DOCUMENT RESUME

ED 462 394 TM 024 911

AUTHOR	Bishop, John
TITLE	The Economics of Employment Testing. Working Paper.
INSTITUTION	Cornell Univ., Ithaca, NY. Center for Advanced Human Resource Studies.
SPONS AGENCY	National Center for Research in Vocational Education, Berkeley, CA.; National Commission on Testing and Public Policy.
REPORT NO	CAHRS-WP-88-14
PUB DATE	1989-03-24
NOTE	90p.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Economic Factors; Estimation (Mathematics); Females; *Labor Force; Minority Groups; *Occupational Tests; *Personnel Selection; Productivity; Standardized Tests; *Test Use
IDENTIFIERS	General Aptitude Test Battery; Output Measurements

ABSTRACT

Greater use of employment tests for selecting workers will mean that rewards for developing competencies measured by the tests will rise, increasing the supply of workers with the skills. Greater use of tests to select workers will also change the sorting of workers across jobs. The impact on total output will depend on the extent to which the developed abilities measures by employment tests have larger impacts on worker productivity in some occupations than in others. This question was examined by analyzing General Aptitude Test Battery revalidation data for 31,399 workers in 159 occupations and by reviewing the literature on how the standard deviation of worker productivity varies across occupations. The analysis found that differentials do exist and that reassigning workers who do well on a test to occupations where the payoff to talent is particularly high will increase aggregate output. The magnitude of the output effect was esimated, taking into account effects on women and minorities. Ways in which employment tests can simultaneously strengthen incentives to learn, improve sorting, and minimize adverse impacts on minority groups are discussed. Appendix A is a worker evaluation chart. Appendix B contains four tables of output variability, and Appendix C is a table of weights for revalidation data. (Contains 3 figures, 6 tables, 91 references for the text, 5 appendix references, and 57 sources for the appendix tables.) (SLD)

3/24/89
Draft--Not for Quotation

THE ECONOMICS OF EMPLOYMENT TESTING

John Bishop
Cornell University
Working Paper # 88-14

Center for Advanced Human Resource Studies
New York State School of Industrial and Labor Relations
Cornell University
Ithaca, New York 14851-0925
607-255-2742

This is a draft of a paper that is to appear in Testing and Public Policy edited by Bernard Gifford, Chair of the Commission on Testing and Dean of the Graduate school of Education at the University of California-Berkeley. The research that has culminated in this paper was sponsored by the Center for Advanced Human Resource Studies, the National Center for Research in Vocational Education and the Commission on Testing and Public Policy. I would like to thank Peter Mueser, Marc Bendick, David Levine, Paul Ong, Andrew Weiss, Sheldon Zedeck and Nambury Raju for helpful comments on earlier versions of the paper. The opinions and conclusions expressed herein are solely those of the author and should not be construed as representing the opinions or policies of any agency of the United States Government. This paper has not undergone formal review or approval of the faculty of the ILR school. It is intended to make results of Center research available to others interested in human resource management in preliminary form to encourage discussion and suggestions.

ABSTRACT

Greater use of employment tests for selecting workers will have important effects on the economy. First, the rewards for developing the competencies measured by the tests will rise and this will increase the supply of workers with these competencies. Employment tests predict job performance because they measure or are correlated with a large set of developed abilities which are causally related to productivity and not because they are correlated with an inherited ability to learn. Our economy currently under-rewards the achievements that are measured by these tests and the resulting weak incentives for hard study have contributed to the low levels of achievement in math and science.

Greater use of tests to select workers will also change the sorting of workers across jobs. Its impacts on total output depends on the extent to which the developed abilities measured by employment tests--academic achievement, perceptual speed and psychomotor skills--have larger impacts on worker productivity in dollars in some occupations than in others. This question is examined by analyzing GATB revalidation data on 31,399 workers in 159 occupations and by reviewing the literature on how the standard deviation of worker productivity varies across occupations. The analysis finds that indeed such differentials exist and therefore that reassigning workers who do well on a test to occupations where the payoff to the talent is particularly high will increase aggregate output. The magnitude of the output effect was estimated by reweighting the GATB revalidation data to be representative of the 71 million workers in the non-professional and non-managerial occupations and then simulating various resorting scenarios. Selecting new hires randomly lowered aggregate output by at least $129 billion or 8 percent of the compensation received by these workers. An upper bound estimate of the productivity benefits of reassigning workers on the basis of three GATB composites is that it would raise output by $111 billion or 6.9 percent of compensation. Reassignment based on tests had an adverse impact on Blacks and Hispanics but greatly reduced gender segregation in the work place and substantially improved the average wage of the jobs held by women. These results are based on a maintained assumption--the models of job performance which were estimated in samples of job incumbents are after corrections for measurement error and selection on the dependent variable yield unbiased estimates of true population relationships--that is almost certainly wrong. The biases introduced into the calculation by this assumption lower the estimated costs of introducing random assignment of workers to jobs, exaggerate the benefits of greater test use and exaggerate the changes in demographic composition of occupational work forces.

The paper concludes with a discussion of ways in which employment tests can simultaneously strengthen incentives to learn, improve sorting and minimize adverse impacts on minority groups.

4/3/89

THE ECONOMICS OF EMPLOYMENT TESTING

Employment testing appears destined to have a growing role in the allocation of workers to jobs. The competencies measured by these tests are becoming increasingly important. Unskilled manufacturing jobs are moving to Asia, Africa and Latin America. If they are to remain in the US, manufacturers must automate and this in turn necessitates a more skilled and flexible workforce (Adler 1986; Hirschhorn 1984). Employers are complaining that many new hires and long service employees do not have the reading, math and reasoning skills necessary to learn the demanding jobs being generated by the new information technology. At the same time that the demand for more skilled workers is rising, the supply appears to be contracting. The test scores of high school students fell during the 1970s and while they have rebounded somewhat, they have not yet returned to their former level.

These forces are causing American manufacturers to become more selective when they hire new workers. At the same time, the legal impediments to the use of aptitude tests may be diminishing (McDowell and Dodge 1988). Even if the trend of court decisions accepting the claims of validity generalization were to be reversed, employers and society can gain most of the benefits of improved selection by top down hiring from a ranking generated by race normed test scores (Schmidt 1988; Wigdor and Hartigan 1988). Consequently, there is no necessary conflict between minority interests and greater use of tests in employment selection. As a result, test use appears to be growing. A 1985 American Society for Personnel Administration survey (BNA 1986) found that 24 percent of the firms responding had increased testing in the past year and another 44 percent were considering an increase in the amount of testing they do.

Greater use of employment tests will effect the economy in two ways. First, it increases the rewards for developing the skills and competencies assessed by the tests and, as a result, their supply of workers with these skills is likely to increase. Students will see a benefit to devoting more time and energy to their studies and parents will see a stronger connection between the quality of local schools and their child's career success.

Secondly, the sorting of workers across jobs and occupations will change. Employment tests yield information on the probable job performance of job applicants that is not available from other sources (Dunnette 1972; Ghiselli

1973; Hunter 1986; Schmidt 1988). If a trait measured by a test has a larger effect on dollars of output in occupation A than in occupation B, recruiting people who do well on the test into occupation A will increase national output. Greater use of tests for selection is also likely to change the gender breakdown and ethnic makeup of particular occupations.

These two effects of employment testing are the subject of the paper. Incentive effects are examined in Part I and sorting effects are examined in Part II. The paper concludes with a discussion of the incentive and sorting efficiency effects of different methods of selecting workers for jobs and then recommends an approach to employment testing which simultaneously strengthens incentives to learn and improves the sorting of workers across jobs.

PART I. INCENTIVE EFFECTS

> General ability or "intelligence" refers to a repertoire of information-processing skills and habits....These skills and habits must be developed.(p. 29)
>
> ...intelligence tests...is an unfortunate label. It is too easily misunderstood to mean that intelligence is a unitary ability, fixed in amount, unchanged over time, and for which individuals can be ranked on a single scale. (p.28)
>
> Achievement and aptitude tests are not fundamentally different. They both measure developed ability, they often use similar questions, and they have often been found to yield highly related results. Rather than two sharply different categories of tests, it is more useful to think of "aptitude" and "achievement" tests as falling along a continuum. (National Academy of Sciences Committee on Ability Testing, 1982 p. 27).

The professional consensus appears to be that employment tests measure **abilities, skills and habits which must be developed** and which are, therefore <u>malleable</u>. How malleable depends on the nature of the skill and the power of the educational intervention. Evidence of the malleability of the skills measured by employment tests can be found in a variety of literatures. Adoption studies have found that children adopted by upper middle class parents have significantly higher IQ and academic achievement than the siblings who remain with their lower class parents (Schiff et al 1978, 1982, Dumaret 1985, Duyme 1985). Other studies have shown that scores on academic achievement tests improve over the course of the school year and then decline during the summer vacation (Heyns 1987), improve more rapidly for those in school than for drop outs (Husen 1951; Hotchkiss 1984) and improve more rapidly if the student pursues a rigorous college prep curriculum (Bishop 1985; Hotchkiss 1984). The important effects of environment on these developed abilities is also demonstrated by the upward trend of national mean scores on IQ tests (Tuddenham 1948; Flynn 1987), by the large fluctuations in scores on broad spectrum achievement tests (scores of Iowa seniors on the Iowa Test of Educational Development rose .58 standard deviations between 1942 and 1967 and then fell by .35 standard deviations between 1967 and 1979, Forsyth 1987) and by the rapidly closing gap between black and white achievement in National Assessment of Educational Progress data (see Table 1). In the early NAEP assessments black high school seniors born between 1952 and 1957 were 6.7 grade level equivalents behind their white counterparts in science proficiency, 4 grade level equivalents behind in mathematics and 5.3 grade level equivalents behind in reading. The most recent National Assessment data for 1986 reveals that for blacks born in 1969, the gap has been cut to 5.6 grade level equivalents in science, 2.9 grade level equivalents in math and 2.6 grade level equivalents in reading (NAEP 1988, 1989). Koretz's (1986 Appendix E) analysis of data from state testing programs supports the NAEP findings.

Since the abilities measured by employment tests are malleable, it is important to take into account the effects of employment testing on the supply of skilled people. Greater use of tests measuring competence in reading and mathematics for selecting workers will increase the rewards for having these skills. This is likely to have two effects: students will devote more time and energy to developing these skills and parents will become more willing to pay higher taxes to achieve higher standards in their local schools. This judgement follows from four propositions which will be defended below:

1. The American labor market under-rewards the developed abilities measured by these tests. Even though academic achievement has substantial effects on worker productivity, most employers do not base hiring decisions on achievement in high school because grades are not comparable across high schools, transcripts are hard to obtain in a timely manner and administering employment tests risks costly litigation.

2. Young people would devote more time and energy to developing these abilities if the rewards were greater.

3. Parents would be more likely to demand higher standards of their local schools and to support the tax increases necessary to pay for better schools if their child's future depended more directly and visibly on how much is learned in high school.

4. The substantially better performance of European, Canadian, Australian and Asian secondary school students on international mathematics, science and geography exams results in part from the substantially greater economic rewards these societies give learning achievements in high school.

The first of these propositions is defended in the section 1.1. The labor market fails to appropriately reward effort and achievement in high school primarily because employers do not have access to reliable information on the academic effort and achievements of recent high school graduates. Section 1.2 addresses the second proposition by examining student incentives to study hard in high school. Section 1.3 analyzes incentives to upgrade local schools. Section 1.4 examines incentives to learn in Europe, Australia and Japan and concludes that labor market rewards for achievement in high school are much stronger in these societies than in the US; this is one of the reasons why their students study longer hours and learn much more math and science than American students.

1.1 The Absence of Major Economic Rewards for Effort in High School

Signals of learning such as years of schooling which are visible to all are handsomely rewarded. In 1986 25 to 34 year old male (female) college graduates working full time full year earned 44 (49) percent more than high school graduates and high school graduates earned 22 (23) percent more than high school dropouts. Schooling also reduces the risk of unemployment. These rewards have significant effects on student enrollment decisions. When the payoff to a college degree for white males fell in the early 1970s, the college attendance rates of white males fell substantially (Freeman 1976b). When the payoff to college rose again during the late 1970s and 1980s, male college attendance rates rose as well. Years of schooling is only a partial measure of learning accomplishment, however.

In contrast to years spent in school, the effort devoted to learning in high school and the actual competencies developed in high school are generally not well signaled to colleges and employers. Consequently, while students are generously rewarded for staying in school, the students who do not aspire to attend selective colleges benefit very little from working hard while in high school. This is in large measure a consequence of the failure of the labor market to reward effort and achievement in high school.

Students who plan to look for a job immediately after high school generally see very little connection between their academic studies and their future success in the labor market. When 10th graders were asked which math and science courses they needed "to take to qualify for their first choice of job", only 18 percent checked trigonometry or calculus, 20-23 percent checked physics, chemistry, biology and geometry and 29 percent checked algebra (Longitudinal Survey of American Youth 1988). Statistical studies of the youth labor market confirm their skepticism about the economic benefits of taking the more difficult courses and studying hard:

- For high school students, high school grades and performance on academic achievement/aptitude tests have essentially no impact on labor market success. They have -
  --no effect on the chances of finding work when one is seeking it during high school, and
  --no effect on the wage rate of the jobs obtained while in high school.(Hotchkiss, Bishop and Gardner 1982)

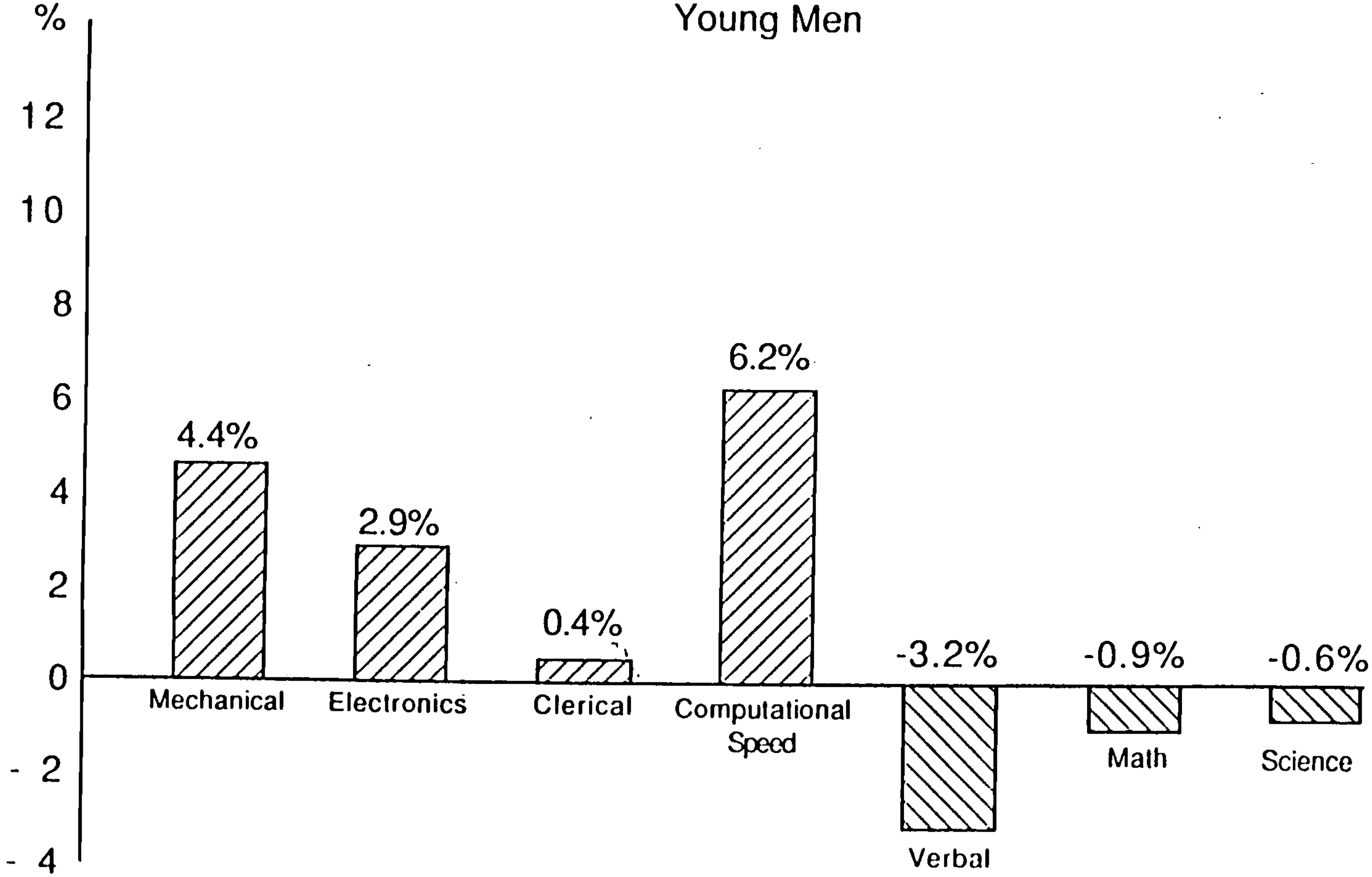
- As one can see in table 2, for those who do not go to college full-time, high school grades and test scores had -
  --no effect on the wage rate of the jobs obtained immediately after high school in Kang and Bishop's (1985) analysis of High School and Beyond seniors and only a 1 to 4.7 percent increase in wages per standard deviation (SD) improvement in test scores and grade point average in Meyer's (1982) analysis of Class of 1972 data.
  --a moderate effect on wage rates and earnings after 4 or 5 years [Gardner (1982) found an effect of 4.8 percent per SD of achievement and Meyer (1983) found an effect of 4.3 to 6.0 percent per SD of achievement],
  --a small effect on employment and earnings immediately after high school.

[Figure 1 and 2 about here]

- Results of an analysis of the Youth Cohort of the National Longitudinal Survey are summarized in figures 1 and 2 (Bishop, 1988). It was found that during the first 8 years after leaving high school, young men received no rewards from the labor market for developing competence in science, language arts and mathematical reasoning. The only competencies that were rewarded were speed in doing simple computations (something that calculators do better than people) and technical

Figure 1

Effect of Competencies on Wage Rates, 1983-1986 Young Men



Source: Analysis of NLS Youth data. The figure reports the effect of a one population standard deviation increase in Armed Services Vocational Aptitude Battery subtest while controlling for schooling, school attendance, age, work experience, region, SMSA residence and ethnicity.

competence (knowledge of mechanical principles, electronics, automobiles and shop tools). For the non-college bound female, there were both wage rate and earnings benefits to learning advanced mathematics but no benefits to developing competence in science or the technical arena. Competence in language arts did not raise wage rates but it did reduce the incidence of unemployment among young women.

- In almost all entry-level jobs, wage rates reflect the level of the job not the worker's productivity. Thus, the employer immediately benefits from a worker's greater productivity. Cognitive abilities and productivity make promotion more likely, but it takes time for the imperfect sorting process to assign a particularly competent worker a job that fully uses that greater competence -- and pays accordingly.

