ESTIMATING THE IMPACT OF THE GED ON THE EARNINGS OF YOUNG DROPOUTS USING A SERIES OF NATURAL EXPERIMENTS

John H. Tyler
Richard J. Murnane
John B. Willett

Working Paper **6391**

NBER WORKING PAPER SERIES

ESTIMATING THE IMPACT OF THE GED ON THE EARNINGS OF YOUNG DROPOUTS USING A SERIES OF NATURAL EXPERIMENTS

John H. Tyler
Richard J. Murnane
John B. Willett

Working Paper 6391
http://www.nber.org/papers/w6391

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 1998

We gratefully acknowledge the research support provided by the National Center for the Study of Adult Learning and Literacy (NCSALL)** and the Rockefeller, Russell Sage and Spencer Foundations. The first author is grateful to Jeffrey Kling for many helpful conversations that contributed to this work, to Joshua Angrist, Lawrence Katz, and Carline Minter Hoxby for their particular insights that pushed the work forward, to participants at the NBER Summer Institute Labor Studies workshop, the Harvard Labor Economics Seminar, and the MIT Labor/Public Economics Seminar for helpful comments, and to numerous state GED-program directors and practitioners for their willingness to share their knowledge and experience. Special thanks go to Janet Baldwin of the GED Testing Service of the American Council on Education, Linda Headley Walker of the State Education Department of New York, Leslie Averna of the Connecticut Department of Education, and John Sojat of the Florida Department of Education for their assistance in providing data on GED testers, and to personnel of the Office of Research and Statistics at the Social Security Administration's Baltimore Headquarters, particularly Peter Wheeler and Russell Hudson, for providing and assisting with Social Security earnings data. The support of these individuals and organizations does not imply their endorsement of the contents of this paper.

** Work supported by NCSALL was supported under the Educational Research and Development Centers Program, Award Number R309B60002, as administered by the Office of Educational Research and Improvement/National Institute on Postsecondary Education, Libraries, and Lifelong Learning, U.S. Department of Education. The contents do not necessarily represent the positions or policies of the National Institute on Postsecondary Education, Libraries, and Lifelong Learning, the Office of Educational Research and Improvement, the U.S. Department of Education, or the National Bureau of Economic Research, and you should not assume endorsement by the Federal Government.

Estimating the Impact of the GED on the Earnings
of Young Dropouts Using a Series of Natural
Experiments
John H. Tyler, Richard J. Murnane and John B. Willett
NBER Working Paper No. 6391
February 1998
JEL No. J1

ABSTRACT

The General Educational Development (GED) credential has become the primary "second chance" route to high school certification for school dropouts in the United States. Despite the widespread use of the GED, however, bias due to self-selection has limited our knowledge about the effects of the credential on the labor market outcomes of dropouts. This paper uses a series of natural experiments created by interstate variation in GED passing standards to reduce self-selection bias in estimates of the impact of the GED on the earnings of young dropouts. To exploit the natural experiments, we use a unique merged data set containing the GED test scores and Social Security earnings of a sample of 16-21 year-old dropouts who attempted the GED in 1990. As a result of our research design and the fact that lower-scoring GED candidates receive very little post-secondary education, our results primarily measure the labor market signaling value of the GED. For dropouts who have indicated a desire to acquire the credential and whose skills place them on the margin of passing the GED exams, we find that acquisition of a GED increases the 1995 earnings of young white dropouts by 10-19 percent. These results are robust to experiments that use different treatment and comparison groups, and they withstand sensitivity analyses that explore possible violations of our identifying assumptions. We find no statistically significant evidence that the GED increases the earnings of young nonwhite dropouts, a result that we attribute to a "programmatic effect" discussed in the paper.

John H. Tyler
Harvard Graduate School
of Education
6 Appian Way
Cambridge, MA 02138
tylerjo@hugse1.harvard.edu

Richard J. Murnane
Harvard Graduate School
of Education
6 Appian Way
Cambridge, MA 02138
and NBER
murnanri@hugse1.harvard.edu

John B. Willett
Harvard Graduate School
of Education
6 Appian Way
Cambridge, MA 02138
willetjo@hugse1.harvard.edu

Estimating the Impact of the GED on the Earnings of Young Dropouts Using a Series of Natural Experiments

Introduction

The General Educational Development (GED) credential has become the primary "second chance" route to high school certification for school dropouts in the United States. The GED program was initiated in 1942 as a means of providing World War II veterans lacking a high school diploma with an opportunity to earn a secondary credential. Civilians first took the tests in 1952. Since 1949, 12.6 million adults have been awarded a GED. In 1996 a record 758,500 school dropouts attempted the GED and 524,500 of these achieved scores high enough to be awarded the certificate. Currently, all fifty states, the District of Columbia, U.S. territories, 10 Canadian provinces and territories, and several foreign countries use and recognize the GED as an alternative high school credential.[1]

The growth of the GED program is due to two overarching reasons. First, the GED is recognized as a valid secondary-completion certificate by most post-secondary institutions and training programs. Second, governmental policies have both encouraged school dropouts to pursue a GED and have funneled federal dollars into GED preparation programs.[2]

Within the last ten years at least thirteen studies have examined the impact of the GED on labor market outcomes.[3] However, all of these studies are potentially compromised by selectivity bias. The self-selection problem arises because unmeasured variables (e.g., motivation) that determine who selects to acquire a GED may also determine earnings. Failure to control for these

---

[1] Information in this paragraph is from the GED Testing Service (1991b) and (1997).

[2] Cameron and Heckman (1993) argue that (a) the 1966 Adult Basic Education Act and (b) a variety of federal programs related to post-secondary education and training (such as qualification requirements for Pell grants) created a demand for high school credentials that fueled the post-1963 growth in the number of GED-certified secondary completers. We note that many welfare and job training programs also have components that encourage GED acquisition.

[3] See Boesel, Alsalam, and Smith (1997) for an extended listing. See also Murnane, Willett, and Boudett (1995; 1997).

variables will confound the effects of the credential on earnings with the effects of the unmeasured variables on earnings.

We are able to reduce selectivity bias and provide more credible estimates of the treatment effect of the GED on earnings in two ways. First, we use a research design that exploits a series of "natural experiments"[4] that arise because different states have different standards regarding what constitutes a "passing" score on the GED exams. This interstate variation in GED passing standards allows us to compare the mean earnings of dropouts who have the same GED test-scores, but who differ in GED-status depending on their state of residence. The result is that we have a plausible source of exogenous variation in GED-status. Second, our data—containing basic demographic information, GED test-scores, and Social Security-taxable earnings—are from a large, national sample of dropouts *who have all attempted the GED battery of exams*. Using these data, we control for all treatment and comparison group differences that originate in the decision to acquire a GED.[5] Our research design and data mean that our estimates can be interpreted as the effect of the GED on the annual earnings of young dropouts who desire a GED and whose cognitive skills place them at the margin of being able to pass the GED exams. For such individuals, our estimates show that GED acquisition is associated with a 10 to 19 percent increase in the earnings of young, white dropouts five years after obtaining the credential.[6] We find no evidence that the GED affects the earnings of young, non-white dropouts. As we will

---

[4] For a discussion of the use of "natural experiments" in an economic context see Meyer (1995), Heckman and Smith (1995), and Besley and Case (1994).

[5] As noted by a reader of an earlier draft, this is true under the assumptions that (1) the costs of taking the GED are the same across states, and (2) the decision to take the exam is unaffected by the passing standard. Assumption (1) is effectively satisfied since the out-of-pocket costs for taking the GED in 1990 ranged from $0 to $40. The second assumption is an important identifying assumption in our work, and we discuss possible violations of this assumption later in the paper.

[6] To put the 19 percent upper bound estimate in perspective, if this were entirely a wage effect, then the $1,500 increase to the $7,800 baseline earnings of the comparison group, would represent an extra 75 cents an hour (from $3.90 to $4.65) for a full-time, full-year worker. On the other hand, analyses that we discuss in the appendix suggests that a substantial part of the treatment effect we measure, especially for females, may represent movement from persistent non-employment to employment.

show later, our estimates are primarily measures of the labor market signaling value of a GED.

This paper is organized as follows. In Section 1 we discuss the GED and its relation to the labor market. Next, we discuss the natural experiments that we use to reduce selectivity bias. In Sections 3 and 4 we discuss the data and the methodology we employ to obtain our estimates, and in Section 5 we present our results and discuss possible explanations for the white–nonwhite differences we find. Section 6 concludes and summarizes.

