percentage of white and black young adults reaching various levels of competence on tests of workplace-oriented document reading tasks. Also shown is the score predicted for the black group assuming that the mean black score is one standard deviation below the mean white score. Source: National Center for Educational Statistics (1993).

sorts of documents one has to deal with in the normal workday. The National Academy panel suggested the following nonstatistical points of reference (Jaynes and Williams, 1989, pp. 353–354):

Documents test:
200 = ability to select and match store coupons.
300 = ability to follow directions, using a map.

Quantitative test:
200 = ability to add simple checkbook entries.
300 = ability to enter and add accounting items.

The National Academy panel summarized the results by saying that the same percentages of blacks and whites could accomplish fairly simple reading and quantitative analysis tasks, but that whites showed a greater advantage as the tasks became harder.

It is possible to go a good bit beyond the National Academy’s statement. Figure 4.5 shows the percentage of white and black young adults who achieved various levels of competence on a document reading test. Figure 4.6 is a similar figure for the evaluation of quantitative skills. Also shown is the percentage of blacks who would be predicted to
Figure 4.6 Percentage of white and black young adults reaching various levels of competence on tests of workplace-oriented quantitative tasks. Also shown is the score predicted for the black group assuming that the mean black score is one standard deviation below the mean white score. Source: National Center for Educational Statistics (1993).

achieve each level of competence, on the assumption that (a) both black and white scores are distributed normally, with the same standard deviation, and (b) that the mean of the black population is one standard deviation unit below the white mean. (In the terms of the example computations, if the NAEP test had been scored as the WAIS is, the white scores would serve as the referent group and the black scores as the –15 group.) As can be seen, the match between data and prediction is quite close.

A similar picture applies to assessments of explicitly taught school topics. Knowledge of science is particularly important, because most observers believe that the possession of elementary knowledge about science will be a central skill in the coming workplace. Figure 4.7 shows how 17-year-olds fared in the 1989–1990 school year. The relevant levels are:

1. Understand simple scientific principles.
2. Apply basic scientific information to problems.
3. Analyze scientific procedures and data.
4. Integrate specialized scientific information in comprehension.
These scientific skills are quite different from document reading and elementary quantitative skills, but the relative picture across ethnic groups is very much the same. The distribution of black and Hispanic scores is approximately what would be expected if the skill level of the minority populations was about one standard deviation unit below the skill level of the white population.

These studies, and many others that there is not time to review [cf. Humphreys, 1988], strongly indicate that in most tests of cognitive competencies the black population mean is about one standard deviation level below the white population mean, and that the Hispanic population mean is slightly above the black mean. These intergroup differences seem to have been constant since 1975. There are some older data, which I have not cited, indicating that there were even greater differences in educational and cognitive achievement before 1970. The push for minority education has not been a failure, but progress toward intergroup equality seems to have stopped sometime in the 1970s. This conclusion is consistent with, although somewhat more detailed than, the conclusion of the National Academy panel on the status of black Americans (Jaynes and Williams, 1989). Furthermore, the result is definitely not confined to schoolchildren and young adults. In 1992, the Educational Testing Service, at the request of the Department of
Figure 4.8  The mean effectiveness within job classes, as a function of the decile level of the job class and the mean difference between population subgroups. The figure assumes that .8 of the population fall into the more effective subgroup (A) and .2 fall into the less effective subgroup (B).

Comments are made that the entire workforce is being dragged down by an "incapable" group of minority workers. Indeed, T. Nakasone, the prime minister of Japan in the late 1980s, caused something of a furor by a statement very much like this. (He later denied that he said it, but the press had a transcript of his remarks.) Such statements are simply not true.

Figure 4.8 shows the relevant data. This figure plots the mean effectiveness of the population, under the conditions described previously. Look at how close the lines are to each other. For the topmost job classes, effectiveness is virtually constant, regardless of the value of $d$, that is, regardless of the size of the between-groups differences. Intergroup differences do affect performance levels in the less desirable jobs, but they are sizable only for $d = 1$, the highest estimate we consider for the size of intergroup differences. Even if this estimate is correct, the effect may be of little economic importance, since there is relatively little difference in value between high and low performance in the less desirable job classes.