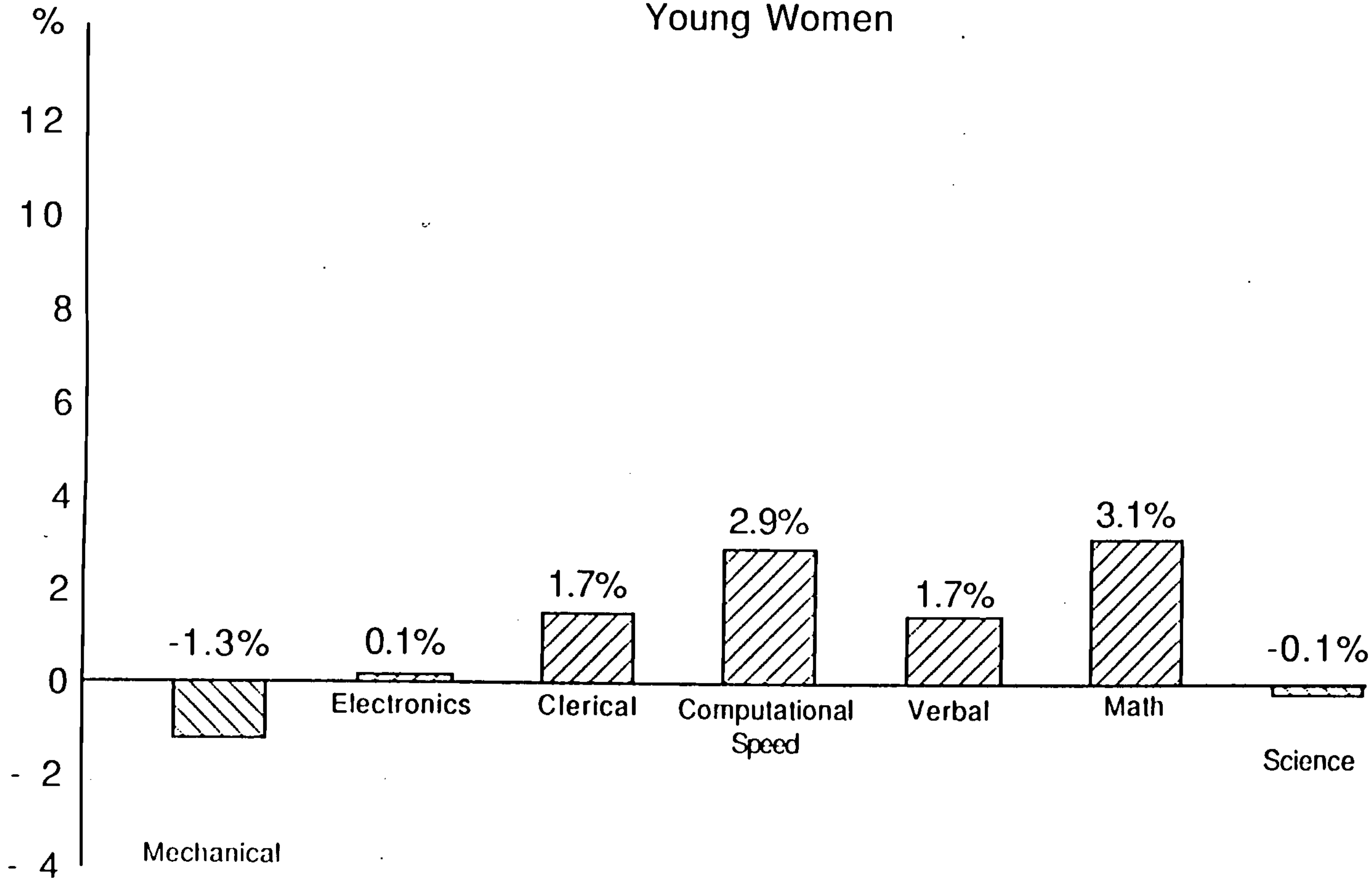
The long delay before labor market rewards are received is important because most teenagers are "now" oriented[1], so benefits promised for 10 years in the future may have little influence on their decisions.

Although the economic benefits of higher achievement are quite modest for young workers and do not appear until long after graduation, the benefits to the employer (and therefore, to national production) are immediately apparent in higher productivity. This is the implication of the finding that tests of mathematical, verbal and problem solving ability are valid predictors of job performance in most civilian and military jobs (Ghiselli 1973; Hunter 1983; Hunter, Crosson and Friedman 1985). A recent study of Marine recruits found, for example, that holding a battery of other tests constant that a one standard deviation increase in two mathematical reasoning subtests increased a work sample measure of job performance by .183 SD in skilled technical jobs, .24 SD in skilled electronic jobs, .34 SD in general maintenance jobs, .447 SD in clerical jobs, .22 SD for missile battery operators and food service jobs and .416 SD in field artillery jobs. Verbal and science subtests also had significant effects on job performance. Holding other tests constant, a standard deviation increase on four subtests measuring mechanical and technical knowledge resulted in a job performance gain of .415 SD in skilled technical jobs, of .475 SD in skilled electronics jobs, of .316 SD in general maintenance jobs, .473 SD in mechanical maintenance jobs, of .450 SD for missile battery operators and food service workers, of .345 SD in combat occupations and .270 SD in field artillery (Bishop 1988b).

Figure 3 compares the percentage impact of mathematical and verbal achievement [specifically a one standard deviation difference in GPA (.7

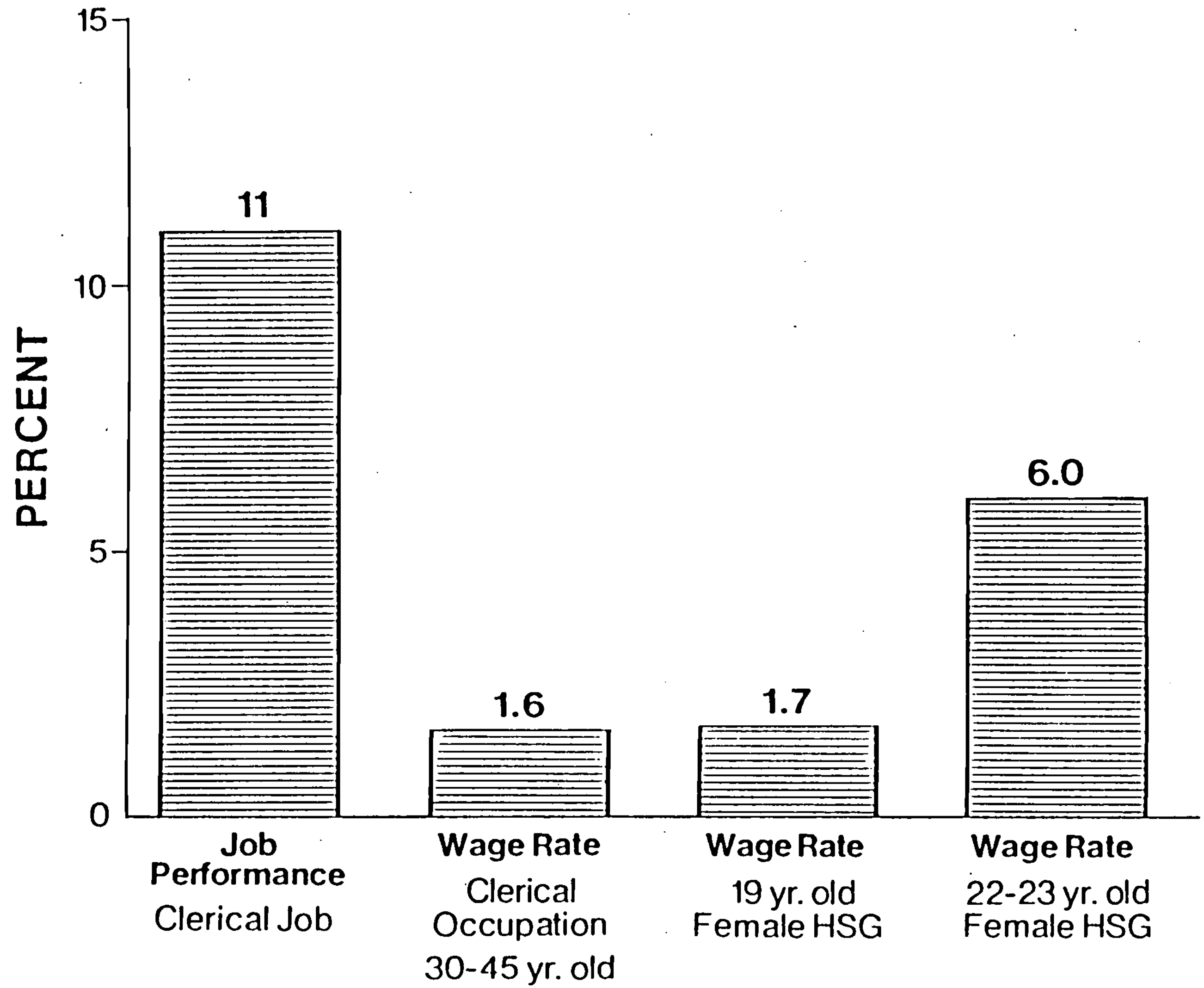
Figure 2

Effect of Competencies
on Wage Rates, 1983-1986
Young Women



Source: Analysis of NLS Youth data. The figure reports the effect of a one population standard deviation increase in Armed Services Vocational Aptitude Battery subtest while controlling for schooling, school attendance, age, work experience, region, SMSA residence and ethnicity.

Figure 3


PERCENT
15
10
5
0
11
1.6
1.7
6.0
Job Performance
Clerical Job
Wage Rate
Clerical Occupation
30-45 yr. old
Wage Rate
19 yr. old
Female HSG
Wage Rate
22-23 yr. old
Female HSG

points) and test scores (3.5 grade level equivalents)] on the productivity of a clerical worker, on wages of clerical workers, and on the wages of all workers who have not gone to college[2]. Productivity clearly increases more than wage rates. Since achievement in mathematical reasoning, science and language arts has no effects on the wage rates of young men, the contrast between wage and productivity effects is greater for young men. This implies that when a non-college-bound student works hard in school and improves his or her academic achievements the youth's employer benefits as well as the youth. The youth is more likely to find a job, but not one with an appreciably higher wage. In the next sub-section we explore the reasons for the discrepancy.

Reasons for the Discrepancy between Wage Rates and Productivity on the Job

Employers are presumably competing for better workers. Why doesn't competition result in much higher wages for those who achieve in high school and have strong basic skills? The cause appears to be the lack of objective information available to employers on applicant accomplishments, skills, and productivity.

A 1987 survey of a stratified random sample of small and medium sized employers who were members of the National Federation of Independent Business (NFIB) found that aptitude test scores had been obtained in only 2.9 percent of the hiring decisions studied (Bishop and Griffin, forthcoming). Top down hiring on the basis of test scores is even more unusual. Prior to 1971, employment testing was more common. The cause of this change was the fear of costly litigation over the business necessity and validity of employment tests. The EEOC's codification of the American Psychological Association's professional testing standards and its theory of situational and subgroup differences in validity into federal law made the required validation studies so costly it discouraged almost all employers from undertaking the effort (Friedman and Williams 1982).

Other potential sources of information on effort and achievement in high school are transcripts and referrals from teachers who know the applicant. Both these means are under used. In the NFIB survey, transcripts had been obtained prior to the selection decision for only 14.2 percent of the high school graduates hired. If a student or graduate gives written permission for a transcript to be sent to an employer, the Buckley amendment obligates the school to respond. Many high schools are not, however, responding to such requests. The experience of Nationwide Insurance, one of Columbus, Ohio's most respected employers, is probably representative of what happens in most communities. The company obtains permission to get high school records from all young people who interview for a job. It sent over 1,200 such signed requests to high schools in 1982 and received only 93 responses. Employers reported that colleges were much more responsive to transcript requests than high schools. High schools have apparently designed their systems for responding to requests for transcripts around the needs of college bound students not around the needs of the students who seek a job immediately after graduating.

There is an additional barrier to the use of high school transcripts in selecting new employees--when high schools do respond, it takes a great deal of time. For Nationwide Insurance the response almost invariably took more than 2 weeks. Given this time lag, if employers required transcripts prior to making hiring selections, a job offer could not be made until a month or so after an application had been received. Most jobs are filled much more rapidly than that. The 1982 NCRVE employer survey of employers found that 83.5 percent of all jobs were filled in less than a month, and 65 percent were filled in less than 2 weeks.

The only information about school experiences requested by most employers is years of schooling, diplomas and certificates obtained, and area of specialization. Probably because of unreliable reporting and the threat of EEOC litigation, only 15 percent of the NFIB employers asked the applicants with 12 years of schooling to report their grade point average. Hiring on the basis of recommendations by high school teachers is also uncommon. In the NFIB survey, when a high school graduate was hired, the new hire had been referred or recommended by vocational teachers only 5.2 percent of the time and referred by someone else in the high school only 2.7 percent.

Consequently, hiring selections and starting wage rates often do not reflect the competencies and abilities students have developed in school. Instead, hiring decisions are based on observable characteristics (such as years of schooling and field of study) that serve as signals for the competencies the employer cannot observe directly. As a result, the worker's wage tends to reflect the average productivity of all workers with the same set of educational credentials rather than that individual's productivity or academic achievement. A study of how individual wage rates varied with initial job performance found that when people hired for the same or very similar jobs are compared, someone who is 20 % more productive than average is typically paid only 1.6 % more. After a year at a firm, better producers received only a 4% higher wage at nonunion firms with about 20 employees, and they had no wage advantage at unionized establishments with more than 100 employees or at nonunion establishments with more than 400 employees (Bishop, 1987a).

Employers have good reasons for not varying the wage rates of their employees in proportion to their perceived job performance. All feasible measures of individual productivity are unreliable and unstable. In most cases measurement must be subjective. Workers are risk averse and reluctant to accept jobs in which the judgement of one supervisor can result in a large wage decline in the second year on the job (Hashimoto and Yu 1980; Stiglitz 1974). Most productivity differentials are either specific to the firm or not visible to other employers, and this reduces the risk that not paying a particularly productive worker a comparably higher salary will result in her going elsewhere (Bishop, 1987a), Pay that is highly contingent on performance can also weaken cooperation and generate incentives to sabotage others (Lazear 1986). Finally, in unionized settings, the union's opposition to merit pay will often be decisive.

Despite their higher productivity, young workers who have achieved in high school and who have done well on academic achievement tests do not receive higher wage rates immediately after high school. The student who works hard must wait many years to start really benefiting and even then the magnitude of the wage and earnings effect--a 1 to 2 percent increase in earnings per grade level equivalent on achievement tests--is considerably smaller than the actual change in productivity that results.

1.2 Will Larger Economic Rewards for Learning Induce Students to Study Harder ?

Learning that is certified by a credential is rewarded handsomely. The magnitude of the earnings payoff to a credential has been shown to have significant effects on the numbers of students entering college and choosing specific majors (Freeman 1971, 1976a, 1976b). Learning not certified by a credential is either not rewarded or only modestly rewarded. Consequently, there are strong incentives to stay in school; but much weaker incentives to study hard while in school. If students are to be motivated to devote more time and energy to learning, they must believe their effort will be rewarded. If parents are to be induced to demand better schools and to spend the time supervising homework, they too must believe that better teaching, a more rigorous curriculum and hard study produces learning which will be rewarded in the labor market. When, however, the only signals of learning accomplishment that are available--eg. GPA and rank in class--describe one's performance relative to close friends, the motivation to study and to demand better schools is undermined.

The Zero-Sum Nature of Academic Competition in High School

The second root cause of the lack of real motivation to learn is peer pressure against studying hard. Students report that "in most of the regular classes... If you raise your hand more than twice in a class, you are called a 'teachers pet.'" Its OK to be smart, you cannot help that. It is definitely not OK to study hard to get a good grade. An important reason for this peer pressure is that the academic side of school forces adolescents to compete against close friends. Their achievement is not being measured against an absolute or an external standard. In contrast to scout merit badges where recognition is given for achieving a fixed standard of competence, the only measures of achievement that receive attention in American schools are measures of one's performance relative to one's close friends such as grades and rank in class. When students try hard and excel in school, they are making things worse for friends. Since greater effort by everyone cannot improve everyone's rank in class, the group interest is for everyone to take it easy. At that age peer friendships are all important, so informal pressure from the peer group is able to induce most students to take it easy. All work groups have ways of sanctioning "rate busters." High school students call them "brain geeks", "grade grubbers" and "brown nosers".

Young people are not lazy. In their jobs after school and at football practice they work very hard. In these environments they are part of a team where individual efforts are visible and appreciated by teammates. Competition and rivalry are not absent, but they are offset by shared goals, shared successes and external measures of achievement (i.e. satisfied customers or winning the game). On the sports field, there is no greater sin than giving up, even when the score is hopelessly one sided. On the job, tasks not done by one worker will generally have to be completed by another. In too many high schools, when it comes to academics, a student's success is purely personal.

Another reason for peer norms against studying is that most students perceive the chance of receiving recognition for an academic achievement to be so slim they have given up trying. At most high school awards ceremonies the recognition and awards go to only a few--those at the very top of the class. By 9th grade most students are so far behind the leaders, they know they have no realistic chance of being perceived as academically successful. Their reaction is often to denigrate the students who take learning seriously and to honor other forms of achievement--athletics, dating, holding your liquor and being "cool"--which offer them better chances of success.

The Consequences of Student Apathy

Studies of time use and time on task in high school show that students actively engage in a learning activity for only about half the time they are scheduled to be in school (Frederick, Walberg and Rasher 1979). In the 1980 High School and Beyond Survey, high school students reported spending an average of only 3.5 hours per week on homework. When homework is added to engaged time at school, the total time devoted to study, instruction, and practice is only 20 hours per week. By comparison, the typical senior spent 10 hours per week in a part-time job and 24 hours watching television (A. C. Neilsen unpublished data). Thus, TV occupies more of an adolescents time than learning.

Even more important is the intensity of the student's involvement in the process. Theodore Sizer described American high school students as **"docile, compliant, and without initiative"** (Sizer 1984, p. 54). John Goodlad (1983) described "**a general picture of considerable passivity among students...** (p. 113)". The high school teachers surveyed by Goodlad ranked "lack of student interest" and "lack of parental interest" as the two most important problems in education. The student's lack of interest makes it very difficult for teachers to be demanding.

Some teachers are able to overcome the obstacles and induce their students to undertake hard learning tasks. But for most mortals the lassitude of the students is too demoralizing. In too many classrooms an implicit agreement prevails in which the students trade civility for lowered academic demands (Sizer 1984). Most students view the costs of studying hard as greater than the benefits, so they pressure the teacher to go easy. All too often teachers are forced to compromise their academic demands.

1.3 Incentives to Upgrade Local Schools

Students are not, however, the only group that is apathetic. Even though American children are far behind Taiwanese and Japanese children in mathematics capability, American mothers are much more pleased with the performance of their local schools than Taiwanese and Japanese mothers. When asked "How good a job would you say ___'s school is doing this year educating___", 91 percent of American mothers responded "excellent" or "good" while only 42 percent of Taiwanese and 39 percent of Japanese parents were this positive (Stevenson 1983). Clearly, American parents hold their children and their schools to lower academic standards than Japanese and Taiwanese--as well as European -- parents.

The apathy of parents, school boards and local school administrators regarding the academic standards of local schools is another negative outcome of the absence of external standards for judging academic achievement and the resulting zero sum nature of academic competition in school. Parents can see that setting higher academic standards or hiring better teachers will not on average improve their child's rank in class or GPA. The Scholastic Aptitude Test does not assess knowledge and understanding of science, history, social science, trigonometry, statistics and calculus or the ability to write an essay. Consequently, improving the teaching of these subjects at the local high school will have only minor effects on how my child does on the SAT, so why worry about standards? In any case, doing well on the SAT matters only for those who aspire to attend a selective college. Most students plan to attend open entry public colleges which admit all high school graduates from the state with the requisite courses. Scholarships are awarded on the basis of financial need, not academic merit.

The parents of children not planning to go to college have an even weaker incentive to demand high standards at the local high school. They believe that what counts in the labor market is getting the diploma, not learning algebra. They can see that learning more will be of only modest benefit to their child's future, and that higher standards might put at risk what is really important--the diploma.

Only when educational outcomes are aggregated, at the state or national levels, do the real costs of mediocre schools become apparent. The whole community loses because the work force is less efficient, and it becomes difficult to attract new industry. Competitiveness deteriorates and the nation's standard of living declines. This is precisely why employers, governors, and state legislatures have been the energizing force of school reform. State governments, however, are far removed from the classroom, and the instruments available to them for inducing improvements in quality and standards are limited. If students, parents and school board officials perceive the rewards for learning to be minimal, state efforts to improve the quality of education will not succeed.

1.4 Incentives to Learn in Other Nations

The tendency to under-reward effort and learning in school appears to be a peculiarly American phenomenon. Grades in school are a crucial determinant of which employer a German youth apprentices with. In Canada, Australia, Japan, and Europe, educational systems administer achievement exams which are closely tied to the curriculum. Performance on these exams is the primary determinant of admission to a university and to a field of study. The resumes of recent secondary school graduates customarily contain a list of the examinations taken and the grade on each exam. Good grades on the toughest exams--physics, chemistry, advanced mathematics--carry particular weight with employers and universities.

In Japan, clerical, service and blue collar jobs at the best firms are available only to those who are recommended by their high school. The most prestigious firms have long term arrangements with particular high schools to which they delegate the responsibility of selecting the new hire(s) for the firm. The criteria by which the high school is to make its selection is, by mutual agreement, grades and exam results. In addition, most employers administer their own battery of selection tests prior to hiring. The number of graduates that a high school is able to place in this way depends on its reputation and the company's past experience with graduates from the school. Schools know that they must be forthright in their recommendations because if they fail just once to make an honest recommendation, the relationship will be lost and their students will no longer be able to get jobs at that firm (Rosenbaum and Kariya 1987).

Japanese teenagers work extremely hard in high school, but once they enter college, many stop working. For students in non-technical fields a country club atmosphere prevails. The reason for the change in behavior is that when employers hire graduates with non-technical majors, they base their selections on the reputation of the university and a long series of interviews and not on teacher recommendations or other measures of academic achievement at the university. Students in engineering and other technical programs work much harder than their liberal arts counterparts largely because job opportunities depend entirely on the recommendation of their major professor. Studying hard is not a national character trait, it is a response to the way Japanese society rewards academic achievement.