Section 1.
The GED and the U.S. Labor Market

The GED battery of exams consists of five tests covering mathematics, writing, social studies, science, and "interpreting literature and the arts."[7] Since 1988, the writing component of the test has included an essay; the rest of the test battery is multiple choice. Examinees may take the test in English, Spanish, or French,[8] and they may choose to take any or all of the five sub-tests in a particular sitting. An additional feature of the GED program is that individuals with low scores on any of the tests may retake these tests,[9] and later in the paper we discuss the implications of this retesting feature for how we use our data. There are state restrictions on minimum age (usually

[7] This latter test is often referred to as a "reading" test.

[8] This is an interesting feature of the GED since it means that an individual could obtain this high school equivalency credential in the U.S. without being able to read English. In 1990, about 5 percent of all testers took the Spanish language version of the tests, while less than one-half of one percent took the tests in the French language version. In recent years some degree-granting post-secondary institutions have made the decision to not accept GED credentials that were earned on a non-English version of the test (Batts 1997; Erwin 1997). It is not clear, however, how widespread this practice is.

[9] There is state variation in retesting guidelines, with most states requiring individuals to wait two to six months before retesting. Also, individuals cannot retest on the same GED test form in a given year. Since there are 3 forms, this means that an individual may take the GED up to three times a year. Our tabulations of the Connecticut and Florida data (the only data that allow us to identify multiple testers) show that 18 percent of white dropouts and 25 to 30 percent of nonwhite dropouts had taken the battery prior to their 1990 attempt.

17-18) and residency (GED Testing Service 1991b).

The GED test battery is periodically normed using a national sample of graduating high school seniors; the last norming was completed in 1987.[10] The Commission on Educational Credit and Credentials of the American Council of Education (ACE)[11] has traditionally recommended that state minimum score requirements be set such that about 70 percent of the norming sample would qualify for the credential. However, the GED program is jointly supervised by the GED Testing Service (GEDTS) and the individual state education agencies, and each state is free to set standards higher than the ACE-suggested minimum. Most have chosen to do so, and this fact creates the natural experiments that are critical to our study.

As stated earlier, the purpose of this study is to isolate and estimate the impact of obtaining a GED on the earnings of young dropouts. There are three primary routes through which acquisition of a GED could, in fact, lead to improved labor market outcomes.[12]

First, for those dropouts with low levels of basic cognitive skills or with English language deficiencies, raising their skills to levels required to pass the GED battery would increase their store of human capital. The evidence on the average amount of time spent in GED preparation activities does not give a clear indication of the potential importance of this human capital component of the GED.[13]

---

[10] The norming is conducted on a sample of individuals who have no stake in the outcomes of the test. This may or may not be a critical determinant in the effectiveness of the norming process. The GEDTS has conducted a study that indicates that the GED scores of the norming sample of high school seniors are highly correlated with their regular school grades (Auchter and Skaggs 1994).

[11] The GED Testing Service is a division of the American Council on Education.

[12] Prior work on the GED has generally included some discussion of the mechanisms through which a GED could affect labor market outcomes. See for example Cameron and Heckman (1993), Cameron (1994), Murnane, Willett, and Boudett (1995; 1997), and Murnane, Willett, and Tyler (1996).

[13] We know that the median time spent in GED preparation in 1989 was only 30.5 hours (GED Testing Service 1991a). Cameron and Heckman (1993) use these figures to suggest that, on average, there could be little human capital acquisition associated with pre-GED preparation, since little time was spent preparing for the GED battery. However, these figures on time spent

Spence's (1973) signaling model offers a second mechanism through which a GED could affect the earnings of dropouts. In this case, dropouts with higher levels of productive traits such as ambition or cognitive skills need a way to distinguish themselves in the labor market from other dropouts. These more skilled, motivated, or "mature" dropouts obtain a GED as a way to signal to employers that they possess these positive attributes. If employers value the GED as a signal of otherwise unobservable or costly-to-observe productive traits, then dropouts with a GED will have better employment opportunities, on average, than dropouts without the credential. As a result we would likely see GED-holders with higher earnings than non-credentialed dropouts.[14]

Finally, it could be that the GED indirectly affects labor market outcomes through its function as a "gatekeeper" to many post-secondary education and training programs and the funding that students often use to pay for these programs (e.g., Pell grants). Virtually all post-secondary institutions and many training programs require a high school degree or a high school equivalency credential for admittance to degree-granting programs,[15] and the GED is by far the

"...studying for the tests" (GED Testing Service, 1991, p. 2) may tell us little about a potentially large segment of the GED population that may have spent many months in Adult Basic Education (ABE) or English as a Second Language (ESL) classes bringing their basic skills and language up to levels sufficient to pass the test. Because of the way the "study time" questions have been worded in GEDTS surveys, it is unlikely that respondents included long-term time spent in basic skills preparation as a part of their response. Thus, the median "study" time for the GED could potentially be higher than what has previously been reported, and thus, GED-acquisition could well be associated with substantial human capital formation.

[14] A critical component of Spence's model is that it must be less costly for the more productive dropouts to obtain the credential than it is for dropouts with lower average levels of the productive traits in question. The fees for taking the GED exam are so low (ranging from no charge to $30 in 1991) as to effectively be zero. Thus, the costs required for the signaling hypothesis to hold must be in terms of the opportunity costs associated with time spent in GED preparation and/or in the "psychic" costs associated with motivating oneself to prepare for and take the seven and three-quarter hour battery.

[15] Boesel, Alsalam, and Smith's extensive survey of the literature suggests that over 90 percent of post-secondary institutions permit students to matriculate in degree-granting programs without a traditional high school diploma, but they do require alternative certification such as the GED (1997). It is important to distinguish degree-granting programs from non-degree-granting programs at post-secondary institutions, since many community colleges have, for example, GED preparation programs, and other offerings that are not to be considered as true post-secondary

most common credential obtained by dropouts (Cameron and Heckman 1993 p. 4). To the extent that dropouts use a GED to access post-secondary education, training, and funding that would otherwise be denied to them, and to the extent that post-secondary education and training impact labor market outcomes, GED-acquisition will affect labor market outcomes.

Our tabulation of *High School and Beyond* data indicate that about 29 percent[16] of dropouts with a GED enroll in a post-secondary institution. However, individuals that drive the estimates in this paper are those who are on the margin of passing the GED exams, and so a more appropriate comparison group in the *High School and Beyond* would be those GED-holders who score in the lowest quartile of the sophomore math test in the *High School and Beyond* data. In this group, about 21 percent of the GED-holders get any post-secondary education and they obtain, on average, about 1 year of post-secondary education. Using these figures along with a 5 percent return to a year of post-secondary education,[17] this "gatekeeping" function can, at most, only explain about 1 percentage points out of our estimated 10–19 percent treatment effect. Furthermore, as we will show in the next section, our natural experiment methodology results in there being no human capital component to our estimates. As a result, our estimates primarily represent the signaling value of the GED in the labor market, and as such, they represent the only empirical estimates of the returns to a labor market signal.[18]

To establish a baseline for the relationship between GED acquisition and labor market

---

education.

[16] The *National Longitudinal Survey of Youth* (NLSY), which first surveyed a sample of 14-21 year-olds in 1979, has self-reported information on post-secondary enrollment. Previous research using these data suggests that between 30 and 38 percent of GED recipients had enrolled in a two- or four-year college by age 28 (Cameron and Heckman 1993; Gareth, Jing, and Kutner 1995).

[17] See Kane and Rouse (1995).

[18] Lang and Kropp (1986) use compulsory schooling laws to test a human capital vs. sorting explanation of years of completed schooling. However, their work does not attempt to estimate a return to the signal provided by another year of schooling. For other discussions of human capital versus signaling explanations of educational attainment see Lang (1994), Weiss (1995), Spence (1973), and Stiglitz (1975).

outcomes, we turn to the October 1993 CPS.[19] Table 1 displays the results of log wage and employment regressions fitted separately for young white and nonwhite dropouts.[20] In estimates that control for gender and region of the country, the second row of Table 1 indicates that young white GED-holders have wages approximately 14 percent higher than white dropouts without a GED, and they have an 8 percentage point greater chance of being employed. Meanwhile, young nonwhite GED-holders enjoy no measurable advantage over nonwhite dropouts without a GED in either wages or probability of employment. While these baseline correlations between GED attainment and labor market outcomes are interesting and informative in their own right, potential self-selection bias precludes a causal interpretation. To reduce or eliminate potential self-selection bias in estimates of the effect of the GED on outcomes we must turn to our natural experiment framework.