Virtually all discussions of the economic effects of intergroup differ-
ences in talent focus on mean differences. Differences in the within-group variation in ability can be quite important. Figure 4.9 shows mean effectiveness levels for different job groups under the assumptions that \( d = .5 \), with the ratio (SR) of group B:group A standard deviations varying from 1.0 (equality) to 2. For comparison, the case of a uniform population \( (d = 0, \text{ ratio of standard deviations} = 1) \) is also shown. Clearly, differences do appear as within-group variation increases. In fact, if we assume extreme variation \( (\text{SR} = 2) \), performance in the best jobs [9th decile] actually exceeds the effectiveness observed in the equality condition. On the other hand, as variation increases, the effectiveness level in the lowest job classes decreases, relative to the equality condition.

Arguments about the effects of within-group variation are more than mathematical/statistical niceties. The term “minority” lumps together a number of disparate groups. Blacks do not show the same test score distributions as Hispanics, and various Hispanic groups differ among themselves. Since the 1960s a substantial number of blacks have moved into the economic middle class, while, as is well known, the economic and social conditions of poorer blacks have not improved.
Figure 4.10  The effect of an EIO policy. Ratio of group A to group B members in different job classes as a function of the difference between group A and B members in mean effectiveness ($d$) and relative within group variation in effectiveness (SR). The case of equality ($d = 0$, SR = 1) is shown for comparison.

All these effects would produce greater within-group variation in group B, and, as Figure 4.9 shows, these effects are not small.

The reason for the effects shown in these two figures can be understood without a lengthy excursion into statistics. Under an EIO policy, reduced effectiveness in the smaller group does not influence the more desirable job classes [Figure 4.8] because very few group B members are assigned to those jobs. The reason that increases in within-group B variation influence effectiveness in the extreme job classes [Figure 4.9] is that as the variation increases some very talented B group members begin to occupy the top job classes, thus increasing effectiveness in those jobs, while some quite untalented group B members must be assigned to the least desirable jobs.

The upshot of these analyses is that under a pure EIO policy the effects of intergroup differences in mean performance are, on the average, quite small. Perhaps more important, these effects are negligible for the most desirable jobs. Under an EIO policy, effectiveness is not the problem. Perceived social justice is.

The EIO policy maintains effectiveness within job classes by permitting large discrepancies in the representation of group A and B members across classes. Group B members virtually disappear from the top job classes and are heavily overrepresented at the bottom. This is shown in Figure 4.10, which plots the ratio of group A:group B members for
the two subpopulations in the overall population. If the employer uses the referrals, the EGO policy is complete.

The EGO policy does not necessarily force job qualifications upward or downward over the entire range of the workplace. In the top job classes group B members are offered positions over more qualified group A members, which reduces top-level effectiveness. On the other hand, effectiveness in the lower job classes should increase, as better qualified group A members are forced into them. Conceivably the two effects could balance each other out.

In practice such fortunate balancing is unlikely, because the relative worth of a highly skilled person compared with a barely qualified person is usually greater in the top job classes than in the bottom classes. The difference between an excellent and an adequate surgeon is greater than the difference between an excellent janitor and an adequate one.

In spite of our ability to generate such apocryphal examples, under the conditions of the total U.S. workforce, the differences in effectiveness under EGO and EIO policies are not large. Figure 4.11 displays the mean effectiveness within each job class under each policy. EIO performance is slightly higher in the top job classes, while EGO performance is slightly higher in the lower job classes.

A problem with this analysis is that it depends upon a rather esoteric
Figure 4.12  The fraction of the workforce that would be qualified for each job class under an EIO policy but would be assigned to a lower job class under an EGO policy.

measure, effectiveness measured relative to the variation in effectiveness in group A. Although this measure can be justified on statistical grounds, it does not have an intuitive interpretation. An alternative approach, which is more intuitive, is to calculate the extent to which implementation of an EGO policy forces people in the more skilled group to accept jobs below the level they would have obtained under an EIO system. One can argue that these individuals are being underutilized, since they are not allowed to contribute commensurate with their talent level. Thus, this analysis stresses suboptimal utilization of whatever cognitive talent is present in the workforce.