American students, in contrast, work much harder in college than in high school. This change is due, in part, to the fact that academic achievement in college has important effects on labor market success. When higher level jobs requiring a bachelors or associates degree are being filled, employers pay more attention to grades and teacher recommendations than when they hire high school graduates. The NFIB survey found that college graduates were hired, 26 percent of the employers had reviewed the college transcript before making the selection, 7.8 percent had obtained a recommendation from a major professor and 6.3 percent had obtained a recommendation from a professor outside of the graduates major or from the colleges's placement office.

Parents in Australia, Canada, Europe and Japan know that a child's future depends critically on how much is learned in secondary school. National and regional exams are the yardstick, so achievement tends to be measured relative to everyone else's in the nation or region and not just relative to the child's classmates. As a result, parents in most other Western nations demand more and get more from their local schools than we do and yet are, nevertheless, more dissatisfied with their schools than American parents. Students in other nations spend much less time watching TV: 60% less in Switzerland and 44% less in Canada (Organization of Economic Cooperation and Development, Table 18.1, 1986) and are much less likely to work part time during the school year. School years are longer. Japanese 5th graders spend 32.6 hours a week in academic activities while American youth devote only 19.6 hours to their studies (Stevenson, Lee and Stigler 1986). Forty-five percent of Japanese junior high school students attend Juku, private schools which provide tutoring in academic subjects (Leestma 1987). By the time they graduate from high school Japanese have spent the equivalent of three more years in a classroom and studying than American graduates (Rohlen 1989).

The greater effort yields greater achievement. In Stevenson, Lee and Stigler's (1986) study of 5th grade math achievement, the best of the 20 classrooms sampled in Minneapolis was outstripped by every single classroom studied in Sendai, Japan and by 19 of the 20 classrooms studied in Taipeh, Taiwan. The nation's top high school students rank far behind much less elite samples of students in other countries. In math and science the gap between Japanese, English, Finnish and Canadian high school graduates and their white American counterparts is more than four US grade level equivalents.

In summary, the lack of true engagement in learning in US high schools and the apathy of local political systems regarding the quality of local schools is to an important degree a consequence of the failure of employers to reward students for real learning achievements. The solution would appear to be for employers (particularly those with attractive jobs) to use measures of academic achievement such as grades, Regents exams and broad spectrum achievement test batteries (eg. the ASVAB) as a selection criterion when hiring recent high school graduates. Such a policy will also increase the validity of employee selection protocols and thus increase the efficiency by which workers are matched with jobs. It is to these sorting effects that we now turn.

PART II. SORTING EFFECTS

Hunter and Schmidt (1982) employ Brogden's formula to calculate the effect of test use on the efficiency of the economy's matching of workers to jobs. In this context Brogden's formula can be viewed as a way of representing for a specific job the derivative of a worker's true productivity ($P^t{}_i$) measured in dollars with respect to a test score (T_i):

$$(1) \quad \frac{\partial P^t{}_i}{\partial T_i} = \frac{Cov(P^t{}_i, T_i)}{Var(T_i)} = r_{TP} \frac{SD(P^t)}{SD(T)}$$

where r_{TP} = true validity, the correlation between true productivity in that job and the test when employees are randomly selected.

$SD(P^t)$ = the standard deviation of output in dollars if the workers had been randomly selected.

$SD(T)$ = the population standard deviation of the test.

They point out that tests are more valid predictors of job performance (eg. have higher r_{TP}) in the more complex jobs that are traditionally better paid and, therefore, probably also have larger standard deviations of productivity in a dollar metric, $SD(P^t)$. When this is the case, output will increase if high scoring individuals are recruited into the most complex jobs and low scoring individuals are recruited into the less complex jobs. They make a simplifying assumption that the ratio of the standard deviation of output in dollars to the wage is the same in all jobs but argue it is quite large, about 40 percent of salary. Under this assumption, they calculate that distributing all workers across four major occupational categories on the basis of a single measure of academic ability will raise productivity 4 percent above the level resulting from random assignment of workers to major occupational category. They also report that assigning workers on the basis of a simple multi-variate selection model involving tests of perceptual speed and spatial ability as well as academic ability would increase productivity by 8 percent relative to random assignment.

However, since people are already recruited into high status jobs on the basis of years of schooling, SAT scores, college major, grades, previous work experience and performance in past jobs (which have independent associations with job performance and together explain much of the variance of test scores), greater use of tests by employers would probably have much smaller effects on national output than those calculated by Hunter and Schmidt. Hunter and Schmidt acknowledge this when they say, "Employers do not select randomly from among applicant pools.... many of these [selection] procedures have low validity, but average productivity levels associated with current methods are certainly above those that would result from random selection from applicant pools, though less effective than our univariate selection strategy (p. 270)". Michael Rothschild (1979) has proposed two other sources of upward bias in their estimate. He argues that the assumption of optimal placement is unreasonable. Tests would never be used by all firms, for all jobs and optimally in every case, so the full benefits calculated would never be realized. A second source of bias, in Rothschild's view, is the possibility that errors in measuring productivity may be positively correlated with test score, and that consequently the estimates of true validity and the standard deviation of true output used in the analysis may be biased. Hunter and Schmidt argue to the contrary that their estimates are conservative because they assume that (1) coefficients of variation of productivity are the same for all occupations, (2) at most three test scores are used to reassign workers and (3) only 4 categories of occupations are analyzed. They point out that these features of the calculation cause it to understate the effects of greater test use on national productivity.

The only way to determine whether the net effect of the offsetting biases makes the H/S estimates too high or too low is to change as many of the problematic assumptions as possible and then redo the calculation. That is what will be attempted in this part of the paper. The objective is an improved estimate of the magnitude of the efficiency gains that may result from greater test use, not a definitive estimate. In the current state of knowledge, a definitive estimate is infeasible for some important sources of bias cannot be eliminated. There is no way of knowing, for example, how effectively tests will be incorporated into selection decisions and whether the measurement errors of job performance are correlated with test scores or not, so it will not be possible to formally address two of Rothschild's objections to H/S's estimates. Most of the factors that Hunter and Schmidt argue cause their estimates to be conservative are dealt with, however, so the resulting estimates are probably upper bounds on the likely impact of greater test use on the productivity of the economy.

Greater use of tests will increase aggregate output either if tests are more valid predictors of job performance in some jobs than others or if improvements in job performance measured in standard deviation units have larger effects on output valued in dollars in some occupations than others. I begin, therefore, by examining how test validity varies across occupations. This is accomplished by estimating "structural" models of relative productivity as a function of three tests scores (general academic achievement, perceptual speed and psychomotor skills), years of schooling, age, total occupational experience, tenure, gender, race and Hispanic background for 8 different occupational categories in the United States Employment Service's General Aptitude Test Battery Revalidation Individual Data File.

The next step is a review of the literature on how variable output is across workers doing the same job and how this variability differs across jobs. The major finding here is that the standard deviation of output is substantially higher in the more cognitively complex and better paid jobs.

The effect of alternative ways of assigning workers to jobs is calculated by simulating such changes in the USES Individual Data File after reweighting it to be representative of all workers outside of professional, managerial and sales representative occupations. The parameters of the "structural" models are used to predict the productivity (in standard deviation units) during the first ten years on the job of all 31,399 workers in the data set in each of the 8 occupational categories analyzed. The mean predicted productivity of workers who currently occupy each job is then compared to the productivity that would result from (1) a random assignment of new hires to jobs and (2) a resorting of new hires across jobs based on the productivity predictions generated by regression equations similar to the structural models but absent data on gender, race and Hispanic background. These results are then translated into a dollar metric by multiplying changes in mean productivity in standard deviation units by estimates of the standard deviation of productivity in dollars obtained from the literature review. The impact of reassignment based on test scores on the gender, racial, and Hispanic composition of each occupation is also simulated and discussed. Part 2 then concludes with a critique of the estimated "structural" models of job performance and the resulting estimates of productivity gains from resorting the workforce on the basis of employment tests.

2.1 Analysis of GATB Validation Studies

Data on the relative productivity of a large and reasonably representative sample of workers is available from the US Employment Service's program for revalidating the General Aptitude Test Battery (GATB). This data set contains data on job performance, the 9 GATB "aptitudes" and background data on 36,614 individuals in 159 different occupations. Professional, managerial and high level sales occupations were not studied but the sample is quite representative of the rest of the occupational distribution. It ranges from drafters and laboratory testers to hotel clerks and knitting-machine operators. The simulations of the effect of changes in selection policies are also conducted in this data set after it has been reweighted to be representative of the 71,132,000 workers who are employed in these occupations.

Since a major purpose of these validation studies was to examine the effects of race and ethnicity on the validity of the GATB, the firms that were selected tended to have an integrated workforce in that occupation. Firms that used aptitude tests similar to the GATB for selecting new hires for the job being studied were excluded. The employment service officials who conducted these studies report that this last requirement did not result in the exclusion of many firms. A total of 3052 employers participated.

Each worker took the GATB test battery and supplied information on their age, education, plant experience and total experience. Plant experience was defined as years working in that occupation for the current employer. Total experience was defined as years working in the occupation for all employers. The dependent variable for this study is a sum of two separate administrations (generally two weeks apart) of the Standard Descriptive Rating Scale. This rating scale (See Appendix A), obtains supervisory ratings of 5 aspects of job performance (quantity, quality, accuracy, job knowledge and job versatility) as well as an "all around" performance rating. Some studies employed rating scales specifically designed for that occupation and in one case a work sample was one of the job performance measures. None of the studies used ticket earnings from a piece rate pay system as the criterion. Studies which used course grades or tests of job knowledge as a criterion were excluded. Firms with only one employee in the job classification were excluded, as were individuals whose reported work experience was inconsistent with their age.

Academic achievement is the sum of two GATB composites, G and N, that have been put into a population SD metric by dividing by 38.8. The G composite is an average of normalized scores on a vocabulary test, an arithmetic reasoning test and a 3-dimensional spatial relations test. The mathematical achievement index (N) is an average of normalized scores on the same arithmetic reasoning test and on a numerical computations test. These two GATB composites were aggregated together because previous analyses had found that when both were entered simultaneously into models predicting relative job performance, the coefficients on both composites were very similar (Bishop 1987). Perceptual Speed is the sum of the P and Q aptitudes of the GATB divided by 36.72 to put it in a population SD metric. Psychomotor Ability is the sum of the K, F and M aptitudes of the GATB divided by 51.54 to put it in a population SD metric.

Because wage rates, average productivity levels and the standards used to rate employees vary from plant to plant, mean differences in ratings across establishments have no real meaning. Only deviations of rated performance ($R^m_{ij}-R^m_j$) from the mean for the establishment (R^m_j) were analyzed. The variance of the job performance distribution was also standardized across establishments by dividing ($R^m_{ij}-R^m_j$) by the standard deviation of rated performance, ($SD_j(R^m_{ij})$, calculated for that firm (or 3 if the sample SD is less than 3).[4] Two models were estimated for each major occupation. They were specified as follows:

$$(2)\quad \frac{R^m_{ij}-R^m_j}{SD_j(R^m_{ij})} = R_{ij} = a_0 + a_1(\underline{T}_{ij}-\underline{T}_j) + a_2(S_{ij}-S_j) + a_3(\underline{X}_{ij}-\underline{X}_j) + v_1$$

$$(3)\quad R_{ij} = \beta_0 + \beta_1(\underline{T}_{ij}-\underline{T}_j) + \beta_2(S_{ij}-S_j) + \beta_3(\underline{X}_{ij}-\underline{X}_j) + \beta_4(\underline{D}_{ij}-\underline{D}_j) + v_2$$

where R_{ij} = ratings standardized to have a zero mean and SD of 1.

$\underline{T}_{ij}$ = a vector of the three GATB composites

S_{ij} is the schooling of the i^{th} individual.

$\underline{X}_{ij}$ = a vector of age and experience variables--age, age^2, total occupational experience, total occupational experience^2, plant experience and plant experience^2.

$\underline{D}_{ij}$ = a vector of dummy variables for black, Hispanic and female.

$\underline{T}_j$, S_j, $\underline{X}_j$ and $\underline{D}_j$ are the means of test composites, schooling, experience variables and race and gender dummies for the j^{th} job/establishment combination. In the first model, standardized ratings are predicted by test composites, schooling and six experience variables. Gender, race and Hispanic are excluded. Because it is illegal for firms to select workers on the basis of gender, race and ethnicity, the selection process must be assumed to ignore this information so the simulation exercise conducted in section 2.3 assumes that workers are assigned to jobs on the basis of performance predictions generated by estimates of equation 2.

In equation 3, normalized ratings deviations are predicted by deviations from the firm's mean for gender, race, Hispanic, age, age squared, plant experience, plant experience squared, total occupational experience, total occupational experience squared, schooling and test composites. The calculation of the effects on aggregate output of reassigning workers to jobs will be based on the predictions of this model. It should be recognized that because of the selectivity of the application and hiring process and of turnover and promotions, the results obtained from fitting this model are not estimates of the true structural relationships prevailing in the full population (Brown 1978; Mueser and Maloney 1987). Since no data sets exist which would enable analysts to model these selection processes, estimates of the true population relationships do not appear to be feasible. An effort will be made in section 2.4 to discuss how the simulation results would probably change if better estimates of true population relationships were available.

The results of estimating equation 2 are presented in Table 3. When test scores are controlled, years of schooling appear to have very small and sometimes negative effects on job performance.[5] The effects of the three test score composites are reported in columns 2-4 of the table. When the metric of job performance is within-job standard deviations, academic achievement has roughly comparable effects on job performance in all occupations except operatives and sales clerks. The effect of academic achievement on the performance of operatives is highly significant but only about two-thirds of the size of the other occupations. Perceptual speed has smaller effects on job performance, but the coefficients are nevertheless significant in all but technical and sales clerk (where the sample is quite small) occupations. Psychomotor skills are significantly related to performance in all occupations but in the better paid and more complex jobs the magnitude of the effect is only about one-third of that of academic achievement. The effect of psychomotor skills is larger in the three least skilled occupations--operatives, sales clerks and service except police and fire. For operatives and sales clerks the impact of psychomotor skills is roughly comparable to the impacts of academic achievement. These results are consistent with previous studies of this data set (Hunter 1983). Models were estimated containing squared terms for academic achievement and psychomotor skills but these additions did not produce significant reductions in the residual variance. Estimating equation 3 by adding dummy variables for gender, race and Hispanic to the equation 2 specification, tends to reduce test score coefficients a little but the pattern remains the same.

The effects of occupational experience and tenure are also quite substantial for all occupations except for sales clerks. The negative coefficients on the square terms for occupational experience and tenure imply they are subject to diminishing returns. For workers who have no previous experience in the field, the expected gain in job performance is about 12-13 percent of a standard deviation in the first year and about 8-9 percent of an SD in the fifth year. The effect of tenure on job performance stops rising and starts to decline somewhere between 16 and 24 years of tenure. Increases in occupational experience lose their positive effect on performance even later--at 37 years for operatives, at over 55 years for craft workers and high skill clerical workers and at 19-31 years for other occupations. Except for technicians, age has large curvilinear effects on job performance as well.

The substantial effects of age and previous occupational experience on job performance are consistent with current hiring practices which give great weight to these job qualifications. These results suggest that a job applicant who has age and relevant work experience in his favor but low test scores may nevertheless be preferable to a young applicant who has high test scores but no relevant work experience. This is particularly likely to be the case if turnover rates are high for the productivity benefits of age and previous relevant work experience are large initially but diminish with time on the job. These results point to the desirability of studying the effects of test scores on job performance in the context of a multivariate model which includes controls for as many other factors as possible. They also remind us that employment tests should not be the sole criterion by which workers are selected. Tests should supplement not displace other criteria for selecting the best job candidate.

2.2 A Review of Studies of Output Variability

The second determinant of the payoff to using tests to select workers is the extent of the variability across workers in their productivity on the job. A search for studies of output variability yielded 49 published and 8 unpublished papers covering 94 distinct jobs. Recent reviews of the literature on SDY by Boudreau (1987) and Hunter, Schmidt and Judiesch (1988) were the source of most the data. The results are summarized in column 1 of table 5 and column 2 of table 4. (The detailed results are reported in Appendix tables 1 through 4). Most of the studies reviewed measured physical amounts of output produced over periods generally lasting one to four weeks and report a ratio of the standard deviation of output to mean output, coefficient of variation or CV. Relative output levels vary over time, so coefficients of variation for a one or five year period are inevitably smaller than the coefficients of variation for a one or two week period. Hunter, Schmidt and Judiesch (1988) review a number of studies which provide evidence on the correlation between output levels over time and how these correlations vary with the length of the time interval studied. This information was then used to construct estimates of the output CVs for periods of a year or more. It is these corrected estimates of the CV which are reported. For semi-skilled factory jobs paid on an hourly basis the coefficient of variation averaged about 14 percent. Output variability is greater in the higher paid technical and precision production jobs. The coefficient of variation averages 27.6 percent in craft jobs and 33.8 percent in technical jobs.

Clerical jobs were divided into high skill and low skill categories. The description of the job in the Dictionary of Occupational Titles was reviewed and jobs which appeared to require greater skill or involve discretion and decision making were classified as "high skill clerical." The jobs which were included in this category were stenographer, computer operator, administrative clerk, supply specialist, claims processor, head teller, ticket agent, customer service representative and teacher aide. Jobs categorized as "routine" were key punch operator, hotel clerk, cashier-checker, telephone operator, mail carriers, file clerks, stock clerk, typists, and toll ticket sorters. This distinction appears to be a real one for the high skill clerical jobs were generally better paid than the routine clerical jobs and the workers in these jobs scored one third of a standard deviation higher on the GATB academic achievement composite than those who occupied the more routine clerical jobs. Furthermore, the variability of job performance appears to be substantially greater in the jobs that require decision making. The coefficient of variation was 25.5 in the high skill clerical jobs and 16.7 percent in the routine jobs.

Data was available for only three service occupations. These three jobs represent too small a sample to produce reliable estimates of the CV for all service jobs except police and fire fighting so the estimate of the service CV employed in the paper is an unweighted average of the CVs for operatives, low skill clerical workers and 20.6, the average for the three service jobs for which there is data on the variability of output. For sale clerks records of sales transactions were employed to calculate the CV and the result was an estimate of 29.8 percent.

When a firm expands by hiring extra workers, it incurs significant fixed costs. It must rent space, buy equipment, hire supervisors and recruit, hire, train, and payroll the additional production workers. If output can be increased by hiring more competent workers, all of these costs can be avoided and the firm's capital becomes more productive. These factors tend to magnify the effects of work force quality on productivity. They imply that the ratio of the standard deviation of worker productivity in dollars (SD$) to average worker compensation is much larger than the productivity CV for that job (Klein, Spady and Weiss 1983; Frank 1984).

Estimates of productivity standard deviations (SD$) in 1985 dollars are reported in column 2 of the table 4. In most cases the author of the study made no attempt to estimate SD$'s, so estimates of SD$ were derived as a product of the CV, the mean compensation for that job and 1.52, the ratio of value added to compensation for private non-farm business excluding mining, trade, finance and real estate. The value added to compensation ratio in retailing and in real estate is much too high to be used as an adjustment factor. So for all sales occupations, it was assumed that SD$ = CV times average compensation. The SD$ that result are $13,668 for technicians, $12,399 for craft workers, $5062 for semiskilled factory jobs, $8925 for high-skill clerical jobs, $4934 for routine clerical jobs, $4068 for service workers other than police and fire fighters and $5228 for sales clerks. While it is possible to debate the accuracy of specific estimates and the reliability of the 15th, 50th, and 85th percentile method of measuring SD$, the basic pattern of rapidly increasing standard deviations of output as one moves up the occupational distribution is unlikely to be disturbed by new data or a revised methodology.