Section 2.
Interstate Variation in GED Passing Standards

For a simple example of how we exploit interstate variation in GED passing standards, consider New York and Connecticut. Since 1985, New York has maintained a higher passing standard than has Connecticut. In any given year a subset of dropouts who took the GED in Connecticut achieved scores just high enough to obtain the credential that year. In the same year, a subset of dropouts who took the GED exams in New York achieved exactly the same scores as the Connecticut subset, but none of these individuals obtained a GED under the higher New York standards. Thus, individuals in the Connecticut and New York subsets have the same GED test-scores and individuals in both groups all chose to try to acquire a GED. However, dropouts in the Connecticut subset possess the GED credential while dropouts in the New York subset do not,

[19] The October CPS is the only Current Population Survey or Decennial Census survey that distinguishes GED-holders from regular high school graduates.

[20] We fit separate regressions for whites and nonwhites because, as we explain later, this is the approach used for the main analyses reported in this paper.

and as a result, among this subset, we can compare similarly skilled individuals who all desire a GED, but who differ in GED status as a result of their state of residence. Once we have accounted for potential differences in state labor markets, we can use this variation in GED-status to learn about the impact of the GED on earnings by comparing the mean earnings of these two groups. In this case we are learning about the average effects of the GED on the earnings of dropouts who express a desire to obtain a GED and whose scores place them on the margin of passing the exams.

All of the states in our sample define their GED passing standards in terms of a combination of the minimum and the mean (or total) scores. For example, in 1990 New York and Florida had relatively high passing standards by requiring that dropouts have a minimum score of at least 40 *and* a mean score of at least 45. Texas, Louisiana, Mississippi, and Nebraska had the lowest standards in 1990, requiring dropouts to have either a minimum of at least 40 *or* a mean score of at least 45. The other, medium, GED passing standard used in this paper was the most widely used standard in 1990, and it required a minimum of at least 35 and a mean of at least 45. Our research requires that we use different minimum-mean score combinations to construct different GED score-groups. Table 2 shows the results of that process.

The rows of Table 2 represent different, sequential ranges of minimum test scores and the two columns represent two divisions of mean scores. To fit our research design, we have constructed ten GED "score groups" from certain combinations of these minimum and mean score ranges. Score group 1, the lowest ranking score group, consists of individuals whose minimum GED score out of the five tests was between 20 and 34, inclusive, and whose mean score was less than 45. Meanwhile, score group 10 contains the highest scoring test-takers, those whose minimum score is 53 or above and whose mean is greater than or equal to 45. The rank ordering of the score groups is obvious from the layout of the table, except for score groups 3 and 4, but a moments glance makes it clear that both score groups 4 and 5 are higher hurdles than score group 3.

For our research, the critical score groups in Table 2 are score groups 3 and 4. These are the "affected score-groups" for two different natural experiments we will exploit. In each of these two scoring ranges individuals in lower passing standard states—the "treatment states"—have a GED, while individuals in the same scoring range in "comparison states" do not have a GED due to higher state passing standards. When we generate estimates that use score group 3 to form the experiment, the treatment states are Texas, Louisiana, Mississippi, and Nebraska. For simplicity, we will call this Experiment 3. Score group 3 is composed of individuals whose 5 scores are hovering around the 45 mark, but with no scores high enough to bring their overall mean score up to 45. Score group 4 is composed of individuals whose overall mean is 45 or above, but who have at least one relatively low score in the 35-39 range.[21] When our estimates use an experiment where score group 4 is the affected score-group, called Experiment 4 for simplicity, 28 states are included in the treatment-state group.[22]

In experiments that generate our primary estimates, individuals from New York and Florida comprise the comparison group. We explain the reasons for using New York and Florida as the comparison states in the data section of this paper.

In Table 2, score groups 1 and 2 are unambiguously composed of individuals with low cognitive skills, as measured by the GED battery, and none of the individuals scoring in these ranges would have a GED in any of the states. Score groups 5-10 are composed of individuals with increasingly higher levels of cognitive skills, and individuals in these score groups would have a GED in every state. There is only variation in GED status in score groups 3 and 4, and this variation depends upon the state in which one takes the GED.

---

[21] On average, the lowest scores for individuals are on the math and the writing tests (Baldwin 1992, and the authors' tabulation of the data in this study).

[22] Although Connecticut would qualify as a treatment state in Experiment 4, we do not include data from this state in forming our estimates. The reason is that we have the universe of test-takers from Connecticut, but not from the other Experiment 4 treatment states, and we believe that it is important to keep the definition of the treatment-state sample consistent. We note, however, that the point estimates and standard errors change very little in analyses where we have included Connecticut. Furthermore, we do compare outcomes in Connecticut and New York in a later section of the paper as a robustness check on our primary estimates.

In our pseudo-experiments using dropouts who have all indicated a desire to acquire a GED, we use score groups 3 and 4 to compare the earnings of individuals whose scores on the exams are the same, but who differ in GED status as a result of the state where they attempted the exams. As a result, individuals who contribute to our estimates are plausibly balanced on pre-test factors such as motivation to take and prepare for the tests and average levels of skills as measured by GED test-scores. Furthermore, as we showed at the end of Section 1, lower-scoring GED-holders—the individuals driving our results—engage in very little post-secondary education. Thus, our estimates of the impact of the GED on earnings primarily measure the signaling value of the GED in the labor market.

Section 3.
The Data

The data required to exploit the natural experiments described above must contain basic demographics, GED test-scores, year tested, state in which tested, and an outcome measure such as earnings. To obtain a data set containing all of this information, we have constructed a merged data file that combines information from GED test files and from Social Security earnings files. The GED test information comes from administrative data of the GED Testing Service and the state education agencies of New York, Florida, and Connecticut. Using Social Security numbers GED test-takers supplied at the time of the test,[23] the GED files were merged with Social Security annual earnings data by programmers at the Social Security Administration (SSA).[24]

---

[23] While providing SSNs on the GED test form is voluntary, our examination of the Florida and Connecticut data indicate that about 98 percent of examinees 21 and younger (the target group in our study) provide their Social Security numbers.

[24] Angrist's work, particularly his research into the effects of voluntary military service on labor market outcomes (1997), has been very influential and informative in the development and execution of our own project. Our paper is conceptually close to the 1997 Angrist paper, and his experience in coordinating with Social Security Administration programmers in efforts linking individuals to their Social Security earnings, grouping the resulting micro-data to satisfy SSA

3.1 Descriptive Summary of the Underlying Individual-Level Data

Table 3 presents details of the individual-level GED administrative data on which our analytic sample is based.[25] In data obtained from New York, Florida, and Connecticut, we possess information on the universe of dropouts who attempted the GED tests in each of these states in 1990, subject to Social Security number validation.[26] For the data obtained from the GED Testing Service (GEDTS), we have—from each of 42 states—a sample of dropouts who attempted the GED battery in 1990. Later we discuss the critical role played by the universe of 1990 GED candidates in New York and Florida in our research design. For now, we focus on the fact that according to Table 3, most GED candidates are young, (around 50 percent of the dropouts who attempt the GED are between 16 and 21 years of age, with most being 18, 19, or 20), mostly white (two-thirds), and about equally divided between males and females. Minimum GED test-scores in rows 14-18 of Table 3 support earlier GED Testing Service reports showing that dropouts find the math and the writing tests the most difficult tests in the battery. The percentage of 1990 test-takers who scored high enough to obtain a GED (in row 19) is roughly comparable in the different data sets to the 70 percent reported passing by the GED Testing Service in 1990 (GED Testing Service 1991a). Finally, the last two rows in Table 3 indicate the percentage of each data set with valid Social Security numbers (SSNs)—a necessary condition for use in our analyses.

---

confidentiality requirements, and then using the released aggregated data to answer research questions has been invaluable to use. We are indebted to him for his help.

[25] As we will explain shortly, we ultimately use aggregate data in our analyses as a result of SSA regulations regarding the nature of data that they will release.

[26] We explain the SSN validation process in the Data Appendix. As Table 3 shows, SSN validation is quite high for the Connecticut, Florida, and New York data, and thus for expository simplicity, we discuss these data in the paper as if we had the entire universe of test-takers from those states.

Based on information provided to us by the SSA regarding valid ranges of SSNs extant in the U.S., we identified SSNs that were bogus and eliminated these observations before we sent the data to SSA. Row 20 gives the percentage of potentially valid or "good" SSNs in the data after these initial inspections.

After preliminary analyses and identification of "good" SSNs, we sent the data to SSA, which attached Social Security-taxable earnings. As indicated in the last row of Table 3, except for the special case of New York (whose lower match rate is explained in the Data Appendix), the programmers at SSA were able to provide valid matches[27] and locate earnings for 90 to 95 percent of the individuals in the administrative GED data. The resulting micro data set contained information collected at the time dropouts in the sample attempted the GED, together with Social Security earnings records for each year from 1980 to 1995.