Figure 4.12 shows the extent of suboptimal utilization under an EGO policy. Two different cases are shown: the extremely pessimistic situation \(d = 1, \ SR = 1\) case and the more realistic \(d = .5, \ SR = 1.25\) situation. The total fraction of the workforce suboptimally utilized is .41 in the pessimistic scenario and .19 in the realistic one. These two statistics should be disturbing to anyone who is concerned with workforce efficiency. Quite aside from any moral or legal considerations, underutilization of 20–40 percent of the workforce is a serious matter.

If we were to look only at effectiveness levels (Figure 4.11), we might be tempted to say that the difference between an EIO and an EGO policy is trivial. If we look at suboptimal utilization (Figure 4.12), it appears that the difference is substantial. The reason for this apparent
Figure 4.13 "Jealousy" under an EGO policy as a function of job classes. The jealousy quotient for a job class is the fraction of the group A workforce who have higher effectiveness scores than a group B member assigned to that job class, but who are themselves assigned to a lower job class.

overrepresented in the more desirable job classes. Similarly, a group that is less talented, on the average, will be disproportionately represented in the less desirable job classes. Furthermore, the disparity in average ability does not have to be very great to produce a marked disparity in representation at the extremes.

There is a somewhat surprising codicil to this conclusion. If there is greater variance in the less talented group than in the more talented group, then under some circumstances the less talented group may actually have higher representation in the very best job classes than it does in the upper middle range of job classes.

The disparities in job assignment within an EIO model could easily lead to the perception that assignment has been made on the basis of group membership. Consider how the situation might look to a majority group member who is a committed EIO advocate and is in the upper job class—that is, one of the "movers and shakers" in our society. This person would look around his or her own job class and see a reasonable number of minority group members. Therefore, on the basis of personal experience, the society leader would be justified in concluding that there is no group difference in cognitive skills. The society leader then discovers that minority group members are hardly ever found in the
**Figure 4.14** The effort (interviews per successful candidate) required to select a person with requisite Gf and Gc skills in various job classes: 2000 relative to 1985.

---

**PROJECTIONS**

Workforce skills associated with verbal intelligence and crystallized problem solving will increase during the next decade. The increase is almost entirely due to the aging of the workforce. Any absolute defects in the starting point, that is, the skill levels as they existed in 1985, will be retained. In particular, if we regard our workforce as competing against European and Japanese workforces, we must realize that those workforces are also aging and will gain the same benefits in crystallized intelligence as our own workforce.

A competitive advantage for our workforce could be obtained only if we either embarked on a retraining program for those in the workforce or if we markedly improved our school system and, through it, the qualifications of the entering workforce. The problems with the schools have already been discussed. Historically, the United States has been much less receptive to worker training than have the European and industrialized Asian countries. The analysis offered here does not provide room for complacency.

One fact emerges again and again. Any cognitive problems that are projected for the overall workforce are compounded at the top. The worst projections for Workforce 2000 are for those jobs that require a high degree of competence and emphasize problem solving rather than the application of previously learned solution methods. The issue of how we might teach problem solving in the schools and how much it might cost will be addressed in Chapter 7.
Figure 4.A1  The distribution of levels of achievement on a test in a population consisting of two groups, A and B. A is assumed to contain 80% of the total population. Group A has a mean test score of 0 and a standard deviation of 1, Group B has a mean test score of -1 and a standard deviation of 1. Above a score of 1.67, the population consists almost entirely of group A members. Below scores of -1.8 group B members predominate.

the distribution of scores for groups A, B, and the entire population. The latter figure is obtained by adding the group A and group B population densities at each level of performance. The positions and relative sizes of the two distributions follow from the assumptions. The group A distribution is larger than the group B distribution because group A is larger than group B. In Figure 4.A1, I assumed $q = .8$, which means that the area under the curve for group A must be four times as large as the area under the curve of group B. The group A distribution is centered on zero, and the group B distribution is centered on -1, reflecting the prior assumptions.

To find the fraction of group A and group B members that fall above or below some arbitrary score level, one calculates the relative area of each curve that lies beyond [to the right] of that score level. (These values can be obtained by consulting a table of density functions for the standard normal curve.) In Figure 4.A1, 5 percent of the area under group A’s curve falls beyond [to the right] of +1.67. On the other hand, only four-tenths of 1 percent of the area under group B’s curve falls to the right of 1.67.