What about jobs where capital equipment controls the pace of work? It has been argued that in automated continuous process industries the amount and quality of output is determined by technology and computer programs not by the skills and talents of the workers. In fact, however, programs cannot be written to handle all contingencies and machines are never completely reliable so human operators have an important role to play (Hirschhorn 1984; Adler 1986). In capital intensive industries with high rates of energy and materials consumption, small errors can cause substantial losses. Small adjustments which increase fuel efficiency can save a utility or refinery millions of dollars a week. This has been demonstrated by a very careful study of the variability of the job performance of the operators of electric utility plants(see Table B2). In the study of the operators of electric generating plants commissioned by the Edison Electric Institute, committees of technical experts were organized and asked to make consensus estimates of the frequency and costs of the most common types of operator errors. Once the relationship between specific operator errors and the purchase costs of replacement power was established, the experts estimated what would be expected (in dollar terms) from an operator at the 15th, 50th and 85th percentile of job performance. The study concluded that the standard deviation for the productivity of control room operators is about $278,000 in 1985 dollars at nuclear plants and $115,000 at fossil fuel plants (Dunnette et al 1982).[6] When the results of Wroten's study of output variability among refinery operators is combined with the results of the Dunnette et al study, the estimated SD$ for this small but very important set of jobs is $91,020. The

SD$ of plant operators is more than 6 times larger than any of the other occupations in the USES Individual Data File. As a result, resorting to maximize total output implies that workers who would be above average producers in all occupations should be assigned to this occupation.

2.3 Simulation Results

The question posed in this section is "What will happen to aggregate output and to the gender and ethnic composition of various occupations, if firms are allowed and/or encouraged to use employment tests to select new hires?" To simulate the effect of changes in the allocation of workers across jobs on aggregate output, one needs estimates of how the effects of test scores and other worker characteristics on productivity vary across jobs. If the data were available, we would want to estimate, for random samples of the population, linear regressions in which the true relative productivity in dollars, $P^t_{ij}-P^t_j$, of the i^{th} worker in the j^{th} job is a function of the worker's characteristics. Unfortunately, in most studies the only indicators of productivity are supervisory ratings which are not defined on a ratio scale and have only limited reliability.

If, however, outside estimates of the standard deviation of true productivity among job incumbents, $SD_j(P^t_{ij})$, are available and assumptions are made about the measurement error in these ratings and about selection effects, estimates of the effect of test scores on true productivity in that occupation can be derived from regression models in which ratings are predicted by test scores and other worker characteristics. The measurement assumptions implicitly made by Hunter and Schmidt and most other contributors to the literature are:

$$(4)\quad R_{ij} = \frac{R^m_{ij}-R^m_j}{SD_j(R^m_{ij})} = \sqrt{r_{pp}}\left[\frac{P^t_{ij}-P^t_j}{SD_j(P^t_{ij})}\right] + v$$

where r_{pp} = the reliability of supervisory ratings (eg. the correlation between independent ratings by two different supervisors in the selected sample of job incumbents).

$SD_j(P^t_{ij})$ = the standard deviation of true productivity in the selected sample of incumbents in job "j".

v is uncorrelated with true productivity.

In other words, the ratings of relative job performance are assumed to be cardinal measures of productivity that are linearly related to true productivity and that errors in assessing productivity are negatively associated with true productivity. This assumption implies that measurement error in the dependent variable attenuates the true relationship. Since the upper bound on the reliability of job performance measures like the Standard Descriptive Rating Scale appears to be .6 (King, Hunter and Schmidt, 1980), the impact of a right hand side variable on true productivity in standard deviation units can be calculated by multiplying the coefficients reported in Table 3 by 1.29, the inverse of the square root of criterion reliability. It is further assumed that $SD_j(P^t_{ij})$ is equal to the $SD\$_j$, the standard deviation of productivity in dollars discussed in section 2.2. While these assumptions may seem reasonable, there do not appear to be any studies which have demonstrated that errors in assessing job performance are negatively correlated with true productivity and only a few studies establishing the reasonableness of the assumption that $SD_j(P^t_{ij}) = SD\$_j$ (Vineberg and Taylor 1972; Corts et al. 1977; Trattner et al 1977). To facilitate comparisons with previous literature, the calculations of output effects presented below are based on the assumptions detailed above.

The second problem that must be dealt with is the fact that job performance outcomes have been used to select the sample used in the analyses. Since incompetent workers are fired or induced to quit and high performing workers are promoted to jobs of a higher classification, job incumbents are a restricted sample of the people originally hired for a job (Bishop 1988a). The systematic nature of attrition from the job substantially reduces the variance of job performance and biases coefficients of estimated job performance models toward zero. When all variables are multivariate normal, the ratio of the coefficients estimated in the selected sample to the true coefficient estimated in an unselected population is equal to:

$$(5) \quad \beta^*/\beta \;=\; VR/(1-R^2(1-VR)) \;=\; VR + R^{*2}(1-VR)$$

where VR is the ratio of the variance of y in the selected sample to its variance in the full population, R^2 is the multiple coefficient of determination of y on $\underline{x}$ in the full population and R^{*2} is the multiple coefficient of determination of y on $\underline{x}$ in the selected population (Goldberger 1981). Estimates of VR, the ratio of incumbent job performance variance to new hire job performance variance can be derived from the NCRVE employer survey analyzed in Bishop (1987a, 1988a). Using reported productivity in the 3rd through 13th week after being hired for two different workers as the data, a variance ratio was calculated by dividing job performance variance of incumbents (pairs of workers both of whom were still at the firm at the time of the interview a year or so after being hired) by the job performance variance of a group of very recent hires (pairs of workers both of whom stayed at least 13 weeks but who may or may not have remained at the firm through the interview). The resulting estimate of VR was .486.[7] Assuming multivariate normality and noting that the R^2 of the models in table 3 averages about .16, our estimate of B/B^*, the multiplier for transforming the coefficients estimated in the selected sample into estimates of population parameters, is 1.76. The reader is reminded that while these corrections deal with some bias problems, others remain, so even with these corrections the simulations presented below are not definitive. The likely effects of the biases that remain will be discussed after the simulation results are presented.

The Productivity Loss from Random Assignment of Workers to Jobs

The first simulation exercise is a comparison of the mean predicted productivity of workers who currently occupy each job to the productivity that would result from a random assignment of new hires to jobs. The parameters of the equation 3 model were used to predict the productivity (in standard deviation units) during each of the first ten years on the job of all 31,399 workers in the data set in each of the 8 occupational categories analyzed.

$$(6)\quad \hat{R}_{ijt} = \hat{B}_{j1}\underline{T}_i + \hat{B}_{j2}S_i + \hat{B}_{j3}\underline{X}_{it} + \hat{B}_{j4}\underline{D}_i + C_j$$

where $\underline{X}'_{it}$ = a vector of age and total occupational experience variables: $(age_i - tenure_{ij} + t)$, $(age_i - tenure_{ij} + t)^2$, (total occupational experience$_{ij}$ - tenure$_{ij}$ + t), and (total occupational experience$_{ij}$ - tenure$_{ij}$ + t)^2.

tenure$_{ij}$ = the plant experience of the i^{th} worker in the j^{th} job/establishment at the time of the GATB study.

t = time since being hired. It ranges from 0 to 10.

total occupational experience$_{ij}$ - tenure$_{ij}$ is the worker's experience

> in the occupation prior to coming to work at the establishment. If the worker is reassigned to a different broad occupational category, this previous occupational experience is set at zero.

The effects of age and previous occupational experience at the time of hire were included along with test scores, schooling, gender and ethnicity. An annualized present discounted value of each worker's predicted productivity during the first ten years was then calculated under the assumption of a 6 percent real interest rate and a monthly turnover rate of 1 percent (which yields a yearly retention rate of .8869).

$$(7) \quad APV_{ij} = \sum_{t=.5}^{9.5} \hat{R}_{ijt}(.8869/1.06)^t / \sum_{t=.5}^{9.5} (.8869/1.06)^t$$

Based on occupation, race and Hispanic status, each worker was assigned a weight so that the USES Individual Data File would become representative of all 71,132,000 workers in these 8 occupations (see Appendix Table C1 for a description of how these weights were derived). The weighted mean annualized present value of predicted productivity resulting from random assignment of new hires to occupations was then subtracted from the weighted mean annualized present value of predicted productivity during the first ten years on the job for the current set of individuals in that occupation. This was then translated into dollars by multiplying first by 1.29, second by 1.76 and then by the $SD\$_j$ for that occupation.

The results of this simulation exercise are presented in Table 4. The loss in productivity that would result from random assignment of workers to jobs is estimated to be about $1800 dollars per worker per year or 8 percent of mean compensation. The aggregate yearly loss is $129 billion in 1985 dollars. The reductions in productivity primarily occur because: (1) workers who had higher than average productivity during their early years at the firm due to previous experience in the occupation are often randomly assigned to an occupation where this previous experience is of no value and (2) workers with high test scores are much less likely to be assigned to high skill jobs which use their talents than is the case currently. These results are clearly an extreme lower bound estimate of the benefits (relative to random assignment) of the current process of matching workers to jobs. If other worker characteristics such as occupationally specific education, tastes and talents for particular occupations and performance in previous similar jobs had been included in the model, estimates of productivity loss resulting from random assignment of workers to occupations would have been substantially greater.

The Productivity Gains from Re-Sorting Workers on the Basis of Test Scores

The effect of greater use of employment tests to select workers on productivity was explored by simulating the effects of reassigning new hires on the basis of the productivity predictions derived from equation 2. A annualized present discounted productivity (averaged over the first ten years on the job) was calculated for each worker in each occupation. The reassignment scheme employed a variant of the "cut and fit" or successive selection technique (Thorndike 1949; Guion 1965). The 8 occupations were arrayed in a hierarchy according to the magnitude of the dollar change in productivity that results from a unit change in academic achievement. Plant operators were at the top of the hierarchy. The computer program sorted all workers by the present discounted value of their predicted productivity as plant operators (based on equation 2) and then assigned just enough people from the top of that ranking to fill all 228,000 of the nation's plant operator jobs. The remaining workers were then sorted by their productivity in technical occupations and those found at the top of the ranking were assigned to these occupations until all 5,261,000 technical jobs were filled. This procedure was repeated next for craft jobs, then for high skill clerical jobs, for low skill clerical jobs, for service jobs, and for operative jobs. Those left over after operatives were selected became sales clerks.[8]

The simulated effects of this reassignment scheme on productivity are presented in Table 5. Output rises by $1561 per worker per year or by 6.9 percent of mean compensation. The total gain from applying this plan to the 71 million workers represented in the data base is $111 billion per year. There are major improvements in the productivity of plant operators, technicians and craft workers which more than offset large declines in the productivity of operatives and sales clerks.[9]

The testing is costly, however, so the net benefits of greater testing will be somewhat smaller. The firm's costs are generally assumed to be about $10.00 per administration. The tests generally take 3 hours to take, so I will assume that the value of the job applicant's time is $24.00 on average. If each employer were to do its own testing and to test 10 applicants for every position filled, the total yearly costs of the testing would be $10.7 billion [.48*10*$34*(71,132,000-5,682,000) assuming a monthly new hire rate of 4 percent and no testing of sales clerks]. An alternative approach which reduces the testing burden would have labor market intermediaries or testing organizations (eg. the Employment Service, private employment agencies, the Educational Testing Service) administer the battery of employment tests and then report the scores to potential employers when requested by the worker. Twenty seven percent of the work force change jobs in a year (Horvath 1981). If each job changer were to take 3 tests on average and one fifth of those with more than a years tenure were tested yearly as well, the total yearly costs of testing would be $2.3 billion [$34(.27*3+.73*.2)*71,132,000]. The projected social costs, therefore, probably lie somewhere between 2 and 10 percent of the projected social benefits.

The Distributional Effects of Resorting on the Basis of Test Scores

The simulated effect of the reassignment scheme on the mean test scores, schooling and demographic character of each occupation is presented in the even numbered columns of Table 6. The characteristics of those who are currently in each occupation are presented in the odd numbered columns. Currently workers in technical and high skill clerical occupations have the highest academic achievement and operatives and service workers have the lowest. The simulation results in the workers with the strongest academic achievement being reassigned to plant operator, technical and craft occupations and the workers with the weakest academic achievement being reassigned to operative and sales clerk occupations. Some of the changes are truly dramatic--the mean test score of plant operators rises by 2 population standard deviations and the mean score of sales clerks falls by 1.6 population standard deviations. This outcome is a result of placing the plant operator occupation at the top of the hierarchy and the sales clerk occupation at the bottom. The simulation also produces an increase in the schooling of plant operators and a decline in the mean schooling of sales clerks.

Reassigning workers on the basis of test scores, age and previous work experience but not gender or ethnicity produces large changes in the demographic composition of some occupations. Women end up with most (77 percent) of the plant operator jobs and roughly half of the craft jobs.

Occupations which have historically been predominantly female become more evenly split between men and women. As anticipated, black representation decreases in plant operator, technical, craft, clerical and service occupations and increases in operative and sales clerk occupations. Similar but more modest changes occur for Hispanics. Since, however, employers know the minority status of job applicants, the adverse impact on minorities of using tests to select employees can be eliminated by within-group scoring of the tests or by other affirmative action efforts.

Comparison with Hunter and Schmidt

How do these results compare to those of Hunter and Schmidt (1982)? The estimated total effect of going from random selection of new hires to optimal use of tests, age and previous work experience is 15 percent of the compensation of workers subject to reassignment. This is much larger than the 8 percent figure H/S obtain in their three test score selection model when SD$ is 40 percent of each occupation's mean compensation. The reasons for the difference are: (a) the estimates of differences in SDY across occupations are much larger than the one's assumed in their simulation, (b) the restriction of range correction (which was based on actual data on the reductions in job performance variance resulting from the selective nature of turnover) is larger than the one they assumed, (c) job assignment is based on a composite of test scores, schooling, age and previous occupational work experience that has greater validity than test scores alone and (d) 8 rather than 4 occupational categories are analyzed.

2.4 A Critique of the Simulations

The simulation results just presented are based on a maintained assumption that the models of relative job performance described in section 2.1 (which were estimated in samples of job incumbents) are, after the correction for errors in measurement of the criterion and the selective nature of turnover (ie. restriction of range), unbiased estimates of true population relationships. This assumption is almost certainly incorrect and this results in the findings of the simulation exercise being biased as well. The underlying performance model is biased for two reasons: omitted variables and the selection process that determines which members of the population are hired for the job.

While equation 2 and 3 are more complete specifications of the background determinants of job performance than is typically found in the literature, they lack controls for important characteristics of the worker which are often known by hiring decision makers and which are associated with worker productivity. Examples of things left out of the model are occupationally specific schooling, grades in relevant subjects in school, reputation of the school, the amount and quality of previous on-the-job training, performance in previous jobs, interview performance, physical strength and a desire to work in the occupation. Quite clearly, if random assignment of new hires to jobs involves ignoring all of this additional information as well as information on schooling and years of experience in the occupation, the loss in productivity would be substantially larger than the numbers reported in table 4.

The omission of so many important determinants of job performance also biases the simulations of the impact of greater test use. If these variables had been included in the job performance models, the coefficients on test scores would probably have been smaller and adding test scores to the factors considered in hiring selections would have resulted in fewer workers being reassigned. This in turn reduces the output gain that results from greater use of employment tests for selection and exaggerates the predicted changes in demographic composition of occupational work forces.

The other source of problems is selection effects. The selectivity bias caused by turnover and promotion decisions that depend on realized levels of job performance has already been discussed and corrected for. Another form of selectivity bias is introduced by the selection that precedes the hiring decision. If hiring selections were based entirely on X variables included in the model, unstandardized coefficients such as $\hat{\beta}$ would be unbiased and correction formulas would be available for calculating standardized coefficients and validities. Unfortunately, however, incidental selection based on unobservables such as interview performance and recommendations is very probable (Thorndike 1949; Olson and Becker 1983; Mueser and Maloney 1987). In a selected sample like accepted job applicants, one cannot argue that these omitted unobservable variables are uncorrelated with the included variables that were used to make initial hiring decisions and, therefore, that coefficients on included variables are unbiased. When someone with 10 years of formal schooling is hired for a job that normally requires an associates degree, there is probably a reason for that decision. The employer saw something positive in that job applicant (maybe the applicant received a particularly strong recommendation from previous employers) that led to the decision to make an exception to the rule that new hires should have an associates degree. The analyst is unaware of the positive recommendations, does not include them in the job performance model and, as a result, the coefficient on schooling is biased toward zero. This phenomenon also causes the estimated effects of other worker traits used to select workers for the job such as previous relevant work experience to be biased toward zero. Variables which were not used to select new hires such as the GATB test scores will probably have a positive correlation with the unobservable. Since the unobservable probably has its own independent effect on job performance (ie it is not serving solely as a proxy for test scores), test score coefficients are likely to be positively biased. Mueser and Maloney (1987) experimented with some plausible assumptions regarding this selection process and concluded that coefficients on education were severely biased but that test validities were not substantially changed when these incidental selection effects are taken into account.

Consequently, the estimates of the effects of greater use of the GATB presented in Table 5 probably exaggerate its true effect. If the simulations had been conducted using the true structural model of job performance rather than the biased one that was available, fewer people would have been reassigned and productivity gains would have been smaller. Still another problem with the simulations is that they took no account of turnover risks. The large effects of tenure on the productivity of plant operators, technicians and craft workers implies that specific training is particulary important in these occupations and that minimizing turnover should be an important goal of a firm's hiring selections. Some of the workers assigned to plant operator jobs in the simulation might have been college students working part time who would have been unlikely to remain long in the job.

Greater use of employment tests is not the same thing as greater use of the GATB. The GATB lacks measures of technical, scientific and advanced mathematical competence and is, therefore, not the best employment test available. If these subtests were added to the GATB there would be a substantial increase in validity and classification efficiency (eg. workers with a strong technical background would be assigned to craft jobs rather than clerical jobs and workers strong in math and English but weak in the technical arena would be assigned to clerical jobs). If a fully optimal sorting routine had reassigned workers across 100 occupations on the basis of a test battery with separate verbal, mathematical and technical ability as well a perceptual speed and psychomotor ability, the sorting efficiency gains would have been larger than those simulated. These abilities are not all that highly correlated and studies of the classification problem in the military find that important increases in utility result when recruits are optimally assigned to jobs on the basis of a test battery like the ASVAB.

On the other hand, Mike Rothschild is correct when he argues that there are many barriers to the complete reshuffling of the work force that would be necessary for employment testing to have its maximum effect (the effect that is simulated in Table 5). Employers would have to become much better informed about employment testing. If they all sought advice from industrial psychologists, long queues would result and consulting fees would skyrocket. If a number of worker aptitudes are to be reliably measured, a couple of hours must be devoted to the testing. This would impose a burden on job seekers in some high turnover labor markets and some low wage industries would, consequently, eschew testing altogether. The simulation model did not ask the workers who were being transferred whether they wanted the higher paying jobs. Some would have refused. The simulation ends gender segregation of occupations and makes wholesale transfers of clerical workers to plant operator and craft jobs. Improved structural models would probably reduce the size of these shifts, but even more modest shifts would be difficult to pull off. Affirmative action goals and/or the use of race normed test scores in selection would also reduce the sorting impacts of greater test use. Clearly, the EEOC regulation of employment testing is not the only barrier to a more efficient allocation of workers across jobs and many of these other barriers would have to fall before testing could have its full effect. Consequently, the likely productivity benefits and resorting effects of allowing employers a free hand with regard to employment testing are smaller than those presented in Table 5.

The simulated effects of substituting random selection of new hires for the current job-worker matching system reported in Table 4 are, by contrast, gross underestimates of the true costs. The selected nature of the sample and the many variables omitted from the "structural" models of job performance, cause very large biases in these simulations. Depending on how far one goes down the road toward random selection, the loss in sorting efficiency might be 2 or even 4 times those estimated.[10] Rates of involuntary separation would increase and this would increase unemployment and waste investments in specific training. In addition, economic incentives to go to school and study hard would be greatly reduced and this would cause further reductions in total output and standards of living. These results suggest that the current system of matching workers to jobs which makes almost no use of tests (tests were given prior to hiring in only 3 percent of hiring events sampled in the NFIB study) is not doing all that bad a job. This conclusion would appear to contrast somewhat with Hunter and Schmidt's (1983) characterization of current selection processes quoted at the beginning of part 2.

On the more important issue of how increased employment testing will effect national output, there is no disagreement with Hunter and Schmidt. The simulations imply that the improvements in the matching of workers to jobs resulting from increased employment testing will significantly increase output. The 6.9 percent figure might fall to 2 or 3 percent of employee compensation once one takes the biases and the barriers to optimal use of tests into account. On the other hand, taking constraints off the use of tests will also reduce tryout hiring and turnover and increase investment in specific human capital. These effects were not part of the simulations. Since total compensation of labor will exceed $3 trillion in 1988, applying the 2 to 3 percent estimate to the nation's entire workforce implies that the productivity gain from unconstrained employment testing would eventually increase gross national product by 60 to 90 billion dollars per year or between 1 and 2 percent of GNP. These effects would not arrive suddenly for the tests only influence hiring decisions. Current employees would not be fired and replaced by new hires selected on the basis of tests because the gains from better selection will seldom be sufficient to justify firing employees who have developed firm specific skills. It would, therefore, be a decade before the full effect of testing on the allocation of workers to jobs would be realized.