3.2 Construction and Descriptive Summary of the Aggregate, Released Data Set

As detailed by Angrist (1997), SSA earnings data have a number of limitations.[28] One important limitation is that the SSA does not release individual earnings to researchers. Therefore, the micro data set produced by the SSA that linked 163,521[29] GED-attempters to their earnings was the starting point for SSA-construction of an aggregated data set that would satisfy both SSA confidentiality requirements and our research design. The initial aggregated data set consisted of 2,138 cells defined by state, GED score, demographic group, age group, and two variables not used in this project. Out of these 2,138 cells, 161 cells failed to meet a last SSA confidentiality

[27] SSNs are not assumed to be valid by the SSA. Rather, when a match in the SSA data base was found for an SSN in the administrative GED records, the demographic information in the two data sets had to match according to a pre-determined algorithm. See the Data Appendix for a discussion of the SSN validation process.

[28] We discuss limitations regarding top coding, covered occupations, and other aspects of the SSA earnings records in the Data Appendix. Our overall conclusion is that these limitations do not provide serious problems for this research.

[29] The SSA actually linked 582,787 individuals to SSA earnings. This higher number represents the extra individuals who took the GED in years 1984-1989 in Connecticut, Florida, and New York. This paper concentrates on those tested in 1990.

edit,[30] and thus the data set released to us by the SSA contains cell-identifiers for 1,977 cells, along with the cell frequency, average FICA earnings, and the standard deviation of earnings for each of the years 1980-1995.

Because the SSA will only release aggregate data, we had to make choices as to how we wanted SSA programmers to aggregate the micro data set. Coarse stratification of the micro data would result in the loss of much individual-level information. If our aggregation scheme stratified the data too finely, many small cells subject to censoring would result. The key stratifying decisions centered on test-taking age and race. The best balance between cell size and retention of individual-level information resulted in stratifying race into two categories (white and non-white) and test-taking age into two categories (those who were 16-21 when the test was attempted and those who were 22-and-older when the test was attempted).[31] The full list of stratifying variables is presented in Table 4.

Because we must stratify rather coarsely on age (2 values), there is the possibility that age heterogeneity may influence results which use the very heterogeneous older age group—those 22 and older in the year of the test. For this reason, our focus in this paper is on the impact of the GED on earnings of dropouts who were age 16-21 when they attempted the GED. In what follows then, all sample references are to the released aggregated data set of dropouts who attempted the GED in 1990, when they were 16-21. Table 5 presents summary statistics for this

---

[30] SSA confidentiality regulations require that no information be released in any cells where there are fewer than 3 observations. The data released by the SSA were also subject to a confidentiality edit that masks earnings information if other cell-related criteria are not met. These additional edits had little effect on our data and are detailed in the Data Appendix.

[31] We chose age 21 and younger as the cutoff both because of the observed distribution of age at testing displayed in Table 3 and because well known evaluations of JTPA programs have made a youth/adult distinction at age 21. See for example the 30-month follow up evaluations of the JTPA in Orr et al (1996).

group.[32]

The first thing to notice in Table 5 is that because we have the universe of young dropouts who attempted the GED in 1990 in New York, Florida, and Connecticut, the relative number of observations contributed by those states is large, as shown in column 1. The analytic sample is considerably more white than nonwhite (column 5)[33] and slightly more male than female (column 6).

Columns 7 and 8 present the mean earnings five years after our sample of dropouts attempted the GED battery, and these columns indicate that the mean earnings of dropouts are generally low.[34] While not reported in this table, the variance of 1995 earnings for any particular subset of our sample are generally large relative to the mean, as is typical with annual earnings. The variability of annual earnings in our data, combined with issues of statistical power, contributed to decisions we made regarding the aggregation of males and females in our analyses. We discuss this aggregation in a later footnote.

---

[32] There are 990 cells with 3 or more observations in this data set, and each cell has attached Social Security earnings information for the years 1980-1995.

[33] While the percentage white is very high in some states, we have the universe of young 1990 test-takers in Florida, a heavily Hispanic state where three-quarters of the young dropouts who attempt the GED are white. We further note that our aggregate percentage-white shown in the last row compares favorably with GEDTS figures from a nationally representative sample of 1989 test-takers that indicate that 70 percent of all attempters that year were white (Baldwin 1990). Finally, Department of Education statistics show that in the mid- to late-1980s, when individuals in our sample would have been dropping out of school, about 66 percent of the secondary school dropouts were white (National Center for Education Statistics 1996, Table 101).

[34] Notice also that the earnings in New York are surprisingly low compared to other states. We have additional data at our disposal to examine the possible causes of these low earnings in New York. In this additional data, we find that the *positive* earnings of dropouts in New York are the highest of any state, indicating that in our data, many young dropouts in New York had zero earnings in 1995.

Columns 2, 3, and 4 in Table 5 designate whether a state falls into the treatment-group or comparison-group category in each of the experiments involving score groups 3 and 4. These columns illustrate how some states contribute to the treatment group in one experiment and to the comparison group in the other experiment. For example, individuals in score group 4 in Arizona are in the treatment category for Experiment 4, while individuals in score group 3 from the same state are in the comparison group of Experiment 3*. Columns 2 and 3 show that the only treatment states in experiments 3 and 3* are Texas and the states in the "Low passing states" group. On the other hand, column 4 shows that in Experiment 4, the only comparison-group states are New York and Florida. There are obviously many different interesting across-state comparisons one could examine. However, contrasting only two states reduces our sample sizes considerably and limits our statistical power for making inferences about treatment and comparison-group differences.

Because of issues having to do with "retaking" the GED that we discuss in the Data Appendix, New York and Florida form the comparison group in two of the three experiments. We single out these states because the New York and Florida data contain only the accumulated best scores of individuals who either have attained a GED or who have "stopped out" as of 1990.[35] The 1990 GEDTS data, however, contain only a one-year "snapshot" of GED-attempters. Thus, some portion of the individuals with non-passing scores in the GEDTS data will have subsequently retested and passed the battery in the years 1991-1995. As a result, any comparison group constructed from the GEDTS data will be contaminated with individuals who, in fact, subsequently obtain a GED. Because of this, estimates of the effect of the GED on earnings where the GEDTS data is used to construct comparison groups will estimate a lower bound treatment effect.[36]

---

[35] By this we mean that we are certain that non-passers in the New York and Florida data have not returned after 1990 to retake the exams and potentially receive a GED as of 1995, when we measure earnings. In these states, anyone who took any portion of the tests after 1990 would not be included in the 1984-1990 data, regardless of their passing status.

[36] For example, using the Connecticut data, we find that out of the 1,641 16–21 year-old dropouts who first tested in 1985, 66 percent passed on the first try. Out of those who did not pass, 18 percent had retested and acquired a GED within two years and after four years a total of

Section 4.
Methods

The goal of the empirical work is to identify the effect of the GED on the earnings of young dropouts. Past research efforts in this area have not been able to effectively control for unobserved differences between dropouts who chose to acquire the credential and dropouts who do not. Since everyone in our sample of dropouts has attempted the GED battery of exams, our sample construction controls for differences between those with and without the credential that originate in the decision to acquire a GED. To further balance the attributes of the comparison groups of interest, we use the natural experiments created by interstate variation in GED passing standards to compare a subset of individuals in the data who have the same GED test-scores, but who differ in GED-status. This comparison is captured in the following simple model describing the annual earnings of dropouts in our data:

$$Y_i = \beta_0 + \beta_1' ST_i + \beta_2' SG_i + \alpha(T_i * ASG_i) + \beta_3 Female + u_i, \tag{1}$$

where

Y = individual i's annual earnings in some year after the GED was attempted,

ST = a vector of state dummies,

SG = a vector of dummies representing the GED score group individual i is in,

T = 1 if individual i is in a (lower-standard) treatment state, and 0 otherwise,

ASG = 1 if individual i is in the affected score-group and 0 if individual i is in a higher

---

25 percent of the original non-passers had acquired a GED. The result is that the pass rate went from 66 percent on the first try in 1985 to an eventual pass rate of 76 percent by the end of 1989.

score group,[37]

Female $= 1$ if individual i is a female, 0 if a male, and

u $=$ an error term with $E[u_i|ST, ASG]=0$ and $u_i \perp (ST, ASG, ST*ASG)$.