To calculate the fraction of the population \(A + B\) with scores above an arbitrary value, \(Z\), it is necessary to consider the fractions of groups
Figure 5.1 The relation between psychometric scores, performance evaluations, and productivity. The worker's ability and knowledge produce psychometric measurements. Performance evaluations and productivity are produced by the combination of the worker's ability and knowledge and the demands of the workplace. Correlational studies, indicated by the double-headed arrow, indicate that the two are related but do not explain why.

Information processing machinery must a person have to solve a problem; what program, or, in more human terms, problem-solving strategy, does the person apply to use the machinery; and what explicit knowledge about the external world does the machinery operate upon?

The machinery question links us to our biological system. The efficiency of functioning of short-term and long-term memory, perception, and motor control are ultimately determined by anatomy and physiology. The knowledge question links us to our social system; we acquire only the information that the culture makes available to us. The programming question falls in between. Problem-solving strategies are limited by our information-processing machinery, but we can develop strategies that use that machinery in an efficient or inefficient way.

Figure 5.1 summarizes the argument. Psychometric tests present people with certain demands; how they use their knowledge and information-processing capacities determines what the resulting score is. Very much the same thing can be said of the workplace. People respond to workplace demands to produce performance that can be evaluated and that results in productivity. Psychometric research can establish a correlation between test scores and performance measures, but it cannot tell us why the correlation exists.

Cognitive psychology seeks to go beyond this, by determining just
Figure 5.2  Schematic of a choice reaction time experiment. The observer places his or her hand on the “home” button. When a number appears on the screen, the observer presses the corresponding button as quickly as possible.

The examinee is supposed to move the finger from the home button to the button with the same number as that on the screen. The measurement of interest is the time between the instant at which the symbol appears on the screen and the time the person’s finger moves off the home button toward the appropriate target button. This measure is called choice reaction time (CRT).

Mail sorting in the post office is a good example of an industrial counterpart to the CRT paradigm: Letters are moved past the sorter’s station on a conveyor belt, and the sorter must read the postal code and punch it into a keypad as the mail goes by. The sorter’s interpretation of the written code is used to route each piece of mail to the appropriate branch point off the conveyor belt and, hence, on to its destination. The sorter must be able to make a decision to punch one key in response to any of the symbols

9 9 9 9

and another key in response to any of the symbols

8 8 8 8.

In theory such decisions could be made by an optical scanner attached to a computer. In practice the U.S. postal service does not expect to
Figure 5.3  A considerably simplified schematic of an air traffic control display.

fast players and then have them go through conditioning workouts, we could try to develop selection and training methods to maximize jobholders' information-processing capacities. Alternatively, we could design the workplace to match the information-processing capacities that the workers have. To continue the athletic analogy, one of the reasons that athletes run faster today than they did in 1920 is because today's athletes have better shoes!

The Air Traffic Control Example

Air traffic controllers use radar displays to manage aircraft movement. Figure 5.3 is a simplified schematic of the display used for sector control, when aircraft are near major airports but are not actually landing. Each marker on the screen indicates an aircraft's location in two-dimensional space. Altitude, speed, and identification numbers are indicated in text beside the marker. In addition, the controller knows the schedule of aircraft that are expected to arrive but have not yet made radar contact. In order to manage the traffic flow, the controller has to work back and forth between verbal and visual information. The controller also has to have memorized information about aircraft capabilities, weather, and control procedures.
mentary physics, recall the necessary equations, and then use a knowledge of mathematics to solve them.

Let us contrast the air traffic control and pulley examples. The visual display facing the air traffic controller is much more complex than the one facing the mechanic. So is the amount of information required. The air traffic controller must be able to recall hundreds of facts. The mechanic who uses a pulley system need only know one principle—that the force on the rope must be the same throughout. However, the pulley problem is by no means trivial. In order to solve it, the problem solver must construct a mental model of the pulley situation and "run" it inside his or her head. Figure 5.4 shows a simple problem where this is easy to do. It is much harder to visualize what is happening when there are five or six pulleys in the system.