The $60 to $90 billion per year estimate is clearly a guess. A better estimate of the effect of greater test use on sorting efficiency requires better estimates of SDY, a better understanding of the magnitude and nature of the biases in job performance models, a model of employment testing's effects on turnover, investments in specific training and unemployment and above all an understanding of how employers would use tests if they were given the opportunity. Clearly, much more research is needed on these topics.

PART III. POLICY RECOMMENDATIONS

The findings presented in the first two parts of the paper imply that improved signaling of worker skills and competencies to employers will probably have significant positive effects on productivity and standards of living. Productivity gains occur both because more valid selection procedures improve the match between workers and jobs and because the supply of workers with the talents measured by the tests or school examinations grows in response to the increase in labor market rewards for the talents. The distributional consequences of greater use of academic achievement for selecting workers are that the better jobs will go to those who studied hard in school and those who attend schools that have good teachers and maintain high standards. Women will gain more access to high paying occupations but the representation of Blacks and Hispanics in occupations where the payoff to cognitive skills is high such as plant operator, craft worker and technician will fall.[11] Adverse impacts on blacks and Hispanics can be avoided by race norming the test scores (as the GATB currently does) and affirmative action. Consequently, impacts on minority groups should not be the basis for deciding whether to use an employment test or which test to use. Other instruments are available for achieving employer and societal goals regarding integration on the job and the representativeness of a firm's workforce. When, however, it comes to generating incentives to develop the skills needed on the job and efficient matching of workers with talents to jobs, there appears to be no other selection instrument that will sort efficiently while generating the correct incentives quite as well as measures of academic achievement. These are the two criteria--incentives and sorting efficiency--by which alternative employee selection policies should be evaluated. That is the task undertaken in the remainder of the paper.

Sorting efficiency will tend to be maximized when the employment tests used in selection for a particular occupation measure developed abilities which have a uniquely high productivity payoff in that occupation (eg. mechanical comprehension for maintenance and repair occupation). In other words, selection/classification protocols should attempt to assign workers to occupations in which they have a comparative advantage. Tests should be used but they should supplement not displace consideration of other factors such as personality, physical strength and occupationally relevant training and experience. If most of the people hired into an entry job move up to other more responsible positions, the criteria applied at the port of entry needs to take the higher level jobs into account.

The analysis presented in the first part of the paper implies that student incentives to learn and parental incentives to demand a quality education are maximized when the following is true: (1) significant economic rewards depend directly and visibly on academic accomplishments, (2) the accomplishment is defined relative to an externally imposed standard of achievement and not relative to one's classmates, (3) the reward is received immediately, (4) everyone, including those who begin high school with serious academic deficiencies, has an achievable goal which will generate a significant reward and (5) progress toward the goal can be monitored by the student, parents and teacher.

We will see shortly that it is not easy to design a system of signaling and certifying academic achievement which satisfies all of these requirements. Consequently, it will generally be desireable to use more than one signal of academic achievement and to use different signals when selecting for different jobs. Let us examine the alternatives.

Diplomas:

High school diplomas and college degrees are effective devices for generating incentives to enroll in school. The standard diploma does not, however, generate incentives to attend regularly or to study hard and thus it fails requirement # 1, the most critical requirement of all. Establishing a minimum competency level for receiving a high school diploma only slightly improves incentives. Some students arrive in high school so far behind and the consequences of not getting a diploma are so severe, minimum competency standards are not set very high (and cannot in good conscience be set too high given the constraints on the system). Once they satisfy the minimum, many students stop putting effort into their academic courses.

Schooling is a valid predictor of job performance but to a great degree its validity derives from its correlation with test scores. The evidence on its incremental contribution to validity once test scores are controlled is more mixed. An analysis of GATB revalidation data by Bishop (1987b) found very weak effects of schooling but this is probably an artifact of the selection biases discussed earlier (Mueser and Maloney 1988). Selection into the military is based explicitly on the test scores and high school graduation, not on unobservables as in the civilian sector. Since selection is based on X variables, selection effects can be corrected for (Dunbar and Linn 1986). Analysis of military data finds that high school graduation has its own unique impacts when test scores are controlled. Weiss's (1985) study of Western Electric employees found that completing high school is a valid predictor of low absenteeism and low turnover but not job performance. Thus even when studies find that graduating from high school has little effect on job performance, it appears to effect retention. Consequently, from a sorting efficiency point of view, the high school diploma belongs on the list of credentials considered by employers even when test scores are available.

Competency Profiles:

Competency profiles are check lists of competencies that a student has developed through study and practice. The ratings of competence that appear on a competency profile are relative to an absolute standard, not relative to other students in the class. By evaluating students against an absolute standard, the competency profile prevents one student's effort from negatively affecting the grades received by other students. It encourages students to share their knowledge and teach each other.

A second advantage of the competency profile approach to evaluation is that students can see their progress as new skills are learned and checked off. The skills not yet checked off are the learning goals for the future. Seeing such a check list getting filled up is inherently reinforcing.

With a competency profile system, goals can be tailored to the student's interests and capabilities, and progress toward these goals can be monitored and rewarded. Students who have difficulty in their required academic subjects can, nevertheless, take pride in the occupational competencies that they are developing and which are now recognized just as prominently as course grades in academic subjects. Upon graduation, the competency profile is encased in plastic and serves as a credential certifying occupational competencies. If the ratings by teachers (and the sponsoring employers of cooperative education students) are reliable indicators of competence, employers will find this information very valuable, and the students who build a good record will benefit.

A great many vocational programs currently use competency profiles both to structure instruction and as a system for articulating with the labor market and further training. Unfortunately, however, most schools do not view mailing out profiles to prospective employers as part of their responsibility. There is a great deal of geographic variation in the format of these documents, the skills and competencies that are assessed and the competency standards used. In many cases only occupational skills are assessed by the profile. These problems make it more difficult for employers to use these profiles and reduces their ability to aid a student's job search. Some thought needs to be given to how to include more generic competencies such as numeracy and writing in these profiles, how some standardization can be achieved and how they can be made more accessible and useful to employers.

Hiring Based on Grades in High School:

Using grades to select new hires results in a very visible dependence of labor market outcomes on an indicator of academic accomplishment. There are, however, two disadvantages. It results in zero-sum competition between classmates and consequently contributes to peer pressure against studying and parental apathy about the quality of teaching and the rigor of the curriculum. The second problem is that it induces students to select easy courses and thus tends to cause grade inflation. These problems can be mitigated somewhat if employers take the rigor of courses into account when evaluating grades, give preference to schools with tough grading standards, and vary the number hired from particular schools in response to the actual job performance of past hires from that school.

From the sorting point of view, the disadvantage of high school GPA is that it has low validity when there are no adjustments made for grading standards and it is difficult for employers to make such adjustments.[12]

Job Tryout and Promotions Based on Performance:

From the point of view of motivating students to study, the problem with job tryout and performance reward systems is that the dependence of labor market outcomes on academic achievements is both invisible and considerably delayed.

From the efficiency point of view, the disadvantages of job tryout are the costs of training workers who end up being fired, its unpopularity with workers who will spend months unemployed if they are fired, and its potential for generating grievances.[13] Performance evaluations are known to be unreliable, and this makes workers reluctant to take jobs in which next year's pay is highly contingent on one supervisor's opinion. Pay that is highly contingent on performance can also weaken cooperation and generate incentives to sabotage others. The benefits of performance reward systems are that they motivate better performance, they tend to attract high performers to the firm, and they tend to induce the high performers to stay at the firm. When these factors are balanced, it appears that most workers and employers choose compensation schemes in which differentials in relative productivity result in relatively small wage differentials (Bishop 1987a).

Job Knowledge Tests:

From the point of view of learning incentives, the disadvantage of job knowledge tests is that they generate no incentives to study history and literature and generate incentives to study math and science only occasionally (i.e. when the student expects to seek a technical job and the job knowledge tests for the job contains math and science questions relevant to the job). They may also induce students to over-specialize in school. If at some point in their career a job in the field for which they prepared is not available, they are left high and dry.

From the point of view of sorting efficiency, job knowledge tests have much to recommend them for they maximize classification efficiency--the assignment of job seekers to jobs which make use of already acquired skills. They are particularly appropriate if applicants vary in their knowledge and background in the occupation and training costs are substantial. If new hires are likely to be quickly promoted into higher level jobs, the job knowledge test should also cover the skills required in these jobs. Job knowledge tests are less useful when none of the applicants has experience in the field and training costs are low.

IQ Tests:

Students, parents and teachers view IQ tests as measuring something that schools do not teach. Even though this public perception is not entirely correct, the perception is not likely to change in the near future, so hiring on the basis of IQ tests fails requirement # 1. Students will not see the connection between how hard they study and higher IQ scores.

General Aptitude Test Battery (GATB):

The cognitive subtests of the current GATB measure only a limited number of very basic skills--vocabulary, reading, arithmetic computation and reasoning. There are no sub-tests measuring achievement in most of the subjects in the standard high school curriculum--science, history, social science, algebra, high school geometry or trigonometry. Greater use of the GATB to make hiring selections would strengthen incentives to learn arithmetic and English but would not strengthen incentives to study other high school subjects. Consequently, hiring on the basis of the GATB fails requirement # 1.

On the other hand, a large body of research suggests that the cognitive subtests of the GATB are valid predictors of job performance in many private sector jobs (Hunter 1983). The results of our analysis suggest that greater use of the GATB in selection decisions would yield substantial sorting efficiency gains. We will see shortly, however, that other selection methods--broad spectrum achievement test batteries and state sponsored exams assessing the student's mastery of the high school curriculum--are able to achieve at least as efficient sorting outcomes as the GATB and generate much better incentive effects.

Broad Spectrum Achievement Tests Batteries:

From the point of view of incentives to study a broad range of academic subjects, broad spectrum achievement test batteries such as the ASVAB are the best of the alternatives reviewed so far. If some of the subtests in the battery include material covered in the standard college prep high school curriculum such as algebra, statistics, chemistry, physics and computers, the use of such tests for selection would generate parental pressure for an upgraded curriculum and encourage high school students to take more rigorous courses. When many employers use achievement tests to select new employees, everyone who wants a good job faces a strong incentive to study, and those not planning to go to college will find the incentive especially strong. The best paying firms will find they can set higher test score cutoffs than low paying firms, so the reward for learning will become continuous. Whether one begins 9th grade way behind or way ahead, there will be a benefit on the margin to studying hard for it will improve one's job prospects.

Broad spectrum achievement test batteries covering science, computers, mechanical principles, economics, business practices and technology as well as mathematics, reading and vocabulary also maximize sorting benefits as well. Test batteries which cover the full spectrum of knowledge and skills taught in high school are more valid predictors of job performance than tests which assess math and verbal skills only. Evidence for this statement comes from examining the relative contributions of various subtests to the total validity of the ASVAB battery. Maier and Grafton's (1981) analysis of hands-on measures of the job performance for Marine Corps recruits found, for example, that validity (corrected for restriction of range) was .46 for auto shop information, .50 for mechanical comprehension, .51 for electronics information, .51 for general science, .50 for word knowledge, .52 for mathematics knowledge, and .51 for arithmetic reasoning. Tests measuring electronics, mechanical, automotive and shop knowledge--material that is generally studied only in vocational courses--have high validity. Analyzing this and other military data sets, Hunter, Crosson and Friedman (1985) concluded that the "general cognitive ability" construct that best predicted performance in all military jobs included subtests in general science, electronics information, mechanical comprehension and mathematics knowledge as well as conventional word knowledge and arithmetic reasoning subtests. The addition of these four subtests to the construct increased validity by 11 percent and the proportion of true job performance variance explained in the Maier and Grafton data from .306 to .372 (Hunter, Crosson and Friedman, 1985, Table 19).

Broad spectrum achievement test batteries also improve classification efficiency. The technical subtests of ASVAB are important predictors of hands-on measures of job performance in technical and maintenance jobs but did not contribute to the prediction of performance in clerical jobs. Verbal subtests contributed to clerical performance but did not correlate with performance in many of the other jobs in the study. Tests measuring understanding of computers, business, economics, marketing and psychology would probably similarly improve the validity of batteries used to select workers for most white collar jobs in the private sector. The conclusion that follows from this analysis is that, on both sorting and incentive grounds, broad spectrum achievement test batteries are better devices for selecting workers than the cognitive subtests of the GATB.

Will the courts allow firms to use broad spectrum achievement tests covering subjects not offered until the final years of high school? My fear is that, since the research on test validity in the civilian sector has used the GATB almost exclusively, everyone may be forced to use reading, vocabulary, and arithmetic reasoning tests that are demonstrably similar to their GATB counterparts. Courts might require that employers demonstrate that each item on a science test have a specific application in each job for which it is a proposed selection device. To avoid having to redesign the test for each job, test developers would dumb the test down and include only simple questions covering scientific principles that are learned in grade school. Litigation costs and the potential liability are enormous so companies are extremely cautious about testing. When choosing an employment test, defensibility in court is a much more important criterion than maximum validity. Given the uncertainty of whether ASVAB research will be accepted as evidence on the validity of similar tests for civilian jobs, broad spectrum achievement test batteries will probably be judged too risky. A well designed validity study can protect a firm using an unconventional test battery but in most cases the potential benefit of finding a more valid selection method will not outweigh the costs of the study and the greater risks of litigation. The fear of litigation has significantly inhibited testing research outside of government. Companies seldom share the results of their validity studies or allow them to be published (even when the company's name is withheld) for fear of revealing their defense strategy to a potential litigant. If things are left as they are, it will be at minimum a decade before tests measuring competence in algebra, science and the technical arena can be used as general selection devices for craft and other blue collar jobs. Firms need to be given a signal by the EEOC that broad spectrum achievement test batteries are acceptable selection devices and in fact preferred over the low level basic skills test that serves as the "g" aptitude of the GATB.

To speed the transition to broad spectrum achievement test batteries, the GATB (which has not changed appreciably since 1950) should be revised. Subtests similar to the technical, mathematical knowledge and science subtests of the ASVAB should be added and the SATBs revised to reflect military research. The employment service should also undertake a major study of the validity of the new GATB in the full spectrum of civilian jobs and undertake to develop subtests assessing knowledge of business, marketing and computers.

To maximize the incentive effects, it is essential that students, parents and teachers be aware that local employers are using tests for selection and what kind of material is included on these tests. Employers should seek out ways of publicizing their use of broad spectrum achievement tests. Unfortunately, the fear of litigation may cause employers to give only limited publicity to their use of tests and so constrain the type of tests that are used that many of the potential beneficial incentive effects of employment testing may never be realized.

Performance on Achievement Exams Taken at the End of Secondary School

In Canada, Australia, Japan and most European countries, the educational system administers achievement test batteries (eg. the 'O' and 'A' Levels in the UK, the Baccalaureate in France) which are closely tied to the curriculum. While the Japanese use a multiple choice exam, all other nations use extended answer examinations in which students write essays and show their work for mathematics problems. Generally, regional or national boards set the exam and oversee the blind grading of the exams by committees of teachers.[14] These are not minimum competency exams. Excellence is recognized as well as competence. In France, for example, students who pass the Baccalaureate may receive a "Très Bien", a "Bien", an "Assez Bien" or just a plain pass. These exams generate credentials which signal academic achievement to all employers and not just the employers who choose to give employment tests. The connection between one's effort in school and performance on these exams is clearly visible to all. Consequently, school sponsored achievement exams like those used in Europe would have much stronger incentive effects than employer administered broad spectrum achievement tests.

This approach to signaling academic achievement has a number of advantages. Because it is centralized and students take the exam only once, job applicants do not have to take a different exam at each firm they apply to and the quality and comprehensiveness of the test can be much greater.

There is no need for multiple versions of the same test and it is much easier to keep the test secure. By retaining control of exam content, educators and the public influence the kinds of academic achievement that are rewarded by the labor market. Societal decisions regarding the curriculum (eg. all students should read Shakespeare's plays and understand the Constitution) tend to be reinforced by employer hiring decisions. Tests developed solely for employee selection purposes would probably place less emphasis on Shakespeare and the Constitution.

The disadvantages of schools administering the achievement exams is that students have fewer chances to demonstrate their competence. If one has an off day, one must typically wait an entire year before the exam can be retaken (in Finland the delay is a few months and retaking the exam is very common). With employer administered exams, having an off day is less damaging for one will shortly have a chance to do better at another employer. Employers may also find it is easier to compare job applicants who have all taken the same employer administered exam.

With regard to validity, there is probably little difference between the two systems. Scores are reported for each subject so employers may focus on the tests which have special relevance to their jobs. School administered tests are more reliable measures of achievement because they sample a much larger portion of the student's knowledge of the field (the ASVAB General Science subtest, by contrast, allows the student 11 minutes to do 24 items). They may also be more valid because they are not limited to the multiple choice format. Thus, even though the topics covered in the school exam are probably less relevant to the firm's jobs, it is probably just as valid a predictor of performance as a specially designed employment test.

It would, therefore, appear desireable for American schools to sponsor tests of competency and knowledge that are specific to the curriculum being studied (e.g. New York State's Regents Examinations, NOCTI's Student Occupational Competency Achievement Tests) and then to provide students with competency profiles certifying capabilities. State Departments of Education are logical sponsors of such a testing and certification program but they are not the only possible sponsor. Testing organizations (eg. the Educational Testing Service) or a new joint educator/employer organization could also sponsor and administer such a program.

NOTES

1. Teenagers can expect much higher levels of income in the future but they are not able to consume at the rate implied by their expected lifetime income because they are unable to borrow against this future income at reasonable rates of interest while they are in high school. They are liquidity constrained (Hubbard and Judd, 1987). This results in the youth placing a much higher value on free time this week than free time 10 years from now. Investments in college are undertaken, nevertheless, because parents and society subsidize the investment, loans are available to relieve liquidity constraints, college life is enjoyable and prestige is derived from attendance.

2. Studies that measure output for different workers in the same job at the same firm, using physical output as a criterion, have found that the standard deviation of output varies with job complexity and averages about .164 in routine clerical jobs and .278 in clerical jobs with decision making responsibilities (Hunter, Schmidt & Judiesch 1988). Because there are fixed costs to employing an individual (facilities, equipment, light, heat and overhead functions such as hiring and payrolling), the coefficient of variation of marginal products of individuals is assumed to be 1.5 times the coefficient of variation of productivity. Because about 2/3rds of clerical jobs can be classified as routine, the coefficient of variation of marginal productivity for clerical jobs is 30 % [1.5*(.33*.278+.67*.164)]. A .5 validity for general mental ability then implies that an academic achievement differential between two individuals of one standard deviation (in a distribution of high school graduates) is associated with a productivity differential in the job of about 11 % (.5*.74*30%). The ratio of the high school graduate test score standard deviation to the population standard deviation is assumed to be .74. This issue is more thoroughly discussed in Bishop (1987b, 1988b).

3. The survey was of a stratified random sample of the NFIB membership. Larger firms had a significantly higher probability of being selected for the study. The response rate to the mail survey was 20 percent and the number of usable responses was 2014.

4. The formula was $SD(R^m_j) = (R^m_{ij} - R^m_j)^2/N-1$. Occasionally employers who had only 2 or 3 employees gave them all the same rating. Consequently, a lower bound of 40 percent of the mean $SD(R^m_j)$ was placed on the value the SD could take. Models were also estimated which did not standardize job performance variance across firms and which instead standardized the variances only across the occupation. None of the substantive findings were changed by this alternative methodology.

5. Mueser and Maloney (1988) argue persuasively that since schooling is a very important factor in the selection process, the coefficients on schooling in estimations like these are negatively biased estimates of true population relationships. This argument probably applies as well to the coefficients on work experience in the occupation but not at the firm.

6. Large as it may seem the estimate for operators of nuclear plants is in fact quite reasonable. In the first 4000 years of world wide operation of nuclear plants there have been two catastrophic accidents caused by operator error each costing over 5 billion dollars. The NRC estimates that improved safety procedures will reduce operator caused catastrophic accidents to about one fifth that rate (one in every 10000 years of plant operation). There are about 5 six person shifts operating each plant, so the standard deviation of output across individual workers that results from just this one risk is about $9 million per year.

7. This estimate of the variance ratio is probably too large for two reasons. First, selective turnover has been operating for only a year. Second, workers who were promoted to better jobs were retained in the calculation not dropped. If a longer period were analyzed and workers had been dropped from the sample when they were promoted, a lower variance ratio would have been obtained and all estimates of sorting effects would have increased proportionately. On the other hand, large establishments were under represented in the study. Since they tend to have less selective turnover than small establishments, this produces a small bias in the opposite direction.