The parameter of interest, α, represents the effect of the GED on the earnings of dropouts whose GED-status is affected by state-of-residence. It is straightforward to show that α is equal to

$$E[Y_i|T_i{=}1,ASG_i{=}1] - E[Y_i|T_i{=}0,ASG_i{=}1] - \big(E[Y_i|T_i{=}1,ASG_i{=}0] - E[Y_i|T_i{=}0,ASG_i{=}0]\big). \quad \textbf{(2)}$$

That is, the effect of GED-acquisition on the earnings of individuals in the affected score-group can be obtained from a contrast of the mean earnings of individuals from treatment and comparison-group states who are in the affected score-group compared to a contrast of the mean earnings of individuals from treatment and comparison-group states who are in a score-group above the affected score-group. Based on (2), a method of moments estimator of α can be constructed from sample means.[38] This difference-in-differences estimator (DD)[39] is $\alpha_{DD}=(\bar{Y}_T-\bar{Y}_C)-(\bar{Y}_{THi}-\bar{Y}_{CHi})$ where,

T indexes individuals in the treatment group, that is, individuals in the affected score-group who are in a low-passing-standard state in a given experiment, and thus have a GED,

C indexes individuals in the comparison group, that is, individuals in the affected score-group who are in a high-passing-standard state, and thus do not have a GED,

THi indexes individuals in treatment (low-standard) states that are in a scoring group

---

[37] We do not use individuals who score in ranges below the affected score-group range in our analyses because, in the GEDTS data, these score-groups are contaminated with individuals who have a GED.

[38] The standard errors for the estimator are obtained in the standard manner, with no assumptions of equality of variances across groups.

[39] DD estimators go back at least as far as Campbell (1969) in psychology and have been used recently in economics by Meyer (1989), Card and Krueger (1994), and Gruber (1994) among others.

higher than the affected score-group, and thus have a GED,

CHi indexes individuals in comparison-group (high-standard) states that are in the same *high scoring* group as the THi group, and thus also have a GED,

$\overline{Y}_T$ = the mean earnings of the treatment group,

$\overline{Y}_C$ = the mean earnings of the comparison group,

$\overline{Y}_{T\,Hi}$ = the mean earnings in the T Hi group, and

$\overline{Y}_{C\,Hi}$ = the mean earnings in the C Hi group.

The identifying assumptions for the DD estimator follow from equation (1). First, our main assumption—that the natural experiments generate exogenous variation in GED-status—is captured in the assumed independence between the interaction term and u_i. Second, equation (1) allows for no higher order interactions. As Meyer (1995) points out, the possibility of interactions is one of the main threats to the validity of inferences from a difference-in-differences design. Finally, the nature of our data forces us to model the determinants of earnings levels, rather than the determinants of the natural logarithm of earnings. Thus, we must assume either that the correct model is specified in levels, rather than in logs, or that the distributions of earnings in comparison groups are such that equation (1) is an appropriate model in either levels or logs.[40] Later in the paper we examine the empirical evidence regarding violations of each of these main assumptions and find that any resulting biases have little effect on our estimates.[41]

---

[40] For a discussion of the relationship between the distributions of outcomes and the suitability of levels versus log models in a difference-in-differences design, see Meyer (1995).

[41] To account for the possibility that different treatment–comparison-group gender distributions or different distributions across states within the treatment or comparison groups could affect our results, we weight the data so that the weighted mean earnings in each group that comprise the DD estimator reflect balance along the key dimensions.

In the context of heterogeneous treatment effects, Angrist (1997) discusses the difference between the weighted α_{DD} estimate of α (which he calls a "controlled contrast") and the regression estimate of α (call it α_r) in the presence of covariates (such as gender). Angrist's discussion is relevant to our work if we assume that there are group-specific constant treatment effects (e.g. α_{males} and $\alpha_{females}$). In this case, Angrist's work can be used to show that the weighted

Section 5.
Results

The main research question addressed in this paper is whether acquisition of a GED has an impact on the earnings of dropouts. We identify this effect using plausibly exogenous variation in GED-status generated by interstate variation in GED passing standards. For a subset of GED test-takers in the data who are affected by this state passing standard variation, we can identify the effect of the GED on earnings, and our main findings for this group are that:

- *acquisition of a GED increases the annual earnings of young white dropouts by 10-19 percent, but it may take up to five years to see gains of this magnitude, and*

- *we find no statistically significant impact of the GED on the earnings of young nonwhite dropouts five years after GED receipt.*

These results are stable across several experiments that use different treatment and comparison groups, and they withstand sensitivity analyses that explore possible violations of our identifying assumptions.

5.1 DD Estimates of the Impact of the GED on the Earnings of Young Dropouts

The DD estimator that we use, $(\overline{Y}_T-\overline{Y}_C)-(\overline{Y}_{THi}-\overline{Y}_{CHi})$, requires a second difference, $(\overline{Y}_{THi}-\overline{Y}_{CHi})$, that is constructed over dropouts in the treatment and comparison states who all have scores above the GED passing threshold. For simplicity, we call these groups the high comparison-groups. We could use any of the score groups 5 through 10 to form the high comparison groups, since all individuals in these score ranges have a GED regardless of the state. We have chosen to use all of the individuals in score groups 5 through 10 to form the high comparison groups in our

---

α_{DD} is not necessarily equal α_r. Our own comparisons of α_{DD} and α_r show that the story is qualitatively the same with either estimator.

Also, in separate analyses, we reproduce all of the results of this paper using different weighting schemes. The results are not sensitive to the choice of weights.

DD estimates to avoid making arbitrary decisions about which score group to use and to maximize the number of observations in the high comparison groups.[42] Using this specification and the DD estimator, the central results of this paper are shown in Table 6. In that table we present separate estimates for white and nonwhite dropouts[43] from three different experiments involving young dropouts who attempted the GED in 1990. For these dropouts, we measure the effect of the GED on their 1995 earnings.

We focus initially on the estimates associated with Experiment 4 and Experiment 3. In these experiments, acquisition of a GED is associated with about a $1,500 increase in annual earnings for young white dropouts, which translates into a 19 percent increase in earnings.[44] The point estimates for whites associated with the two different experiments ($1473 and $1531) are remarkably similar given that the states that comprise the treatment group in each experiment are completely different. Because the estimates in Table 6 are based on a pooled (male and female) sample, it could be that the underlying treatment effects for males and females in the two experiments are very different, even though the pooled estimates are similar. This, however, turns out to not be the case as the figures below show. When estimated separately, the white male and female estimates and standard errors for Experiments 3 and 4 are:

---

[42] Separate analyses verify that our results are not qualitatively different when we use any one of the score groups 5-10 as the high comparison group, as opposed to using all individuals in score groups 5-10 simultaneously. See Table 10 for the results for whites.

[43] Preliminary analyses conducted separately for white males, white females, nonwhite males, and nonwhite females indicated that results for white males and females were similar and that results for nonwhite males and females were similar. Thus, to increase sample size and statistical power, the analyses in this paper aggregate by gender and not by white–nonwhite.

[44] The percentage increase is calculated as (DD Est.)/$\overline{Y}_C$.

	White Males	White Females
Experiment 4	1657*	1195
	(937)	(821)
Experiment 3	1649*	1430**
	(861)	(546)

Both Experiment 3 and Experiment 4 use New York and Florida as comparison-group states. Experiment 3*, like Experiment 3, uses Texas, Louisiana, Mississippi, and Nebraska as treatment states, with the members of score group 3 in these states forming the treatment group. Instead of using New York and Florida as comparison-group states, however, Experiment 3* uses states from the GEDTS data in the comparison groups. As we discussed at the end of the Section 3, we know that when the GEDTS data are used to construct the comparison group, the comparison group is contaminated by some individuals who actually have a GED. Because of this, the 1995 earnings of the comparison group will be "artificially" high and the DD estimate will be a downwardly biased, conservative estimate.

On the other hand, since the scores of the comparison group in Experiments 3 and 4 are based on the accumulated best scores in the New York and Florida data, rather than "one shot" scores, the comparison groups in these experiments may be slightly less skilled than the treatment group. As a result, the estimates in Experiments 3 and 4 may be upwardly biased,[45] and thus, Experiment 3* can be interpreted as a lower bound estimate and Experiments 3 and 4 as upper bound estimates of the GED treatment effect. Taken together, these three sets of results provide strong evidence that we are doing a reasonable job of estimating the true underlying population parameter.[46]

---

[45] However, DD estimates constructed using Connecticut and New York that we present later in Table 8 suggest that there may be little or no upward bias. The relevance of the Connecticut–New York comparison for this discussion is that scores in both of these states are constructed from accumulated best scores.