How complex a computing burden can an individual tolerate? Some problems so tax people's resources that they attempt shortcuts, such as counting the number of pulleys in the system and assuming that each pulley reduces the force required by a factor of two. This simplification ignores the fact that some pulleys simply change the direction
choosing an appropriate response (at the bottom of the figure) when presented with a problem-solving situation (at the top of the figure). In order to do this, the problem solver must develop an internal representation of the problem in working memory (the middle section of the box). Colloquially, the internal representation depicts how the problem solver sees the problem. What is the person’s interpretation of what is going on?

Interpretation depends on an interaction between what is “out there” in the world and what we think is relevant “in here” in our long-term memory. This is shown by the middle and right-hand boxes of the figure. Think of the way a physician diagnoses a patient, by combining observations of patient’s symptoms with knowledge about similar patients and diseases.

Building an internal representation is only a first step. The purpose of problem solving is to build an interpretation that can be recognized as a guide to action. This is depicted in the left-hand box of Figure 5.5, which shows that the internal representation triggers problem-solving rules (procedural knowledge in the argot of cognitive psychology) telling the problem solver what to do when he or she creates a particular interpretation of a situation. Some simple rules for driving in traffic were shown above. Procedural rules for medical and legal problem solving are far more complicated than those for driving, but the principles of execution are the same.
more specialized than the axioms of geometry, followed by a bit of arithmetic. Mistakes are certainly possible.

A third way to solve this problem is simply to remember the Pythagorean theorem and apply it. The task of constructing the internal representation is much simplified but the method is less general than the trigonometric or geometric method, because it works only with right triangles.

Or you could simply whip out a tape measure and measure the distance from X to Y. If the boards are too long to do this physically, draw a scale diagram and measure line C. This method provides only an approximation to the answer, because you will be limited by the accuracy of your measurements. Therefore, it works only in situations in which extreme accuracy is not required. Fortunately, though, most of the world seems to be that way. Hands-on methods of thinking can often reduce the strain on the brain.

Clearly, each method has its advantages and disadvantages. In almost every situation problem solving involves a trade-off between using general solutions that place heavy demands on working memory and using limited, remembered solutions that shift the burden from manipulations in working memory to retrievals from long-term mem-
Figure 5.7 The sentence verification paradigm. An observer sees screen 1, and then screen 2. The task is to indicate whether the sentence on screen 1 is a true or false description of the picture on screen 2. Since the two screens are not seen at the same time, the comparison must be made from memory.

Figure 5.7 shows a schematic representation of the sentence verification paradigm. Screen 1 presents a sentence, "The Star Is Above the Plus," and Screen 2 displays a picture with a plus sign and a star. The observer must decide whether the sentence accurately describes the picture.

The sentence verification paradigm is a cognitive psychology task that tests the ability to process and compare visual information. It involves comparing a sentence with a visual image presented at different times, requiring short-term memory and visual-spatial processing abilities.

Regarding the performance of personnel psychologists, the sentence verification paradigm is a useful tool for assessing cognitive abilities. It can be included in tests like the ASVAB or GATB to evaluate an individual's ability to process and remember information quickly.

Psychometricians represent the performance of person $i$ on task $j$, $X_{ij}$, as

$$X_{ij} = \sum_k w_{jk} f_{ik} + e_{ij},$$

where $k$ is the $k$th of $1..K$ factors (basic abilities) underlying performance, $w_{jk}$ is the importance of the $k$th factor to performance on task $j$, $f_{ik}$ is the level of person $i$'s ability on factor $k$, and $e_{ij}$ is a residual.
outside world and memory itself. Long-term memory contains rules governing how messages will be written on the blackboard.

The rules are stated in the production system notation, as pattern \( \rightarrow \) action rules. These rules should be thought of as active agents who continually examine the contents of the blackboard rather than as passive lists of directions. Each production examines the contents of the blackboard independently of the actions of any other production. When a production matches some part of the blackboard, the action associated with that production is taken. (This is referred to as the production’s firing.) If more than one production can be matched to the contents of working memory, a conflict resolution rule is invoked to determine which production is fired. Different versions of the model use different conflict resolution rules. But we need not be concerned with this level of detail.