8. This hierarchical process for allocating new hires to jobs is not fully optimal. Some workers will not be assigned to the occupation in which they have the greatest comparative advantage. A computer program that assigns all new hires optimally would be much more complex and the task has been left for another paper.

9. If the SD$ for retail clerks had been calculated by multiplying the CV by 1.52 as for other occupations, sales clerks would have been placed above operatives and service workers in the hierarchy and these two occupations not sales clerks would have been assigned the lowest scoring students. This would significantly reduce the productivity decline among sales clerks but produce a substantial decline in the productivity of service workers and increase the decline in the productivity of operatives. The total change in productivity for the economy as a whole from resorting would not be very different, however.

10. The legal theories that have been used to attack employment tests on EEO grounds are equally applicable to other selection criteria. If the theory of differential validity by subgroup and employer were applied to selection criteria like years of schooling, school reputation, GPA and recommendations from previous employers (all of which have adverse impact), these criteria would probably fail court tests for jobs like those in the GATB Revalidation data. If the 1970s trend of court decisions restricting employer prerogatives to select the "best" job applicant had continued rather than being reversed, we might have moved a considerable distance down the road toward random selection of new hires for these jobs. Sandra Day OConnor's concurring opinion in Watson (1988) signals a major shift in the application of the Griggs adverse impact test, so the trend now seems to be in the direction of greater freedom for the employer.

11. This adverse impact results not because tests are unfair but because academic achievement contributes to worker productivity and because there are, unfortunately, real differences in mean levels of academic achievement between groups (Jones 1988). The tests are giving us the unhappy news that educational opportunities and achievement have not been equalized. The cause of the situation is the low quality of the education received by most Blacks and Hispanics. Progress has been made in reducing these quality differentials and achievement gaps are diminishing. This means the problem will diminish over time. If the process of closing the gap is to be speeded there needs to be increased investment in both regular and adult basic education.

12. Most of the published studies of the validity of grades probably used information that had been collected by the firm when hiring decisions were being made. Consequently, most of the validity coefficients reported for grades are probably negatively biased by the selection effects so the true validity of GPA than is generally thought.

13. Mueser and Maloney (1987) develop a model of job tryout hiring which they claim implies that it may be efficient to ignore available information on stable worker competencies signaled by high test scores. They apparently do not recognize that the model also implies that information on education and previous work experience should also be ignored. They acknowledge that "Although employing applicants for long enough to observe performance entails costs of training and lost productivity, it may increase the incentives workers have to apply effort to learning their jobs by enough to compensate for such costs." In fact, however, turnover costs are so large--training costs are generally about one month's wages and fired workers suffer a couple of months of unemployment--, that a sequential decision strategy will always dominate the strategy they consider. It will hardly ever be optimal to hire ten people for one position and then fire 9 of them after a tryout. In any job requiring even a modest amount of specific training or transitional unemployment, the optimal strategy is to use all the inexpensive information available to make an initial selection and then to give those selected a tryout but to plan on seldom having to fire the new employee. It is true, however, that the option of firing the worst performers results in Brogden's formula overstating the private benefits of a selection method.

14. Germany is somewhat exceptional in giving the teacher some influence over the questions asked and their grading. Ingenkamp (1969) has described the system. "The actual responsibility for setting the questions varies in the different Federal States. In some the subjects are set by a central committee for all schools; in others the Gymnasium submits suggestions, from which the representative of the State School Authority, who supervises the whole examination, chooses the subjects to be set. The assessment of the candidates work is likewise more or less centralized. Usually the examination work is scrutinized and marks assessed by the specialist teacher at the Gymnasium, then submitted to another specialist who acts as a co-assessor and to all the other teachers of the Abitur class for their opinion, and is finally sent to the representative of the State School Authority for confirmation.(p. 144)"

BIBLIOGRAPHY

Adler, Paul. "New Technologies, New Skills, California Management Review, Volume XXIX, Number 1, Fall 1986.

Bishop, John H. Preparing Youth for Employment. Columbus, Ohio: National Center for Research in Vocational Education, 1985.

Bishop, John H. "The Recognition and Reward of Employee Performance." Journal of Labor Economics. October, 1987a.

Bishop, John "Information Externalities and the Social Payoff to Academic Achievement." Center for Advanced Human Resource Studies Working Paper # 87-06, Cornell University, Ithaca, New York, 1987b.

Bishop, John H. "Match Quality, Turnover and Wage Growth". Center for Advanced Human Resources Studies Working Paper # 88-03, Cornell University, Ithaca, New York, 1988

Bishop, John "The Productivity Consequences of What is Learned in High School." Center for Advanced Human Resource Studies Working Paper # 88-18, Cornell University, Ithaca, New York, 1988b.

Bishop, John and Griffin, Kelly. Recruitment, Training and Skills of Small Business Employees, (National Federation of Independent Business Foundation, Washington, DC, forthcoming).

Boudreau, J. W. "Utility Analysis Applied to Human Resource Productivity Improvement Programs." 1986, To appear in M.D. Dunnette (Ed.) Handbook of Industrial and Organizational Psychology, (2nd edition).

Brogden, H. E. "When Testing Pays Off." Personnel Psychology, 1949, Vol. 2, pp. 133-154.

Brown, Charles. "Estimating the Determinants of Employee Performance." Journal of Human Resources, Spring 1982, Vol. XVII, No. 2, pp. 177-194.

BNA. "Selection Tests: Employer Interest Increasing." Bulletin to Management, January 2, 1986, p. 2, 7.

Corts, D. B., Muldrow, T. W. and Outerbridge, A. M. (1977). Research base for the written test portion of the Professional and Administrative Career Examination (PACE): Prediction of job performance for customs inspectors. Washington, D.C.: U.S. Office of Personnel Management.

Dumaret, A. (1985) "IQ, scholastic performance and behavior of sibs raised in contrasting environments." J. Child Psychol. Psychiatry. 26:553-80.

Department of Labor. General Aptitude Test Battery Manual 1970, United State Department of Labor, Manpower Administration.

Dunbar, Stephen B. and Linn, Robert L. "Range Restriction Adjustments in the Prediction of Military Job Performance". Prepared for Committee on the Performance of Military Personnel, National Research Council / National Academy of Sciences. September 1986.

Dunnette, Marvin D. Validity Study Results for Jobs Relevant to the Petroleum Refining Industry. Washington, DC: American Petroleum Institute, 1972.

Dunnette, Marvin D.; Rosse, Rodney L.; Houston, Janis S.; Hough, Leaetta M.; Toquam, Jody; Lammlein, Steven; King, Kraig W.; Bosshardt, Michael J. and Keyes, Margaret. "Development and Validation of an Industry-Wide Electric Power Plant Operator Selection System," Personnel Decisions Research Institute, Minneapolis, Minnesota, Report submitted to Edison Electric Institute, 1982.

Duyme, M. (1985) "Scholastic achievement as a function of parental social class: an adoption study." Developmental Psychology, ed. C. J. Brainero, V. F. Reyna, 319-25. Amsterdam: Elsevier Science.

Flynn, James R. "Massive IQ Gains in 14 Nations: What IQ Tests Really Measure." Psychological Bulletin, 1987. Vol. 101, No. 2, 171-191

Forsyth, Robert A. (personnel communication) "Achievement Trends for the Iowa Tests of Educational Development in Iowa: 1942-1985," Iowa City:Iowa Testing Programs, 1987.

Frank, Robert. "Are Workers Paid Their Marginal Product?" American Economic Review, Vol. 74:4, September 1984, 549-571.

Frederick, W.; Walberg, H.; and Rasher, S. "Time, Teacher Comments, and Achievement in Urban High Schools." Journal of Educational Research 73, no. 2 (November-December 1979): 63-65.

Freeman, Richard B. The Market for College-Trained Manpower: A Study in the Economics of Career Choice. Cambridge, MA: Harvard University Press, 1971.

Freeman, Richard. "A Cobweb Model of the Supply and Starting Salary of New Engineers." Industrial and Labor Relations Review, 29, January 1976a, pp. 236-248.

Freeman, Richard. The Overeducated American. New York: Academic Press, 1976.

Friedman, Toby and Williams, E. Belvin. "Current Use of Tests for Employment." Ability Testing: Uses, Consequences, and Controversies, Part II: Documentation Section, edited by Alexandra K. Wigdor and Wendell R. Gardner. Washington, DC: National Academy Press, 1982, pg. 999-169.

Gardner, John A. Influence of High School Curriculum on Determinants of Labor Market Experience. Columbus: The National Center for Research in Vocational Education, The Ohio State University, 1982.

Ghiselli, Edwin E. "The Validity of Aptitude Tests in Personnel Selection." Personnel Psychology. 1973: 26, 461-477.

Goldberger, Arthur S. "Linear Regression after Selection." Journal of Econometrics. Vol.15, 1981, pp. 357-366.

Goodlad, J. A Place Called School. New York: McGraw-Hill, 1984.

Griggs vs. Duke Power Company, 3 FEP 175 (1971).

Guion, R. M. Personnel Testing, New York: McGraw-Hill, 1985.

Hashimoto, M., and Yu, B. "Specific Capital, Employment and Wage Rigidity." Bell Journal of Economics, 11, no. 2, 1980: 536-549.

Hause, J. C. "Ability and Schooling as Determinants of Lifetime Earnings, or If You're So Smart, Why Aren't You Rich." In Education, Income, and Human Behavior, edited by F. T. Juster. New York: McGraw-Hill, 1975.

Hauser, Robert M. and Daymont, Thomas M. "Schooling, Ability, and Earnings: Cross-Sectional Evidence 8-14 years after High School Graduation." Sociology of Education, July 1977, 50, 182-206.

Heyns, Barbara. "Schooling and Cognitive Development: Is There a Season for Learning." Child Development, Vol. 58, No. 5, 1987, pp. 1151-1160.

Hirschhorn, Larry. Beyond Mechanization. Cambridge, Mass.: MIT Press, 1984.

Hollenbeck, K., and Smith B. The Influence of Applicants' Education and Skills on Employability Assessments by Employers. Columbus: The National Center for Research in Vocational Education, The Ohio State University, 1984.

Horvath, Francis W. "Job Tenure of Workers in January 1981." Monthly Labor Review, Sept, 1981, p. 23-32.

Hotchkiss, Lawrence. Effects of Schooling on Cognitive, Attitudinal and Behavioral Outcomes. Columbus: The National Center for Research in Vocational Education, The Ohio State University, 1984.

Hotchkiss, Lawrence; Bishop, John H. and Gardner, John. Effects of Individual and School Characteristics on Part-time Work of High School Seniors. Columbus: The National Center for Research in Vocational Education, The Ohio State University, 1982.

Hubbard, R. Glenn and Judd, Kenneth L. "Liquidity Constraints, Fiscal Policy and Consumption." Brookings Papers on Economic Activity, 1986, # 1, pp. 1-60.

Hunter, John Test Validation for 12,000 Jobs: An Application of Job Classification and Validity Generalization Analysis to the General Aptitude Test Battery. Washington, DC: US Employment Service, Department of Labor, 1983.

Hunter, John. "Cognitive Ability, Cognitive Aptitudes, job Knowledge and Job Performance." Journal of Vocational Behavior, Vol. 29, No. 3, December 1986. pp. 340-362.

Hunter, J. E. and Hunter, R. F. "The Validity and Utility of Alternative Predictors of Job Performance." Psychological Bulletin, 96, no. 1, 1984, pp. 72-98.

Hunter, John and Schmidt, Frank. "Fitting People to Jobs", Human Performance and Productivity, edited by Marvin Dunnette and Edwin Fleishman, Hilsdale, NJ: Erlbaum, 1982, pp. 258-271.

Hunter, John E.; Crosson, James J. and Friedman, David H. "The Validity of the Armed Services Vocational Aptitude Battery (ASVAB) For Civilian and Military Job Performance," Department of Defense, Washington, D.C., August, 1985.

Hunter, John E.; Schmidt, Frank L. and Judiesch, Michael K. "Individual Differences in Output as a Function of Job Complexity." Department of Industrial Relations and Human Resources, University of Iowa, June 1988.

Husen, Torsten. "The Influence of Schooling Upon IQ." Theoria, 1951.

Husen, Torsten. Talent Opportunity and Career, (Almquist and Wiksell, Uppsala, 1969).

Ingenkamp, Karlheinz. "West Germany." Examinations, edited by Joseph Lauwerys and David Scanlon. The World Year Book of Education 1969, New York: Harcourt, Brace and World, 1969. pp. 140-145.

International Association for the Evaluation of Educational Achievement (IEA) The Underachieving Curriculum, Assessing U.S. School Mathematics From an International Perspective. Champaign, IL: Stipes Publishing Company, 1987.

International Association for the Evaluation of Educational Achievement (IEA) Science Achievement in Seventeen Nations, New York, Pergammon Press, 1988.

Jencks, Christopher and Crouse, James. "Aptitude vs. Achievement: Should We Replace the SAT?" The Public Interest, 1982.

Jencks, Christopher et.al. Who Gets Ahead? The Determinants of Economic Success in America. Basic Books, Inc., Publishers, New York, 1979.

Jones, Lyle (paper for Gifford Commission)

Kang, S. and Bishop, J. (1984) "The impact of curriculum on the non-college bound youth's labor market outcomes." In High School Preparation for Employment, 95-135. Columbus: The National Center for Research in Vocational Education, The Ohio State University.

King, L. M.; Hunter, J. W., and Schmidt, F. L. "Halo in a Multidimensional Forced Choice Performance Evaluation Scale." Journal of Applied Psychology, 65, 1980, 507-516.

Klein, Roger; Spady, Richard; and Weiss, Andrew. Factors Affecting the Output and Quit Propensities of Production Workers. New York: Bell Laboratories and Columbia University, 1983.

Koretz, Daniel, et. al. Trends in Educational Achievement. Washington: Congressional Budget Office, 1986.

Lazear, Edward P. "Pay Equality and Industrial Politics", The Hoover Institution, Stanford University, April 1986.

Leestma, Robert, et. al. Japanese Education Today. A report from the U.S. Study of Education in Japan prepared by a special task force of the OERI Japan Study Team, 1987.

McDowell, Douglas and Dodge, Garen. "Employee Selection and Title VII of the Civil Rights Act: The Legal Debate Surrounding Selection Criteria, Validation and Affirmative Action." Paper prepared for the National Commission on Testing and Public Policy, November 1988.

Maier, Milton H. and Grafton, Frances C. Aptitude Composites for ASVAB 8, 9 and 10. Research Report 1308. U.S. Army Research Institute for Behavioral and Social Sciences, Alexandria, Va, May 1981.

Meyer, R. "Job Training in the Schools." In Job Training for Youth, edited by R. Taylor, H. Rosen, and F. Pratzner. Columbus: The National Center for Research in Vocational Education, The Ohio State University, 1982.

Mueser, Peter and Maloney, Tim. "Cognitive Ability, Human Capital and Employer Screening: Reconciling Labor Market Behavior with Studies of Employee Productivity." Department of Economics, U. of Missouri-Columbia, June 1987.

National Academy of Sciences Committee on Ability Testing, Ability Testing: Uses, Consequences and Controversies. Part 1: Report of the Committee. Edited by Alexandra K. Wigdor and Wendell R. Gerner, Washington, DC: National Academy Press, 1982.

National Assessment of Educational Progress. The Reading Report Card. Princeton, New Jersey: Educational Testing Service, 1985.

National Assessment of Educational Progress. Who Reads Best?. Princeton, New Jersey: Educational Testing Service, 1988.

National Assessment of Educational Progress. The Mathematics Report Card. Princeton, New Jersey: Educational Testing Service, 1988.

National Assessment of Educational Progress. Crossroads in American Education. Princeton, New Jersey: Educational Testing Service, 1989.

Nielsen, A. C., Company (1987) Unpublished raw data.

Olneck, M. "On the Use of Sibling Data to Estimate the Effects of Family Background, Cognitive Skills, and Schooling: Results from the Kalamazoo Brothers Study." In Kinometrics: The Determinants of Socio-economic Success Within and Between Families, edited by P. Taubman. Amsterdam: North Holland, 1977.

Olson, Craig A and Becker, Brian E. "A Proposed Technique for the Treatment of Restriction of Range in Selection Validation." Psychological Bulletin. Vol. 93, No 1, 1983, pp. 137-148.

Organization of Economic Cooperation and Development. Living Conditions in OECD Countries: A Compendium of Social Indicators. Social Policy Studies No. 3. Paris, France: Organization for Economic Co-Operation and Development, 1986.

Rosenbaum, James and Kariya, Tobe. "Market and Institutional Mechanisms for the High School to Work Transition in Japan and the U.S. Paper given at the August 1987 American Sociological Association Meeting.

Rothschild, Michael "Social Effects of Employment Testing." Madison, Wisconsin: Department of Economics, University of Wisconsin, 1979.

Schiff, M., Duyme, M., Dumaret, A., Stewart, J., Tomkiewicz, S., Feingold, J. (1978) "Intellectual status of working class children adopted into upper-middle class families." Science. 200:1503-04.

Schiff, M., Duyme, M., Dumaret, A., Tomkiewicz, S. (1982) "How much could we boost scholastic achievement and IQ scores: a direct answer from a French adoption study." Cognition, 12:165-96.

Schmidt, Frank. "The Problem of Group Differences in Ability Test Scores for Employment Selection." Journal of Vocational Behavior, Vol. 29, No. 3, December 1986. pp. 340-362.

Sizer, Theodore R. Horace's Compromise: The Dilemma of the American High School. The First Report from A Study of High Schools, co-sponsored by the National Association of Secondary School Principals and the Commission on Educational Issues of the National Association of Independent Schools, Boston: Houghton Mifflin Co., 1984.

Stevenson, Harold W. "Making the Grade: School Achievement in Japan, Taiwan and the United States." Annual Report of the Center for Advance Study in the Behavioral Sciences, Stanford, 1983.

Stevenson, Harold W.; Lee, Shin-Ying and Stigler, James W. "Mathematics Achievement of Chinese, Japanese & American Children." Science, February 1986, pp. 693-699.

Stiglitz, Joseph E. "Risk Sharing and Incentives in Sharecropping." Review of Economic Studies, 61, no. 2 April 1974, 219-256.

Taubman, P. and Wales, T. "Education as an Investment and a Screening Device." In Education, Income, and Human Behavior, edited by F. T. Juster. New York: McGraw-Hill, 1975.

Thorndike, R. L. Personnel Selection: Test and Measurement Techniques. New York: Wiley, 1949.

Trattner, M. H., Corts, D. B., van Rijn, P. P. and Outerbridge, A. M. (1977). Research base for the written test portion of the Professional and Administrative Career Examination (PACE): Prediction of job performance for claims authorizers in the social insurance claims examining occupation. Washington, D. C.: U.S. Office of Personnel Management.

Tuddenham, R. D. "Soldier Intelligence in World Wars I and II. American Psychologist, 3, 1948, 54-56.

Vineberg, R. and Taylor, E. N. (1972). Performance in four army jobs by men at different aptitude (AFQT) levels. Alexandria, VA: Human Resources Research Organization (HUMRRO).

Weiss, Andrew. "High School Graduates: Performance and Wages" Bell Communications Research, Inc. Economics Discussion Paper # 3, 1985.

Wigdor, Alexandra K. and Hartigan, John A. editors. Interim Report-Within Group Scoring of the General Aptitude Test Battery. Washington, DC: National Academy Press, 1988.

Table 1

Racial Gaps in Science, Math and Reading Proficiency

[In Grade Equivalent Units]

	Test Date								
	1969	1971	1973	1975	1978	1980	1982	1984	1986
Science									
At Age 17	6.7	--	6.6	--	7.1[a]	--	7.2	--	5.6
At Age 13	6.1	--	6.6	--	6.0[a]	--	5.0	--	4.1
Math									
At Age 17	--	--	4.0	--	3.2	--	3.2	--	2.9
At Age 13	--	--	4.6	--	4.2	--	3.4	--	2.4
Reading									
At Age 17	--	5.3	--	5.0	--	4.8	--	3.3	2.6
At Age 13	--	4.2	--	3.9	--	3.3	--	2.8	2.3

Source: National Assessment of Educational Progress, Crossroads in American Education. February 1989, Figure 2A. In science the difference between 17 year olds and 9 year olds was 64.2 points on the NAEP scale, so a grade equivalent unit was defined as 8.025 on the NAEP scale. The Mathematics Report Card. June 1988, Figure 1.2. The difference between 17 year olds and 9 year olds was 80.3 points on the NAEP scale. Consequently, a grade equivalent unit was defined as 10 points on the NAEP scale. The Reading Report Card. 1985, Data Appendix and Who Reads Best?, February 1988, Table 1.1. The difference between the scores of 17 year olds and 9 year olds was 75 points on the NAEP scale used in the report covering 1971 through 1984 and 18 on the scale used in the report on the 1986 assessment. Consequently, a grade equivalent unit was defined as 9.375 points on the NAEP scale used in the 1971-84 report and 2.25 points on the scale used in the report on the 1986 assessment.