[46] In Appendix C we discuss the limited evidence we have regarding how much of these earnings effects work through increased wages vs. through increased employment. Our tentative conclusions in that discussion are that the GED does not affect the persistent non-employment (percentage of individuals with zero earnings in a year) of white males, but that a part of its affect

The results for nonwhite dropouts are very different from the results for white dropouts. The three experiments yield no statistically significant evidence that acquisition of a GED results in higher earnings for nonwhite dropouts. We will return to the nonwhite results later in the paper, but we note at this point that the pattern of estimates we obtain using a natural experiment approach match closely the pattern of estimates displayed in Table 1 using CPS data: namely, substantial treatment effects for whites and small and statistically insignificant effects for nonwhites.[47]

Our estimates for whites in Table 6 are formed over a sample of dropouts who have all indicated a desire to acquire a GED. Furthermore, the effect of the GED on earnings is identified for a subset of the data whose scores place them on the margin of passing the GED exams. As a result, our estimates should be interpreted as the effect of the GED on the earnings of young dropouts who attempted to acquire a GED in the labor-market and GED-program environments of the early 1990s[48] and whose cognitive skills place them on the margin of passing the GED exams. We believe that our research design and data make our results the best estimates to date of the effect of the GED on the earnings of young dropouts, and the first estimates of the signaling value of the credential.

We have shown that our estimates are robust to the use of different treatment and comparison groups. As a further robustness check on our estimates in Table 6, we obtain estimates from experiments that contrast two states that are geographically close. The attractiveness of such experiments are that since only two states are involved, it is very clear what groups are being compared, and it is clear what state effects are being removed by the DD estimator. The first experiment in these dual-state estimates compares Connecticut (a lower

---

on increasing the mean earnings of white females is by decreasing the percent who are persistently unemployed.

[47] It is important to keep in mind that the Table 1 estimates use log wages and probability of employment as the dependent variables, while the main results in this paper use annual earnings as the dependent variable.

[48] The GED program environment has recently changed. As of January 1, 1997, the minimum passing requirements recommended by the ACE were raised to those reflected by the high standard states in this project—a 40 minimum score and a 45 mean score.

standard state) and New York (a higher standard state). This experiment is especially appealing because we have the universe of 1990 test-takers from each of these states, and it can thus serve as a sensitivity check regarding the extent to which our previous estimates might be influenced by the particular samples drawn in the GEDTS data. Two other dual-state comparisons we present are Texas vs. Florida and Texas vs. Arizona.[49] The estimates from these dual-state comparisons are in Table 8, and the similarity of these point estimates to those of the larger experiments are striking. We take this as additional evidence that our natural experiment approach yields estimates of the underlying population parameter of interest.

While small sample sizes in these dual-state experiments leave us with point estimates that are not precisely estimated (except the Texas-Florida experiment), the Connecticut vs. New York and Texas vs. Arizona point estimates are very similar to those in Table 5 for young whites. The overall story of Table 8 is consistent with our earlier results: young whites dropouts receive substantial monetary benefits from GED acquisition.[50]

5.2 Timing of GED Treatment Effects

As we explained in Section 1, our reliance on natural experiments means that we can only measure the effect of the GED on earnings as it works through signaling or gatekeeping mechanisms. As a result, it is likely that it might take some time after receipt of the credential for any effects on earnings to show up in our data. This is the primary reason we have concentrated to this point on the 1995 earnings of young dropouts who attempted the GED in 1990. To

[49] It is important to note that while the dual-state experiments involving Texas–Florida and Texas–Arizona are subsets of the multi-state Experiment 3*, the Connecticut–New York comparison is one that has not been conducted to this point, since Connecticut is not used in Experiment 4 or Experiment 3.

[50] The point estimates and standard errors for nonwhites in these dual-state experiments support the nonwhite results in Table 6.

explore the possibility that benefits accruing from GED-acquisition do indeed take time to develop, we construct DD estimates using earnings in the years 1991-1995 as the comparison measure. These comparisons continue to use 1990 test-takers as the analytic sample, and we concentrate on white dropouts.

Figures 1-3 show the year-by-year DD estimates for experiments 4, 3, and 3*.[51] Figures 1 and 2 suggest that it may take time for the GED to pay off for young white dropouts. Figure 1 shows that in the first two years after GED acquisition, GED-holders actually earn less than dropouts with the same GED scores, but who do not have a GED. Over time GED-holders in the treatment group begin to gain ground on their uncredentialed counterparts in the comparison group, so that by the fifth year after GED-acquisition, they are earning $1473 more per year. We take this pattern of growing treatment–comparison-group differences in earnings to indicate that newly minted GED-holders are engaged in post-secondary education, training, or job search.[52] It appears that the treatment individuals in Experiment 3 are able to make the GED pay off sooner than the treatment individuals in Experiment 4, but the trends are the same.[53]

The timing of payoffs to the GED estimated from Experiment 3* are displayed in Figure 3. These comparisons between treatment and comparison individuals in the GEDTS data also show some growth in the GED premium, but there are substantial payoffs in the first year in these comparisons.

---

[51] Later in the paper we discuss earnings in the pre-treatment years.

[52] We have established that GED-holders get relatively small amounts of post-secondary education on average. However, those that do may take a relatively long time to acquire credits (e.g., via night courses) and they may remain in a relatively low paying "day" job during their post-secondary experience. Such behavior could depress the average earnings of the treatment group during the first years after GED receipt.

[53] The trends are roughly the same in Experiment 3*, but the presence of different numbers of "contaminees" entering the comparison-group each year make interpretations of the time trend of differences in Experiment 3* problematic.

5.3 Threats to Identification

In this section we examine possible violations of our basic assumptions and the implications for our estimates. As a road map to this section, Table 9 summarizes the series of specification tests we discuss. The table presents the potential threats to identification, the likely direction of bias, the method by which we assess the threat, and the results of each particular test.

Pre-treatment earnings comparisons.

In general, a necessary, but not sufficient condition, for the validity of inferences drawn from any experiment is that there be no pre-treatment differences between treatment and comparison outcomes. Figures 1a-3a reproduce Figures 1-3, this time adding the DD estimates from the two years prior to treatment.[54]

Figures 1a and 3a indicate that there are no statistically distinguishable pre-treatment differences in the mean earnings of the treatment and comparison groups in Experiment 4 and Experiment 3*. In particular, the point estimates of the pre-treatment earnings comparisons in Experiment 3* ($86 and $47) are very close to zero. In contrast, the pre-treatment comparisons in Experiment 3 are negative and statistically significant. This suggests that we might expect to see the treatment group faring worse that the comparison group in the absence of treatment, indicating that attempts to estimate a treatment effect using these groups may be downwardly biased. Thus, pre-treatment contrasts either support the comparability of the treatment and comparison groups or suggest that, at least in Experiment 3, estimates of the treatment effect of

[54] In interpreting Figures 1a-3a, it is important to keep in mind the young age of our sample of dropouts. In the two years before treatment, members of our sample were 14–19 and then 15–20 years old, and the mean earnings of any one group in this period was only $1,000–$3,000.

the GED may be conservative.

Assessing the case for endogeneity bias.

There are two types of endogeneity that could bias our results. First, states could set their passing standards in response to factors related to the potential labor market outcomes of dropouts. For example, if low-standard states set low standards because the average skills of their dropouts are low, and if dropouts in these states are paid relatively less well than dropouts in high standard states, then the setting of the standard is obviously not exogenous. This type of endogeneity, however, is simply a state effect in our model and will be accounted for in our DD estimates. More complicated types of endogenous standard setting would appear as state–GED or state–skill interactions, problems we address below.

The second type of endogeneity involves individuals who possibly "endogenize" the GED passing standards. That is, it may be that when faced with different passing standards, individuals react in ways that may affect the composition of our treatment and comparison groups. In particular, individuals may sort themselves, to some extent, into different score groups if faced with different passing standards.[55] This supposition is obviously important since our basic assumption is that individuals in our treatment and comparison affected-groups are, on average, identical except in terms of GED status (i.e., $u_i \perp T_i{*}ASG_i$). The primary concern is that individuals in the affected score group in treatment states may have higher skills than individuals in the affected score group in comparison states *because* of the fact that the states have different standards, a situation that would lend an upward bias to our basic findings.