When a production fires, its action usually changes the contents of the blackboard. You can think of productions as posting notes for each other in order to direct complicated sequences of actions. Since the notes establish a record of what is going on externally, the interpretation of a piece of information currently on the blackboard may be guided by the context of past actions.

The automobile driving example can be extended to illustrate these principles. Here is an initial production system for driving:

1. If the goal is to move \( \rightarrow \) depress the accelerator.
2. If the goal is to stop \( \rightarrow \) remove foot from the accelerator and depress the brake.
3. If the light is red \( \rightarrow \) set a goal to stop.
4. If the light is green \( \rightarrow \) set a goal to move.
that matter, to the carpenter's 3:4:5 rule. The equation plus the definitions tell the problem solver what data to retrieve and how to manipulate the data. In the case of experts, the steps have been incorporated into a memorized schema. In the case of novices, the steps have to be worked out one at a time.

The dairy worker example can also be seen as a case of schematic reasoning. However, there is an important difference between the dairy worker's schema for visualizing milk cartons and the physicist's schema for solving balance of force problems. The dairy worker's schema was defined in terms of the manipulation of concrete objects—crates of milk cartons. Cues for its application are perceivable in the external world. The balance of force schema is defined in terms of abstractions. Therefore, we expect the balance of force schema to be more difficult to apply than the dairy worker schema, because the physicist has no external cue to indicate when the schema can be applied. The physicist must be able to make the leap beyond the perceivable, surface aspects of a problem to see it in its abstract form. (The abstract form is sometimes called the deep structure of the problem.) Learning to do this appears to be an important part of acquiring expertise in physics (Chi, Feltovitch, and Glaser, 1981) and in most other fields.

There is an important trade-off between deep and surface structures. Schemata based on surface characteristics are easy to apply and provide quick solutions to problems. However, they are generally limited in their applicability. The experienced dairy workers' visual schema, for instance, would be defeated if the dairy shifted to metric units, or even
definition, workers who are on the job can perform the tasks that the job requires. But suppose new tasks are added to the job description? Will the worker be able to learn to perform them? This often depends upon how the worker has learned the original task.

Two quite different types of learning are possible, *procedural* and *declarative* learning. Roughly, the distinction between the two is the distinction between learning to do something (especially, make motor movements) and learning to know something, so you can examine your own knowledge. This distinction will be important in discussing the coming workforce, so it will pay to look at a laboratory task that highlights the difference between procedural and declarative learning.

**Procedural and Declarative Learning**

A learner is aware of declarative learning and often uses overt strategies to control what is learned. No student ever learned Shakespeare’s plays without being aware that he or she was learning them! On the other hand, procedural learning can take place without conscious awareness. Some elegant work by Mary Jo Nissen and her colleagues at the University of Minnesota illustrates the distinction.

Figure 5.10 shows the experimental apparatus they used [Nissen and Bullemer, 1987]. College students sat in front of a bank of lights. There was a numbered button below each light. The lights blinked on and off one at a time. The task was to push the button under each light as the light came on.

In fact the lights were turned on in a regular cycle—for example, 1-5-10-9-8-4-3-6-7. As this cycle was repeated, students began to punch the buttons more and more quickly.

This is hardly surprising. What is surprising is that some people
Table 5.1  Job Description Excerpts for the City Attorney, Fort Worth, Texas

Distinguishing Features

This is administrative and legal work of a highly difficult and responsible professional nature in providing counsel for the general administrative offices of the city.

Examples of Work

Directs activities of the legal department with responsibility for the proper and efficient handling of the legal affairs in which the city is interested.

Provides legal advice to the city and all officers and departments.

Prosecutes and defends suits for and on behalf of the city in all courts.

Prepares all contracts, bonds, and other legal instruments for the city.

Required Knowledge, Abilities, Skills, and Other Attributes

Extensive knowledge of city codes and charter provisions and the general laws of the city and state. Extensive knowledge of the sources of legal reference. Extensive knowledge of court procedures and the rules of evidence. Ability to organize, interpret, and apply legal principles and knowledges [sic] to particular circumstances.


problems and is supposed to be ready to justify his or her solutions explicitly, upon demand by judges and juries.