[a] The Science NAEP was administered in 1977 not 1978.

Table 2

Effect of Academic Achievement on the Wage Rates of High School Graduates

Study and Data Set	Date of Graduation	Age	Achievement Measures	Percent Change in Wage Rate Male	Percent Change in Wage Rate Female
Wage Rates					
Kang & Bishop (1985) High School & Beyond	1980	19	Test-Math,Voc,Read	-1.9	-.5
			GPA in Grade 12	.6	2.2
Gardner (1983) NLS Youth	1976-1982	19-24	AFQT	4.8	4.8
Daymont & Rumberger NLS Youth (1982)	1976-1979	19-21	GPA in Grade 9	.3	2.7
Meyer (1982) (Weekly earnings) Class of 1972	1972	19	Class Rank Grade 12	0.0	2.5
			Test Composite	1.2	2.2
Earnings					
Hause (1975) Project Talent (white)	1961	19	IQ,Test-Math	-3.7	--
		23	IQ,Test-Math	6.1	--

The table reports the percentage response of the wage rate or earnings to a one standard deviation improvement in a measure of academic achievement. For high school seniors a one standard deviation differential on an achievement test is about equal to 3.5 grade level equivalents or 110 points on the Verbal SAT. For GPA, one standard deviation is about .7 when C's = 2.0, B's = 3.0 and A's = 4.0.

TABLE 3
DETERMINANTS OF RELATIVE JOB PERFORMANCE

	Yrs of Schooling	Academic Achievement	Perceptual Speed	Psychomotor Skills	Age	Age Square	Occ Exp	Occ Exp Square	Tenure	Tenure Square	R^2	N
Plant Operators	-.013 (.43)	.244*** (3.89)	.112* (1.68)	.117** (2.30)	.048* (1.69)	-.00053 (1.45)	.024 (.51)	-.00039 (.28)	.096* (1.93)	-.002 (1.36)	.181	651
Technician	.028* (1.75)	.277*** (8.25)	.024 (.72)	.117*** (4.35)	-.005 (.33)	-.00008 (.36)	.041*** (2.93)	-.00097** (2.11)	.084** (5.47)	-.0023*** (3.66)	.115	2384
Craft Workers	-.017** (2.48)	.249*** (15.00)	.060** (3.36)	.079*** (5.96)	.046*** (5.86)	-.00065*** (6.51)	.046*** (8.43)	-.00034*** (2.27)	.064*** (11.37)	-.0016*** (8.60)	.141	10061
High Skill Clerical	.013 (.82)	.272*** (8.75)	.085*** (3.17)	.094*** (3.63)	.035** (2.31)	-.00051** (2.55)	.020 (1.35)	-.00017 (.36)	.117*** (7.35)	-.00316*** (5.07)	.145	2570
Low Skill Clerical	-.015 (1.28)	.296*** (11.91)	.107*** (4.43)	.092*** (4.48)	.035*** (3.46)	-.00057** (4.29)	.042*** (3.36)	-.00090** (2.15)	.095*** (6.73)	-.0027*** (4.94)	.135	4124
Service	-.024 (1.45)	.298*** (8.14)	.072** (1.96)	.138*** (4.65)	.045*** (3.43)	-.00056*** (3.28)	.084*** (5.16)	-.0022*** (4.16)	.052*** (2.70)	-.0012 (1.61)	.152	1928
Operatives & Laborers	-.049** (6.59)	.189*** (10.65)	.079*** (4.37)	.140*** (9.53)	.047*** (6.62)	-.00064*** (6.79)	.038*** (3.77)	-.00052 (1.58)	.078*** (7.38)	-.00166*** (4.65)	.137	8167
Sales Clerks	-.024 (.70)	.119 (1.34)	.118 (1.41)	.167** (2.38)	.071*** (2.63)	-.00084** (2.45)	-.009 (.26)	.0012 (1.08)	.026 (.62)	-.0008 (.50)	.087	417

Table 4

LOSS IN PRODUCTIVITY IF
RANDOM ASSIGNMENT WERE SUBSTITUTED
FOR THE CURRENT ALLOCATION OF WORKERS
[LOWER BOUND ESTIMATE]

	Average Compensation per FTE	Standard Deviation of Output	Loss Per Worker	Number of Workers (1000's)	Aggregate Loss (billions)
Plant Operators	$33,808	$91,020	-$9,652	228	-$ 2.3
Technicians	$26,649	$13,668	-$8,672	5261	-$45.6
Craft Workers	$29,655	$12,399	-$3,700	13073	-$48.4
High Skill Clerical	$23,065	$ 8,925	-$4,914	5227	-$25.7
Routine Clerical	$19,472	$ 4,934	-$1,512	12082	-$18.3
Service Exc. Police & Fire	$15,496	$ 4,068	+$ 889	12724	$11.3
Operatives & Laborers	$23,828	$ 5,062	+$ 250	16816	$ 4.2
Sales Clerks	$17,542	$ 5,228	-$ 723	5682	-$ 4.0
All Workers	$22,566	$ 6,708	-$1,815	71,132	-$128.7

Estimates compare the predicted productivity of current members of each occupation with the mean predicted productivity in that occupation of everyone in the USES data set. Predicted job performance was calculated using equation 3, the best fitting model of job performance which included individual variables for gender, race and Hispanic. Dollar impacts were then calculated by first adjusting for the unreliability of the criterion in the standard manner (i.e. dividing by $\sqrt{.6}$), then correcting for restriction of range by multiplying by 1.76 and then multiplying by the standard deviation of output in dollars (column 2 of Table 5).]

Table 5

THE EFFECT OF RE-SORTING
ON AGGREGATE OUTPUT
[UPPER BOUND ESTIMATE]

	Coefficient of Variation	Impact of Resorting on Average Output: Percent	Impact of Resorting on Average Output: Dollars	Aggregate Gain (billions $)
Plant Operators	---	---	$159,282	36.3
Technicians	33.8	17.8	$ 12,667	66.7
Craft Workers	27.6	7.1	$ 5,623	73.6
High Skill Clerical	25.5	.9	$ 579	3.0
Routine Clerical	16.7	.6	$ 190	2.3
Service Exc. Police & Fire	17.3	1.3	$ 537	6.9
Operatives & Laborers	14.0	-3.4	-$ 2,152	-36.3
Sales Clerks	29.8	-23.8	-$ 7,322	-41.5
All Workers			$ 1,558	111.0

Estimates compare the predicted productivity of current members of each occupation with the predicted productivity of those assigned on the basis of equation 2 which does not make use of information on gender and ethnicity. Equation 2 performance predictions were made for each occupation and each worker. Because the standard deviation of output measured in dollars of plant operators was so high, this occupation got first pick. Then came technicians, craft occupations etc. Those not selected for one of the top 7 occupations became sales clerks. Once workers were assigned to occupations on the basis of equation 2, predicted job performance was then calculated using equation 3, the best fitting model of job performance which included individual variables for gender, race and Hispanic. Dollar impacts were then calculated by first adjusting for the unreliability of the criterion in the standard manner (i.e., dividing by $\sqrt{.6}$), multiplying by 1.76 to correct for range restriction and then multiplying by the standard deviation of output in dollars (column 2 of Table 5).]

Table 6

THE EFFECT OF THE RE-SORTING ON THE ABILITY, GENDER AND ETHNICITY OF OCCUPATIONS

	General Ability (Pop SD's)		Education		Percent Female		Percent Black		Percent Hispanic	
	Current Level	Change	Current Level	Change	Current Level	Change	Current Level	Change	Current Level	Change
Plant Operator	.09	+2.03	12.1	2.09	2	+75	11.1	-11	5.1	-5
Technician	.28	1.03	13.7	.22	55	+1	8.1	- 6	3.4	-1
Craft	-.09	.65	11.9	.53	4	+43	7.1	- 4	7.4	-3
High Skill Clerical	.32	.02	12.9	.67	83	-16	10.1	- 4	5.4	-1
Low Skill Clerical	.00	.03	12.6	-.48	82	-18	10.6	- 4	5.7	0
Service Exc. Police & Fire	-.52	.12	11.8	.01	82	-14	18.0	- 6	8.3	0
Operative	-.59	-.46	11.3	-.01	66	0	14.7	+ 7	10.0	+2
Sales Clerk	-.02	-1.59	12.3	-1.00	86	-20	8.2	+24	5.6	+3
All Occupations	-.21		12.1		62		11.9		7.3	

This table reports the gender, ethnicity, schooling and test scores of current members of each occupation and the changes in each of these variables that would result if new hires had been selected on the basis of the equation 2 predicted productivity regressions (which ignore race and ethnicity). The simulation was conducted by first calculating the equation 2 performance predictions for each worker in each occupation. Because the standard deviation of output measured in dollars of plant operators was so high, this occupation got first pick. Then came technicians, craft occupations etc. Those not selected for one of the top 7 occupations became sales clerks.

NAME OF WORKER (Print) (Last) (First)

SEX: MALE________ FEMALE________

Company Job Title: ____

How often do you see this worker in a work situation?

- ☐ All the time.
- ☐ Several times a day.
- ☐ Several times a week.
- ☐ Seldom.

How long have you worked with this worker?

- ☐ Under one month.
- ☐ One to two months.
- ☐ Three to five months.
- ☐ Six months or more.

A. How much can this worker get done? (Worker's ability to make efficient use of time and to work at high speed.) (If it is possible to rate only the quantity of work which a person can do on this job as adequate or inadequate, use #2 to indicate "inadequate" and #4 to indicate "adequate.")

- ☐ 1. Capable of very low work output. Can perform only at an unsatisfactory pace.
- ☐ 2. Capable of low work output. Can perform at a slow pace.
- ☐ 3. Capable of fair work output. Can perform at an acceptable pace.
- ☐ 4. Capable of high work output. Can perform at a fast pace.
- ☐ 5. Capable of very high work output. Can perform at an unusually fast pace.

B. How good is the quality of work? (Worker's ability to do high-grade work which meets quality standards.)

- ☐ 1. Performance is inferior and almost never meets minimum quality standards.
- ☐ 2. Performance is usually acceptable but somewhat inferior in quality.
- ☐ 3. Performance is acceptable but usually not superior in quality.
- ☐ 4. Performance is usually superior in quality.
- ☐ 5. Performance is almost always of the highest quality.

C. How accurate is the work? (Worker's ability to avoid making mistakes.)

- ☐ 1. Makes very many mistakes. Work needs constant checking.
- ☐ 2. Makes frequent mistakes. Work needs more checking than is desirable.
- ☐ 3. Makes mistakes occasionally. Work needs only normal checking.
- ☐ 4. Makes few mistakes. Work seldom needs checking.
- ☐ 5. Rarely makes a mistake. Work almost never needs checking.

MA 7-66
Apr. 1973

D. How much does the worker know about the job? (Worker's understanding of the principles, equipment, materials and methods that have to do directly or indirectly with the work.)

- ☐ 1. Has very limited knowledge. Does not know enough to do the job adequately.
- ☐ 2. Has little knowledge. Knows enough to get by.
- ☐ 3. Has moderate amount of knowledge. Knows enough to do fair work.
- ☐ 4. Has broad knowledge. Knows enough to do good work.
- ☐ 5. Has complete knowledge. Knows the job thoroughly.

E. How large a variety of job duties can the worker perform efficiently? (Worker's ability to handle several different operations.)

- ☐ 1. Cannot perform different operations adequately.
- ☐ 2. Can perform a limited number of different operations efficiently.
- ☐ 3. Can perform several different operations with reasonable efficiency.
- ☐ 4. Can perform many different operations efficiently.
- ☐ 5. Can perform an unusually large variety of different operations efficiently.

F. Considering all the factors already rated, and only these factors, how good is this worker? (Worker's all-around ability to do the job.)

- ☐ 1. Performance usually not acceptable.
- ☐ 2. Performance somewhat inferior.
- ☐ 3. A fairly proficient worker.
- ☐ 4. Performance usually superior.
- ☐ 5. An unusually competent worker.

Complete the following ONLY if the worker is no longer on the job.

G. What do you think is the reason this person left the job? (It is not necessary to show the official reason if you feel that there is another reason, as this form will not be shown to anybody in the company.)

- ☐ 1. Fired because of inability to do the job.
- ☐ 2. Quit, and I feel that it was because of difficulty doing the job.
- ☐ 3. Fired or laid off for reasons other than ability to do the job (i.e., absenteeism, reduction in force).
- ☐ 4. Quit, and I feel the reason for quitting was not related to ability to do the job.
- ☐ 5. Quit or was promoted or reassigned because the worker had learned the job well and wanted to advance.

RATED BY	TITLE	DATE
COMPANY OR ORGANIZATION	LOCATION (City, State, ZIP Code)	

GPO 863.716

MA 7-66
Apr. 1973

APPENDIX B

STUDIES OF OUTPUT VARIABILITY

A search for studies of output variability yielded 49 published and 8 unpublished papers covering 94 distinct jobs.[6] Their results are reported in tables 1 through 4. Table 1 summarizes the studies of output variability among semiskilled factory workers. The jobs known to be paid on a piece rate basis are not included in the table. Schmidt and Hunter (1983) found that such jobs typically have smaller coefficients of variation. Apparently when workers are paid on a piece rate basis, quit rates are more responsive to productivity than when pay is on an hourly basis. The less productive workers self select themselves out of such jobs and the surviving job incumbents become more and more similar in their output.

Estimates of productivity standard deviations (SD$) in 1985 dollars are reported in column 2 of the tables. In most cases the author of the study made no attempt to estimate SD$'s, so the estimate has been calculated from the CV. Such estimates are placed in a parenthesis. The estimates of SD$ were derived as a product of the CV, the mean compensation for that job and the ratio of value added to compensation for that industry. This ratio is 1.52 for private non-farm business excluding mining, trade, finance and real estate. The value added to compensation ratio in retailing and in real estate was much too high to be used as an adjustment factor. So for all sales occupations it was assumed that SD$ = CV times average compensation. The SD$ of semiskilled factory jobs ranged from $1732 to $7811 and averaged $5062 for jobs not known to be paid on a piece rate.

Table 2 reports managerial estimates of coefficients of variation and productivity SD$'s for plant operators and a number of craft occupations. For craft occupations other than plant operators, the average CV is 27.6 percent and the average SD$ is $12,399. These are smaller than for plant operators and larger than those for semi-skilled factory workers. Within the ranks of blue collar workers there is a clear tendency for coefficients of variation and standard deviations of output to rise with the complexity and wage rate of the job.

Output variability is also great in professional and high level managerial occupations. Users of communication satellites, for example, are going to save billions of dollars as a result of a discovery by a scientist at Comsat which has doubled the effective lifetime of satellites.

Exxon had invested a billion dollars in its shale oil operation at Parachute Creek before giving up on the enterprise. A wiser CEO or better staff work might have avoided or reduced this loss. It does not take many such examples to produce a very large standard deviation of output for professional and high level managerial jobs. In most white collar jobs, however, output variability across incumbents is much smaller.

Table 3 reports the results of studies of output variability in clerical occupations. In many of these studies hard measures of output (e.g., cards punched) were the basis for calculating coefficients of variation.

Table 4 contains estimates of CVs and standard deviations of output for the remainder of the occupational distribution: managerial, technical, sales service personnel. For sales personnel the CVs are based on hard data, distributions of actual sales. The variability of output in sales occupations is clearly higher than in most other occupations and the variability appears to rise with the complexity of the product that is being sold and the amount of initiative required to sell large amounts of the product. For high level sales personnel working in finance and manufacturing many of them paid on a commission basis, the coefficient of variation is 62.8 percent while for sales clerks it is 29.8 percent. When multiplied by mean levels of compensation for full time workers in these occupations, these CVs translate into output standard deviations of $15000 and $5228.

For most of the managerial and technical jobs studied physical measures of output were not definable so the supervisors were asked to report dollar amounts of output expected from workers at the 15th, 50th and 85th percentiles of the job performance distribution. Coefficients of variation averaged 36 percent for technicians implying an output standard deviation of $13668. The coefficient of variation was 33 percent for low level managers and 20.6 percent in the only three service occupations for which data is available. It was felt that these three jobs represented too small a sample to produce reliable estimates of the CV for all service jobs except police and fire fighting so the estimate of the service CV employed in the rest of the paper is an unweighted average of the CVs for operatives, low skill clerical workers and 20.6, the average for the three service jobs for which there is data on the variability of output. While the standard deviation of output appears to be substantial (about $4000) in full time full year service jobs, there is clearly a positive correlation between average wage levels and SD$'s.

TABLE 1

UNSKILLED AND SEMISKILLED BLUE COLLAR WORKERS

	C.V. of Output (Incumb)	Standard Deviation in 1985 Dollars[a]	Method	Sample Size	Source
Hourly or Weekly Pay					
Butter Wrappers	18.4	(4129)	PO	8	Rothe (1946)
Machine Operators	20.5	(6411)	PO	130	Rothe (1947)
Electrical Workers	13.2	(3399)	PO	33	Tiffin (1947)
Assembly Worker	12.8	(4035)	PO	294	Barnes (1958)
Coil Winders	15.0	(3782)	PO	27	Rothe & Nye(1958)
Craft	7.5	$2364	PO	61	Rothe & Nye (1958)
Machine Operators	11.7	$3688	PO	37	Rothe & Nye (1959)
Radial Drill Operator	25	$7881	CA		Roche (1961)
Entry Level Steelworkers	13.7	(6064)	WS	249	Arnold et al. (1983)
Entry Level Steelworkers	6.8	$3000	SHMM	NA	Rauschenberger (1986)
Armor Crewman	16.2		WS	374	Vineberg & Taylor (1972)
Pay Form: Unknown					
Machine Operator	9.1		PO	76	Baumberger (1921)
Soap Wrappers	8.9		PO	30	Wyatt (1927)
Tile Sizing & Sorting	19.1		PO	18	Wyatt (1932)
Paper Sorters	8.7		PO	18	Hearnshaw (1937)
Lamp Shade Manufac.	8.6	(2805)	PO	19	Stead & Shartle (1940)
Wool Pullers	15.1	(2256)	PO	13	Lawshe (1948)
Machine Sewers	14.6	(1732)	PO	100	Wechsler (1952)
Electrical Workers	12.7	(3279)	PO	65	Wechsler (1952)
Cable Makers	17.7	(4596)	PO	40	McCormick & Tiffin (1974)
Electrical Workers	14.1	(3638)	PO	138	McCormick & Tiffin (1974)
Assemblers	19.6	(6095)	PO	35	McCormick & Tiffin (1974)
	14.0	$ 5062			

Estimates of standard deviation of the output (SD$) of full time full year workers that are presented in parenthesis were derived from coefficients of variation (CV) for output. For jobs outside of mining,retailing and finance it was assumed that a more capable worker would necessitate proportionately more materials, energy inputs, overhead labor inputs but not necessitate additional capital. This means that the metric of the CV is K-L productivity and thus that in manufacturing where the ratio of value added to compensation is 1.51, a 10 percent gain in K-L productivity has a dollar value equal to about 15 percent of compensation. Consequently, $SD\$_j = CV_j$ (GNP per full time equivalent worker in industry k)($wage_{kj}$/($wage_k$) where $wage_{kj}$ = average wage of occupation j in industry k and $wage_k$ is average wage in industry k. The ratio of occupation "j"s earnings to the industry average was derived from Table 2 of Occupation by Industry Subject Report of the 1980 Census.

Methods used to Estimate the Coefficient of Variation and Standard Deviations of Output

PO - Physical Output - Where a piece rate prevails, ticket earnings are used as the output measure. Where pay is hourly, physical quantity of output or percent of standard output for the job is used as the output measure. CV's are calculated from this data and SD$s are constructed by using value added per employee (adjusted for relative wage rates) to value the productivity of the average worker.

WS - Work Sample - A sample of the job tasks is taken and workers are observed performing these tasks under controlled conditions. To be useful for calculating a CV, the WS must be defined in units that have a ratio scale that corresponds to output such as 50 lb sacks carried from A to B. It measures peak performance and thus probably does not measure effort as actually applied to a real job. SD$s are calculated from CV's in same way they are calculated from PO based CV's.