To assess the seriousness of this type of bias, we make assumptions that generate three different levels of skill-mismatch between the treatment and comparison groups. Using these three different levels of potential bias, we can generate a range of "adjusted" estimates that account for

[55] It is also likely that, faced with higher standards, some individuals on the margin would elect to not attempt the tests. Assuming that this type of behavior is negatively correlated with productivity-enhancing traits such as persistence, self-confidence, etc., then this type of selection would generate a downward bias in our estimates, yielding conservative estimates of a GED treatment effect.

the fact that individuals may endogenize the passing standard. We relegate the detailed discussion of these adjusted estimates to Appendix C. Our conclusions from the exercise outlined in those analyses are that:

(1) this type of endogeneity bias is less of an issue in Experiments 3 and 3* than in Experiment 4, and

(2) the most plausible adjusted estimates move the Experiment 4 estimates from $1,473 to $1,166, the Experiment 3 estimates from $1,531 to $1,374, and the Experiment 3* estimates from $907 to $881.

Our overall conclusion is that the potential bias from this type of individual-level endogeneity is not severe.

Assessing the case for bias arising from mis-specification or omitted interactions.

Mis-specification

In addition to endogeneity assumptions, we also pointed out in Section 4 that identification hinges on two other basic assumptions. First, either the levels-model is the correct specification relating earnings to other variables, or the distributions of earnings in the comparison groups are such the model is appropriate either in levels or logs. And second, we must assume that there are no omitted, higher-order interactions in equation (1).

Regarding the first assumption, it may well be that the correct specification relating annual earnings to the independent variables in equation (1) is a log-linear relationship. However, applying Meyer's (1995) analysis of the difference-in-differences design to our situation, our best-case scenario in the face of mis-specification is one where both of the high comparison groups have distributions of earnings close to the distribution of earnings of the treatment individuals in the affected score-group. A re-examination of Table 6 offers some evidence that this is the case for white dropouts. Between the three relevant comparison groups (both of the high comparison groups and the treatment group) mean earnings differ by only 2 to 3.5 percent in each of the three experiments. In addition, the standard deviation of earnings (not shown in Table 6) differ between the groups by only 2 to 4 percent. While we have no information on higher moments in the distributions, if these higher moments are also similar across the distributions, then a logarithmic transformation of the dependent variable in equation (1) would likely have little effect on the interpretation given to our findings (Meyer 1995).

The best specification test would compare a DD estimate from the relevant sample means of individual log earnings to the DD estimate constructed over means in the levels. However, our data contain only mean earnings in the levels, so we cannot construct a DD estimator using the mean of individual log earnings for the relevant comparison groups. The best we can do is construct a DD estimator using the log of mean earnings in the data we have.[56] The DD estimates that compare the log of mean earnings across groups in our data do, however, support the idea that equation (1) is appropriate for either levels or logs, as Table 7 shows.

The DD estimates using the log of mean earnings are presented in Table 7. The estimates for white dropouts in Table 7 give very similar overall results as compared to the estimates in levels in Table 6. In Table 6, white dropouts in Experiment 4 showed a statistically significant GED premium of 18.8 percent. The comparable figure in Table 7 shows a statistically significant 17.2 log-point GED premium. For Experiment 3 the comparisons are 19.5 percent in the levels (Table 6) versus 18.1 log-points (Table 7), and for Experiment 3* the comparisons are 10.5 percent in the levels versus 10 log-points. These results, combined with evidence discussed above, which suggests that the distribution of earnings in the relevant comparison groups are similar, lead us to conclude that equation (1) is, in this case, an appropriate model whether Y represents earnings or log earnings.

Omitted Interactions

Potential omitted interaction biases of greatest concern have to do with state–score or state–GED-status interactions. It is straightforward to show in the DD estimator that state–score interactions that favor *high-scoring dropouts in high-standard states* would bias our estimates upward, while interactions that favor *high-scoring dropouts in low-standard states* would bias our estimates downward. State–GED-status interactions that favor *GED-holders in high-*

---

[56] In general, of course, $ln(E[x]) \neq E[ln(x)]$. However, $ln(E[x])$ is a first order approximation of $E[ln(x)]$ (Greene 1993, p. 57).

standard states would bias our estimates upward, while such interactions that favor *GED-holders in low-standard states* would bias our estimates downward.

There may or may not be state–GED-status interactions that affect the outcomes of GED-holders. We have no data that would allow us to identify such an effect separately. However, it is plausible that employers in states where the GED-passing standard is relatively high give more credence to and place more emphasis on the credential than do employers in low-standard states. In this case, *ceteris paribus*, GED-holders would earn more on average in high-standard states than similar GED-holders in lower-standard states, and this type of mechanism would cause an upward bias in our estimates.

Turning our attention to possible state–score interactions, we see no compelling argument that state–score interactions act in one direction or another. Thus, *ex ante*, we cannot sign the bias associated with possible state–score interactions. We can, however, bring *ex post* evidence to bear on the question, by looking at the pattern of estimates that result from using higher and higher GED score groups to form the high comparison groups in the DD estimator. Based on equation (1), if there are state–score interactions that favor high-scoring dropouts in low-standard states, then as we use individuals with successively higher GED scores as high comparisons, the resulting DD estimates should fall. If, on the other hand, there are state–score interactions that favor high-scoring dropouts in high-standard states, then using dropouts with higher and higher GED scores as high comparisons would result in rising estimates.

Our results from an exercise where we use different high comparison groups for whites are in Table 10. Without conducting formal tests, the most obvious patterns in that table are associated with the estimates in Experiments 3 and 3*.[57] In these experiments (excepting when score groups 6 and 13 are used to construct high comparison groups), as successively higher score groups are used to form the high comparisons, the DD estimates get larger. These patterns suggest that if there are state–score interactions at work, then they favor high-scoring dropouts in high-standard

[57] While the DD estimates in Experiment 3* have a consistent downward bias effect due to the contaminated comparison group, this effect is constant as we use different high comparison groups, so there is valid information in the relative differences across different high comparison groups in this experiment.

states and the resulting bias is upward.

The results of these *ex ante* investigations of possible state–score interactions, combined with the most plausible state–GED-status interaction that is the most plausible, yield two possible biases, *both of which would result in an upward bias to our estimates.* To assess how serious these combined effects might be, we take advantage of the fact that in Table 6, we have DD estimates from two different experiments.

The way in which we use the estimates in Table 6 to assess the case for interaction bias lies in the assumption that we expect any upward bias to be greater in Experiment 3 than in Experiment 4. The reason is that Experiment 3 uses treatment states with the lowest standards in the U.S. in comparisons with the high standard states of New York and Florida, while Experiment 4 uses treatment states with "medium" standards in comparisons with New York and Florida. Given the types of state–score and state–GED-status interactions we hypothesize—higher-scoring dropouts *and* GED-holders both faring better in high-standard states—any bias would affect the estimates in Experiment 3 more than the estimates in Experiment 4. Thus, the spread between the Experiment 3 and Experiment 4 estimates gives us some measure of how our results are being affected by the total bias attributable to both state–score and state–GED-status interactions.[58] However, the estimates for the two experiments are less than 4 percent apart ($1473 vs. $1531). Therefore, we conclude that neither type of potential bias can go far in explaining our estimates.

In summary, as with almost any pseudo-experimental setting, there are plausible scenarios under which our identifying assumptions would not hold. For each of our basic assumptions, we

---

[58] This result rests not only on our stated assumption regarding the direction of the bias of the state–test-score interactions, but also on the assumptions that (1) any state–test-score interaction is monotonically increasing in test scores and (2) that at least within the narrow scoring range (score groups 3 and 4) represented by the two experiments, the treatment effect is the same. We think both of these assumptions are reasonable, given that we are looking at a relatively well- and narrowly-defined sample: namely, high school dropouts who score above a certain range on the GED battery of exams.

have examined the most plausible violations, and we have found the overall effects on our estimates for white dropouts to be negligible.

5.4 The Results for Nonwhite Dropouts

We now try to shed light on the results that show no statistically significant GED-effect for nonwhites, and we offer some possible explanations. These explanations fall into either out-of-equilibrium or equilibrium-preserving categories. By out-of-equilibrium, we mean explanations that could explain why we observe the white–nonwhite differences, but could not explain why nonwhites continue to pursue a GED in the absence of any payoff. Equilibrium-preserving explanations both explain the differences and explain why nonwhites continue to obtain the GED certificate.

Out-of-equilibrium Explanations for the White/Nonwhite Differences

Nonwhites are relegated to jobs where skills matter little.

One possible explanation for the differences in our results is that young nonwhite dropouts may tend to be employed in jobs for which basic skills matter little. If this is the case, then employers would value neither the human capital nor the signaling components of the GED, and there would be no payoff to acquiring a GED for nonwhites.