The second aspect of the job is its emphasis upon classification. The attorney is supposed to be able to fit actual cases into their appropriate legal pigeonholes. Case studies of judicial decision making have shown that this is one of the most important parts of an attorney's job. Experienced attorneys develop systems of "if \( \rightarrow \) then rules" that guide them in the classification process. Sometimes these are codified into the law itself. For instance, the Washington State law covering driving under the influence of liquor can be rephrased as

1. If the defendant's blood alcohol reading was above .10, assume that the defendant was legally drunk.
2. If the blood alcohol reading was below .05, assume that the defendant was legally sober.
3. If the blood alcohol reading was between .05 and .10, inquire into the defendant's behavior at the time of the incident.
Figure 6.1 The older method of making cardboard boxes. The operator received information about quality of output only after a production run had been completed.

Job consists in telling the computer what sort of flight is desired and monitoring it to be sure that everything is going well. The word "monitoring" is important. An executive in a forest products company gave me a good example of why.

Figure 6.1 shows how paper companies used to make cardboard boxes. Paper, glue, and various chemicals went into one end of a machine, and a box came out the other end. The machine operator set controls causing the machine to cut different lengths or to heat the product to different temperatures during the gluing step. This allowed the machine to produce different products and to compensate for varia-
Figure 6.2  The new methods for making cardboard boxes. The operator and assistant monitor quality control throughout production.

Section in the quality of the input, such as might occur when paper varied in water content. When the boxes came out of the machine, they were stacked by a fairly low-paid assistant.

The box-manufacturing process, like any other industrial process, was embedded in a system. Periodically, boxes would be inspected by a quality control inspector. Hours, days, or more could elapse between inspections. If the product from a particular machine was unsatisfactory, the line foreman would be told. The foreman would then direct the machine operator to reset the machine so that the error would not recur. This could be done effectively only if the foreman and the operator could reconstruct the situation at the time the error occurred.

Figure 6.2 shows how the manufacturing process has been changed.
stripped of detail, so that top management can concentrate on the essentials. Information flowing downward is expanded to suit local circumstances within the general framework set by top management.

Modern communications and computer technologies have reduced the argument for hierarchical control, because they have increased the amount of information that can be handled at each node. The result has been a tendency toward "flattened hierarchies," where the number of nodes reporting to the central node is decreased. In the extreme, the middle management level simply disappears. The effects of this tendency can be seen by examining two very different institutions: universities and banks.

University Organization and the Computer

In most universities day-to-day teaching and research take place in the department headed by a senior professor who serves as department chair. Because the chairholder is a specialist in the department's discipline, he or she is usually conversant with the work of individual professors, who do most of the teaching and research. In fact, department chairs are usually expected to maintain some teaching and research programs while they are serving as administrators. This means that at the first line of management chemists coordinate the activities of
appropriate authorities rather than on command. At times several different nodes may "spontaneously" organize themselves into a unit to deal with a particular problem. The people within the system have to be flexible enough to understand each other's capabilities, and to appreciate what their role is, not in general, but in the task at hand.

There is no substitute for thinking. But what does thinking mean in the face of computer technology?

DEALING WITH COMPUTERS

Shoshana Zuboff [1989] called her analysis of the workplace In the Age of the Smart Machine. The phrase was well chosen. The greatest change between the workplace that was, in 1950, and the workplace that will be, by 2050 everywhere, and in some places much sooner, will be the extent to which people interact with smart machines. What roles are people going to play in this workplace, and what will the psychological demands of those roles be? To answer these questions, we have to consider what a computer is and how it fits into other changes in the workplace.

The phrase "the computer thinks" has come into our language, but it does not mean the same thing as "I think" or "You think." Computers do not deal with the world; computers deal with mathematical models of the world. Formally, a mathematical model of the world is an abstract system of variables. The value of each variable is deter-
The first step is to determine precisely what real-world operations are to be placed under computer control. They can vary from the cut and paste operations in a word-processing program to sending out signals to control the thrust in an aircraft’s jet engine.

Once the real-world system is defined, it has to be translated into a mathematical formulation. Physical operations on physical things, like erasing words in a typescript or moving the concave surfaces on an aircraft, must be redescribed as mathematical operations on variables.