GS - Gross Sales - CV's are the SD of sales across sales personnel divided by the mean level of sales. SD$ equals the CV times the mean compensation of sales personnel. GS(A) is calculated using a weighted average of the sales of different products.

SHMM - Schmidt, Hunter, McKenzie and Muldrow (1979) Method. Managers who supervise job incumbents are asked to place monetary values on the output produced by an employee at the 15th, 50th and 85th percentile of the job performance distribution. The metric in which they are asked to make these judgement is the cost to have an "outside firms provide these products and services." This yields direct estimates of SD$ and a rough estimate of the CV can be calculated from $(P_{85} - P_{15})/2P_{50}$.

S(m) - Schmidt et al (1979) method with supervisors making their judgments after being supplied a mean output derived from company records.

S(T)- Schmidt et al (1979) method with outliers dropped from the calculation.

SE - Supervisor's estimate for actual employees. Supervisors give dollar values for the productivity of a sample of actual employees. The mean and standard deviation is calculated from this distribution.

S(D) - Schmidt et al (1979) method as modified by Dunnette et al (1982). A first round of workshops with supervisors identified examples of unusually effective, unusually ineffective and average levels of job performance by plant operators. Eight dimensions of performance were developed from these examples and supervisors were asked to retranslate and scale the 667 performance examples in a second round of workshops. Finally participants were asked to estimate dollar value of performance at the 85th, 50th and 15th percentile. Negative values were changed to zero.

Table 2

PRECISION PRODUCTION AND CRAFT OCCUPATIONS

	C.V. of Output (Incumb)	Standard Deviation in 1985 Dollars	Method	Sample Size	Source
Plant and System Operators					
Nuclear Control Room Oper.	108	$277,850	S(D)	34	Dunnette et al. (1982)
Fossil Fuel Cont. Room Oper.	72	$155,340	S(D)	48	Dunnette et al. (1982)
Nuclear Plant Operator	105	$ 97,370	S(D)	19	Dunnette et al. (1982)
Fossil Fuel Plant Operator	61	$ 39,455	S(D)	20	Dunnette et al. (1982)
Hydro Plant Operator	53	$ 27,030	S(D)	31	Dunnette et al. (1982)
Refinery Head Operator	--	$ 15,355	SE	19	Wroten (1984)
Outside Operator	--	$ 14,356	SE	19	Wroten (1984)
Pump Operator	--	$ 10,381	SE	17	Wroten (1984)
		$ 91,020			
Other Craft Workers					
Welders-Refinery	37.3	$ 16,775	SE	14	Wroten (1984)
Handcraft Workers	17.1	$ 5,390	PO	NA	Evans (1940)
Drillers	31	$ 9,772	PO	11	Lawshe (1948)
Arc Welder	16.0		WS	49	U.S. Job Service (1966)
Radar Mechanics [1]	40.3		WS	107	Whipple (1969)
Radar Mechanics [2]	20.1		WS	51	Whipple (1969)
Welders	13.7	$ 5,039	PO	25	Rothe (1970)
Repairman	21.4		WS	385	Vineberg & Taylor (1972)
Outside Mechanic	48.4	$ 21.800	SE	12	Wroten (1984)
Electrician	23	$ 12,539	SHMM	104	MacManus (1986)
Sheet Metal Worker	25	$ 11,696	SHMM	22	MacManus (1986)
Plumber	24	$ 11,856	SHMM	66	MacManus (1986)
Painter	24	$ 8,626	SHMM	41	MacManus (1986)
Meat Cutter	26	$ 7,778	SHMM	14	MacManus (1986)
Maintenance & Tool Room Jobs	46	--	SHMM		Bolda (1985)
	27.6	$ 12,399			
Supervisors					
Steel: Foreman (average)	--	$ 67,923	SHMM	11	Rauschenberger (1985)

The data on electric utility industry was collected in 1981 so the inflation factor based on the growth of utility wages and salaries per FTE is 1.30. The petroleum refinery industry inflation factor since 1983 is 1.10. The steel industry inflation factor is 1.084 for 1985 vs. 1982.

TABLE 3

CLERICAL

Routine Clerical Jobs					
Telegraph Operator	13.2		PO	14	Baumberger (1920)
Machine Bookkeepers	8.4		PO	39	Hay (1943)
File Clerks	17.9		PO	61	Gaylord (1951)
Card Punch Operator	11.5	(2488)	PO	NA	Klemmer & Lockhead (1962)
Proof Machine Operator	13.4	(2932)	PO	NA	Klemmer & Lockhead (1962)
Typists	18.6	(3980)	PO	616	Stead & Shartle (1962)
Card Punch Operator (Day)	10.7	(2278)	PO	113	Stead & Shartle (1962)
Card Punch Operator	21.6	(4550)	PO	62	Stead & Shartle (1962)
Card Punch Operator	12.9	(2746)	PO	121	Stead & Shartle (1962)
Proofreader	18.5		WS	57	US Job Service (1972)
Telephone Operator	17.7		WS	1091	Gael et al. (1975a)
Mail Carriers	22.5		WS	374	US Postal Service (1981)
Mail Handlers	22.7		WS	373	US Postal Service (1981)
Clerical	25	$ 5529	S(M)	91	Burke (1985)
Customs Inspector	15.7		WS	188	Corts et al. (1977)
Meter Reader	18	$ 4481	SHMM	14	MacManus (1986)
Toll-Ticket Sorters	14.9		PO	13	Maier & Verser (1982)
	16.7	$ 4934			
Clerical with Decision Making					
Supply Specialist	26.5		WS	394	Vineberg & Taylor (1977)
Mail Distribution	39.2		WS	417	US Postal Service (1981)
Claims Processor	28.5	$ 5111	CA	15	Ledvinka et al. (1983)
Claims Evaluators	24.5	$ 4896	PO	176	DeSimone et al. (1986)
" "	23.8	$ 3876	SHMM	27	" " "
Claims Authorizer	20.5		WS	233	Trattner et al (1977)
Ticket Agent	26	$ 8411	SHMM	9	MacManus (1986)
Head Teller - Bank	(15)	$ 2369	S(T)		Mathieu & Leonard (1986)
	25.5	$ 8925			

Footnotes for Table 3

[a]The Programmer Aptitude Tests raw validity is .38 based on Schmidt, Rosenberg and Hunter's (1980) validity generalization of data on 1299 programmers.

[b]The estimate of GMA job performance raw validities for technical jobs is based on 20 occupations and a total of 2417 cases. The estimate for professional occupations is based on 2 occupations and a total of 109 cases. Schmidt, Mack & Hunter classify the park ranger job as a level 3 job using Hunters (1983) classification scheme. For a level 3 job the raw validity of GMA is .28.

[c]GMA raw validity for managers is a simple average of 9 separate managerial occupations from the GATB manual.

[d]The raw validity estimate is from Churchill et al's "The Determinants of Sales Person Performance: A Meta-Analysis" (1985) and is based on 44 studies which used objective company data with controls for environmental conditions. Since actual sales data were used it is assumed that criterion reliability is 1.0.

[e]Cascio and Silbey estimated the average compensation of sales personnel to be $75 a day or $18000 a year in 1978. This was inflated to 1985 wage levels by multiplying by 1.555 and then multiplied by CV to estimate SD$.

[f]Bobko et al, SHMM type estimate of SD$ was $4967 which is inflated to 1985 wage levels by multiplying by 1.174 the growth of wages and salaries in the industry from 1982 to 1985.

[g]Pearlman, Schmidt, and Hunter 1980.

[h]Validity estimate for sales clerk jobs is an average of Ghiselli's estimate (-.06) and the mean of more recent studies (.14) is reported by Hunter and Hunter (1984).

TABLE 4

MANAGERIAL, TECHNICAL, SALES AND SERVICE WORKERS

Technical					
Computer Programmer	32	$16550	SHMM		Schmidt et al. (1979)
Budget Analyst	(47)	$15062	SHMM		Hunter & Schmidt (1982)
Park Ranger	33	$ 4828	SHMM		Schmidt et al. (1984)
Instrument Tech. - Refinery	(20)	$28720	SE	14	Wroten (1984)
Computer Programmer	47	$15888	SHMM		Rich & Boudreau (1986)
Cartographic Technician	33.5		WS	443	Campbell et al. (1973)
	33.8	$13668			
Managerial					
Convenience Store Manager	51	$13967	SHMM	110	Weekley et al. (1985)
Bank Branch Manager	(35)	$10064	S(T)		Mathieu & Leonard (1986)
Bank Operations Manager	(14)	$ 3122	S(T)		Mathieu & Leonard (1986)
	33.3				
High Level Sales					
District Sales - Food Manu.	32	($ 8958)*	SHMM	4	Cascio & Silbey (1979)
Insurance Salesman	37.5	$ 5219	CA	92	Bobko (1983)
District Sales Rep. Mfg.	41.3	$17529	GS	18	Burke & Frederick (1984)
Real Estate Sales	83	$21271	SHMM	63	MacManus (1986)
Life Insurance Sales	120	$12453	GS		Brown (1981)
	62.8				
Sales Clerk					
Sales Clerks	22.2	(2807)	GS	153	Stead & Shartle (1940)
Cashiers	17.3	(2147)	WS	29	Lawshe (1948)
Sales Clerks	47.3	(5734)	GS	18	Lawshe (1948)
Grocery Checker	19.3		WS	92	US Job Service (1976)
Cashier Checker	43	$11379	SHMM	29	MacManus (1986)
	29.8	$ 5228			
Service					
Cooks	21.4		WS	385	Vineberg & Taylor (1972)
Package Wrappers	24.1		PO	27	Blum & Candee (1941)
Package Packers	16.4		PO	10	Blum & Candee (1941)
Average of 3	20.6				
Average of Service, Low Clerical & Operatives	17.3	$ 4068			

Appendix C

CONSTRUCTION OF WEIGHTS FOR U. S. E. S. GATB REVALIDATION DATA

	Number of Individuals in USES Data Set			Number Employed (1000's)			Weights		
	All	Black	Hisp	All	Black	Hisp	Non Black Non Hisp	Black	Hisp
Plant Oper.	651	162	35	228	25.3	11.6	421	156	331
Technician	2390	583	249	5261	426	178	2989	731	716
High Skill Craft	10252	1676	789	13112	931	970	1440	555	1230
High Skill Clerical	2583	623	172	5220	525	282	2468	843	1639
Low Skill Clerical	4153	1223	289	12089	1281	689	3832	1047	2384
Service exc. Police & Fire	1933	759	125	13445	2144	1117	9451	3180	8936
Operative	8177	2873	653	16816	2472	1683	2723	860	2577
Sales Clerk	422	112	29	5682	466	318	17430	4160	10970

References for Appendix

Hirschhorn, Larry. Beyond Mechanization, (Cambridge, Mass. MIT Press, 1984).

Schmidt, Frank L., and Gast-Rosenberg, Ilene. "Validity Generalization Results for Computer Programmers." Journal of Applied Psychology, 1980, Vol. 65, No. 6, pg. 643-661.

Schmidt, Frank L., and Hunter, John E. "Individual Differences in Productivity: An Empirical Test of Estimates Derived from Studies of Selection Procedure Utility." Journal of Applied Psychology, 68, no. 3, 1983, pg. 407-414.

Schmidt, Frank L. and Hunter, John E. "Impact of Valid Selection Procedures on Work-Force Productivity." Journal of Applied Psychology, 1979, Vol. 64, No. 6, pp. 609-626.

Schmidt, Frank, Hunter, John and Caplan, James. "Validity Generalization Results for Two Jobs in the Petroleum Industry." Journal of Applied Psychology, 1981, 66, pg. 261-273.

Sources for Tables 1 through 4

Arnold, J. D., Rauschenberger, J. M. Soubel, W. & Guion, R. M. (1983). Validation and utility of a strength test for selecting steel workers. Journal of Applied Psychology, 67, 588-604.

Barnes, R. M. (1937). Time and motion study. New York: Wiley.

Barnes, R. M. (1958). Time and motion study. New York: Wiley (4th ed.).

Baumberger, J. P., and Marin, E. G. (1920). Fatigue and efficiency of smokers in a strenuous mental occupation. Journal of Industrial Hygiene, 2, 207-214.

Baumberger, J. P., Perry, E. E., and Martin, E. G. (1921). An output study of users and non-users of tobacco in a strenuous physical occupation. Journal of Industrial Hygiene, 3, 1-10.

Blum, M. and Candee, B. (1941). The selection of department store packers and wrappers with the aid of certain psychological tests. Journal of Applied Psychology, 25, 76-85.

Bobko, P., Karren, R. & Parkington, J. J. (1983). Estimation of standard deviations in utility analysis: An empirical test. Journal of Applied Psychology, 68, 170-176.

Bolda, R. A. (1985b). Individual Productivity: A Sourcing Analysis. Unpublished Manuscript.

Brown, S. H. (1981). Validity generalization and situational moderation in the life insurance industry. Journal of Applied Psychology, 66, (6), 664-670.

Burke, M. J. & Frederick, J. T. (1984). Two modified procedures for estimating standard deviations in utility analyses. Journal of Applied Psychology, 69, 482-489.

Burke, M. J. (1985). An investigation of dimensions employed and percentile ordering effects in estimating performance standard deviations in dollars for clerical occupations. Unpublished manuscript.

Campbell, J. T., Crooks, L. A., Mahoney, M. H. and Rock, D. A. (1973). An investigation of sources of bias in the prediction of job performance: a six year study. Princeton, NJ: Educational Testing Service.

Cascio W. F. & Silbey, V. (1979). Utility of the assessment center as a selection device. Journal of Applied Psychology, 64, 107-118.

Corts, D. B., Muldrow, T. W. and Outerbridge, A. M. (1977). Research base for the written test portion of the Professional and Administrative Career Examination (PACE): Prediction of job performance for customs inspectors. Washington, D.C.: U.S. Office of Personnel Management.

DeSimone, R. L., Alexander, R. A., and Cronshaw, S. F. (1986). Accuracy and reliability of SD_y estimates in utility analysis. Journal of Occupational Psychology, 59, 93-102.

Dunnette, M. D., Rosse, R. L., Houston, J. S., Hough, L. M., Toquam, J., Lammlein, S., King, K. W., Bosshardt, M. J., & Keyes, M. (1982). Development and validation of an industry-wide electronic power plant operator selection system. Edison Electric Institute.

Evans, D. W. (1940). Individual productivity differences. Monthly Labor Review, 50, 338-341.

Gael, S., Grant, D. L. and Ritchie, R. J. (1975a). Employment test validation for minority and nonminority telephone operators. Journal of Applied Psychology, 60, 411-419.

Gaylord, R. H., Russell, E., Johnson, C. and Severin, D. (1951). The relation of ratings to production records: an empirical study. Personnel Psychology, 4, 363-371.

Hay, E. N. (1943). Predicting success in machine bookkeeping. Journal of Applied Psychology, 27, 483-493.

Hearnshaw, L. S. (1937). Selection tests for paper sorters. Journal of Occupational Psychology, 11, 145-153.

Hunter, J. E., & Schmidt, F. L. (1982). Fitting people to jobs: The impact of personnel selection on national productivity. In M. D. Dunnette & E. A. Fleishman (Eds.), Human performance and productivity (Vol. 1, pp. 233-284). Hillsdale, NJ: Erlbaum.

Klemmer, E. T., & Lockhead, G. R. (1962). Productivity and errors in two keying tasks: A field study. Journal of Applied Psychology, 46, 401-408.

Lawshe, C. H. (1948). Principles of personnel tests. New York: McGraw-Hill.

Ledvinka, J., Simonet, J. K., Neiner, A. G., & Kruse, B. (1983). The dollar value of JEPS at Life of Georgia. Unpublished technical report.

MacManus, V. et al (1986). Report on Develop A Procedure For Utility Analysis in Validation Research. Western Test Development Field Center, Salt Lake City, Utah.

Maier, N. R. F. and Verser, G. C. (1982). Psychology in industrial organizations 5th ed., Houghtin Mifflin, Boston.

McCormick, E. J., & Tiffin, J. (1974). Industrial psychology. Englewood Cliffs, N.J.: Prentice-Hall.

Mathieu, J. E. & Leonard, R. L., Jr. (1986). An application of utility concepts to a supervisor skills training program: A time-based approach. Manuscript submitted for review.

Rauschenberger, J. M. (1986). The Utility of Valid Selection Procedures: Dollars and Sense.

Rich, J. R. & Boudreau, J. W. (1986). The effects of variability and risk on selection utility analysis: An empirical simulation and comparison. Manuscript submitted for review.

Roche, U. F. (1961). The Cronbach-Gleser Utility Function in Fixed Treatment Employee Selection. Unpublished doctoral dissertation. Southern Illinois University, 1961. (Portions reproduced in L. J. Cronbach & G. C. Gleser (Eds.), Psychological Tests and Personnel Decisions, Urbana, Ill., University of Illinois Press, 1965, pp. 254-266.

Rothe, H. F. (1946). Output rates among butter wrappers: II. Frequency distributions and an hypothesis regarding the "restriction of output." Journal of Applied Psychology, 30, 320-327.

Rothe, H. F. (1947). Output rates among machine operators: I. Distributions and their reliability. Journal of Applied Psychology, 31, 484-489.

Rothe, H. F. (1951). Output rates among chocolate dippers. Journal of Applied Psychology, 35, 94-97.

Rothe, H. F. (1970). Output rates among welders: Productivity and consistency following removal of a financial incentive system. Journal of Applied Psychology, 54, 549-551.

Rothe, H. F. (1978). Output rates among industrial employees. Journal of Applied Psychology, 63, 40-46.

Rothe, H. F., & Nye, C. T. (1958). Output rates among coil winders. Journal of Applied Psychology, 42, 182-186.

Rothe, H. F., & Nye, C. T. (1959). Output rates among machine operators: II. Consistency related to methods of pay. Journal of Applied Psychology, 43, 417-420.

Rothe, H. F., & Nye, C. T. (1961). Output rates among machine operators: III. A nonincentive situation in two levels of business activity. Journal of Applied Psychology, 45, 50-54.

Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979). Impact of valid selection procedures on work-force productivity. Journal of Applied Psychology, 64, 609-626.

Schmidt, F. L., Mack, M. J., & Hunter, J. E. (1984). Selection utility in the occupation of U.S. park ranger for three modes of test use. Journal of Applied Psychology, 69, 490-497.

Stead, W. H., & Shartle, C. L. (1940). Occupational Counseling Techniques. New York: American Book.

Tiffin, J. (1947). Industrial Psychology, (2nd ed.). New York: Prentice-Hall.

Trattner, M. H., Corts, D. B., van Rijn, P. P. and Outerbridge, A. M. (1977). Research base for the written test portion of the Professional and Administrative Career Examination (PACE): Prediction of job performance for claims authorizers in the social insurance claims examining occupation. Washington, D. C.: U.S. Office of Personnel Management.

U.S. Job Service (1966). Development of USES aptitude test battery for arc welder 810.884. Washington, D.C.: U.S. Employment Service.

U.S. Job Service (1972). Development of USES aptitude test battery for copy holder 209.588 and proofreader 209.688. Washington, D.C.: U.S. Employment Service.

U.S. Job Service (1976). Development of USES aptitude test battery for grocery checker 229.468. Washington, D.C.: U.S. Employment Service.

U.S. Postal Service (1981). Validation report for positions of city carriers, mail handler, distribution clerk, and distribution clerk machine. Employment and placement division, United States Postal Service.

Vineberg, R. and Taylor, E. N. (1972). Performance in four army jobs by men at different aptitude (AFQT) levels. Alexandria, VA: Human Resources Research Organization (HUMRRO).

Viteles, M. S. (1932). Industrial Psychology, New York: Norton.

Weekley, J. A., O'Connor, E. J., Frank, B., & Peters, L. W. (1985). A comparison of three methods of estimating the standard deviation of performance in dollars. Journal of Applied Psychology, 79, 122-126.

Wechsler, D. (1952). Range of Human Capaçities. (2nd ed.). Baltimore, Md.: Williams and Wilkins.

Whipple, J. W., Baldin, R. D., Mager, R. F. and Vineberg, R. (1969). A three hour performance test to evaluate job effectiveness of army radar mechanics. Alexandria, VA: Human Resource Research Organization.

Wroten, S. P. (1984). Overcoming the futilities of utility applications: Measures, models and management. Presented at the American Psychological Association's 92nd Meeting, Toronto, August.

Wyatt, S. (1927). An experimental study of a repetitive process. British Journal of Psychology, 17, 192-209.

Wyatt, S. and Langdon, J. N. (1932). Inspection processes in industry. Industrial Health Research Board, Report No. 63, Great Britain.

NOTICE

REPRODUCTION BASIS

[X] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").