Finally, the mathematical operations themselves are translated into a computer program. The computer program is not what we normally think of as a mathematical object. Instead, it is a list of physical operations, such as “fetch two numbers from locations x and y, add them,
Table 6.1 Estimate of the Percentage of Jobs Created Between 1985 and 2000 Requiring Certain Reading Skills with Entering Workforce Expected to Have These Skills

<table>
<thead>
<tr>
<th>Reading Skill Level</th>
<th>Percentage of Entering Workers at This Skill Level</th>
<th>Percentage of New Jobs Requiring Skill Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.0</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>71.0</td>
<td>38</td>
</tr>
<tr>
<td>3</td>
<td>17.0</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>3.0</td>
<td>31</td>
</tr>
<tr>
<td>5</td>
<td>1.5</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>.5</td>
<td>1</td>
</tr>
</tbody>
</table>


Notes: Skill level 1 = 2,500-word vocabulary, reads 95–125 words/min, writes simple sentences; 2 = 5,000–6,000-word vocabulary, reads 190–215 words/min, writes compound sentences; 3 = reads safety rules and equipment operating manuals, writes simple reports; 4 = reads journals and manuals, writes business letters and reports; 5 = reads scientific/technical journals and financial reports, writes journal articles and speeches; 6 = same skills as 5, at advanced levels.

Table 6.1. Clearly, the new jobs are biased toward a requirement for high skills, while the new workforce is biased toward a prevalence of low skills.¹

Statistical summaries such as Packer's are useful, but they do not communicate a psychological intuition about what has happened. To do this, I have constructed a small fable.

ALEXANDER THE GREAT, KING HENRY V, AND JOHN HENRY, A STEEL-DRIVIN' MAN

A sad fact of human history is that military technology has often outpaced civilian technology. The rate of progression of military technology dramatizes the changing nature of jobs through history.

Alexander the Great (356–323 B.C.) and King Henry V (1387–1422) had brief lives, even for their times, but their military accomplishments influenced world events for hundreds of years after they died. Would they have understood each other's worlds? Suppose that Alexander had been plunked down in fifteenth century England. Would he have been able to lead Henry's army?

Alexander would have had to learn about some new technologies. Englishmen in the fifteenth century used saddles with stirrups, levers, and gears, none of which were known in Macedonian Greece. Alexan-
Figure 7.1 shows a simplified diagram of the progression from student to worker in U.S. society. On leaving high school (with or without a diploma) students enter either the workforce, the college-university system, or the technical-vocational training system. Let us look at what happens at each of these points.

From the psychologist's point of view, technical-vocational trainers face the easiest educational problem in dealing with the entering workforce. The instructional goals are well defined: Can students perform certain technical skills at the end of the course? The students to be educated are relatively homogeneous. They will have passed through the same school system and will have the same (minimal) amount of experience with the subject being taught. Homogeneity and motivation have been further ensured by student self-selection. Thus, the need for individualized instruction is reduced.

The principal problem with U.S. technical-vocational instruction at the entering worker level is that there is not enough of it. Compared with Europe or Japan, the United States funnels a much higher percentage of its high school graduates into the college-university system and a concomitantly lower percentage into the vocational-technical system. Several critics have argued that our resources should be redistributed, but this is an economic rather than a psychological issue.

In 1991, 62 percent of high school graduates enrolled in colleges or universities. This historic high was not the result of a sudden surge in
grades possible, regardless of whether or not they have truly mastered the material.

The problem with many proposals to improve the educational system is that they seem to assume that assessments designed for certification will, almost as a by-product, serve the function of providing feedback for educational improvement. To see that this is not so, let us consider how assessment fits into a traditional, lecture-style school environment. Figure 7.2 shows the results. At risk of parody, it is not too far afield to say that in this environment the curriculum makers (principals, curriculum designers, and vocational program experts) provide students and tell the teacher what to do with them. The instructor starts his or her lecturing machine, and the students stack knowledge into little boxes in their heads. Sometime later, at assessment time, an external examiner determines what those boxes contain.

Figure 7.2 should look familiar. It is a relabeled version of Figure 6.1, which showed the old-style, inefficient way to make cardboard boxes.