A simple way to assess the plausibility of this hypothesis is to use our data to measure the returns to skills for both whites and nonwhites. If the data show that, controlling for GED acquisition, nonwhites who score high on the GED battery earn no more than low-scoring nonwhites, then this would suggest that nonwhites are in jobs where skills do not matter. In analyses detailed in Appendix C, however, we show returns to skills to be just as high, if not higher, for nonwhites as for whites. Thus, this explanation cannot reconcile the white–nonwhite differences in results that we find.

Employers place a higher value on signals other than the GED for nonwhites.

It is likely that in the absence of more complete information about productive attributes, employers use any signals available that have proven to be reliable predictors of productivity in the past. Employers may find signals other than the GED to be more predictive of the productivity of nonwhite dropouts than of white dropouts. One such possible alternative signal is residential address.[59] It could be the case that employers, especially in urban areas, use residential address as a signal of potential productivity for nonwhite job applicants more so than they do for white applicants. Another possible alternative signal that would affect nonwhites disproportionately is language. To the extent that our nonwhite sample is composed of individuals for whom English is a second language, and to the extent that employers value language as a signal over the GED, we could expect the white–nonwhite results found in our estimates.[60]

Equilibrium-Preserving Explanation

Employers discount the signaling content of the GED for nonwhites because many nonwhites acquire the GED as an "incidental" part of a larger program from which they receive benefits.

This equilibrium-preserving explanation of the results for nonwhites hinges on (1) different behavioral assumptions behind why individuals acquire the credential, and (2) different

---

[59] We are indebted to Mary Jo Bane for this point.

[60] For this explanation to hold, there would have to be enough English-second-language GED-holders in our nonwhite sample to offset the presumed positive treatment effects of the GED enjoyed by nonwhite GED-holders for whom English is the first language. Examination of the underlying individual-level data in our nonwhite sample indicates that the data support this necessary, but not sufficient, condition for the "language as signal" explanation. In our data proportionately more nonwhite GED-holders are Hispanic than black, regardless of the state group.

white–nonwhite distributions associated with those behavioral assumptions.[61] Some dropouts acquire a GED because they either place a value on the credential, or they perceive that employers place a value on the credential. There is another set of dropouts who acquire the credential primarily "incidentally" because it is a "quasi-compulsory" part of a program from which they are seeking benefits. Examples of such programs are Aid to Families with Dependent Children (AFCD), Job Training Partnership Act (JTPA) programs, and the Job Corps.

If the different motivations that lead to acquisition of the GED for these two groups are correlated with productive attributes such as reliability, commitment to work, and general motivation, then the GED would be a signal of productive attributes for the group that actively pursues the signal, and it would be a signal of "welfare training participation" for those who acquired it in an incidental fashion. It is likely that employers place different values on these two signals, discounting the value of the "incidental" GED relative to the "actively-pursued" GED. If the distribution of "actively-pursued" and "incidental" GED-holders is different across whites and nonwhites, then we would expect the signaling returns to the credential to be different across the two groups.

Our best evidence of the impact of quasi-compulsory interventions on the labor market outcomes of economically disadvantaged youths come from a series of experimental evaluations. The youth component of the National Supported Work Demonstration, the JOBSTART evaluation, and the youth component of the National JTPA Study all find a consistent lack of program-induced earnings gains for out-of-school youths (Kemple et al (1992), Cave et al (1993), and Orr et al (1996)). Meanwhile, the GAIN experimental evaluation in California found only small education effects (as measured by GED-attainment and test-score gains) for a large-scale program that required mandatory participation in basic education by low-skilled adult welfare recipients (Friedlander and Martinson 1996). It may be that the apparent ineffectiveness of these programs is at least in part due to their quasi-compulsory nature. This explanation is not only consistent with a "programmatic" explanation for our white-nonwhite differences, but it would also explain why these experimental evaluations find no treatment effects associated with the

[61] We are indebted to Caroline Hoxby for the ideas in this section on the "programmatic" effects on the signaling value of the GED.

particular program (including, for some individuals, receipt of a GED), while we find substantial treatment effects associated with the GED for whites.

This "programmatic" explanation for our white–nonwhite differences requires that (1) a substantial number of GED-holders obtained the credential in the "quasi-compulsory" fashion, and (2) that relatively more of these "incidental" GED-holders are nonwhite than white. To explore these assumptions, we use aggregate AFDC, JTPA, Job Corps, and GED Testing Service statistics from 1990,[62] and we assume that:

- the percentage of nonwhite AFDC recipients in the Job Opportunities and Basic Skills (JOBS) training program was slightly higher than their representation in overall AFDC programs;
- within JOBS and JTPA, nonwhites participated in classroom education versus the other job training offerings at a slightly higher rate than did whites;
- the percentage of dropouts in Job Corps was slightly higher for nonwhites than for whites; and,
- one-half of all individuals in the classroom components of these various programs attempted the GED battery.

Under these assumptions, 11 percent of all white dropouts who attempted the GED battery in 1990 would have done so because the credential was a quasi-compulsory component of AFDC, JTPA, or the Job Corps, while 44 percent of all nonwhite dropouts who attempted to obtain a GED did so because it was tied to one of the above programs. While by no means definitive, these numbers support a "programmatic" explanation of the white–nonwhite differences we find.[63]

---

[62] Our source for the statistics to calculate the needed numbers are Tables 5, 31, 36, 42 of Section 7 and Tables 35 and 38 of Appendix P of the *1993 Green Book: Background Material and Data on Programs Within the Jurisdiction of the Committee on Ways and Means* (1993) and Table 1 of the *1990 GED Statistical Report* (1991a).

[63] It should be noted that all of the out-of-equilibrium explanations could hold in the presence of the "programmatic" explanation.

Section 6.
Summary and Conclusions

Our analyses of dropouts who attempted the GED battery of tests in 1990 show that receipt of a GED increased the 1995 earnings of young white dropouts by 10 to 19 percent. We found no statistically significant evidence that the credential had an impact on the 1995 earnings of young nonwhite dropouts. We are able to identify the effect of the GED on earnings as a result of our research design that uses interstate variation in GED passing standards to generate plausibly exogenous variation in GED-status. For the subset of the data affected by this variation—those dropouts who are on the margin of passing the GED exams—our results are interpreted as estimates of the effect of the GED on the earnings of young dropouts who indicate a desire to acquire a GED. Our estimates are robust across quasi-experiments that use different treatment and comparison groups, and sensitivity analyses we conducted indicate that they are affected very little by possible violations of our identifying assumptions.

We calculate that only about 1 percentage point of our estimates can be accounted for by returns to post-secondary education for GED-holders. Furthermore, our research methodology controls for pre-GED human capital accumulation. As a result, we interpret our findings as primarily representing the signaling value of the GED in the labor market—the only direct evidence of returns to a labor market signal of which we are aware. While there are several plausible explanations for the white-nonwhite differences in our estimates, only the "programmatic" explanation is equilibrium-preserving.

The fact that we find substantial private returns to the labor market signaling component of the GED raises important social welfare questions. If the GED signal only serves to redistribute earnings, raising the earnings of dropouts who acquire a GED at the expense of dropouts who do not acquire the signal, then the gross social returns to the credential are zero and the net social returns are negative (since there are costs associated with obtaining the signal). If the GED signal leads to better matches between dropouts and jobs, however, then Stiglitz (1975) shows that the social returns can outweigh the private returns. We have no empirical method of determining whether or not employers use the GED to better match dropouts to jobs, and so the impact of the

GED on total social welfare remains an open question.[64]

If there is a job-matching component to the credential, however, then the social returns to the GED are dependent on the passing standards. To see this just note that a standard either so low as to allow everyone to pass or so high as to allow no one to pass would provide no job-matching information. Furthermore, it could be that maximization of total social welfare requires state-specific passing standards, since the underlying distributions of the skills of dropouts and the demand for skills by employers may vary from state to state.[65] Thus, the decentralized manner in which GED passing standards are currently established[66] may approximate the regime under which total social welfare will be the greatest. Less clear, however, is whether the *current set* of state passing standards—where states as diverse as New York and Iowa have the same GED passing standard—is the set that maximizes the social returns to a GED.

---

[64] Furthermore, it may be that the presence of the GED induces some students, who would otherwise benefit from more formal schooling, to drop out of school, further reducing total social welfare since the productivity of these individuals would have been higher had they completed more years of schooling.

[65] This point was first brought to our attention by Jeffrey A. Smith.

[66] As we have noted elsewhere in this paper, the amount of "decentralization" associated with state passing standards was reduced dramatically as of January 1, 1997 with new GED Testing Service mandated minimum requirements.

Table 1. Weighted log hourly wage and probability of employment regressions for white and nonwhite dropouts ages 17-27 from the October 1993 Current Population Survey.[1] (standard errors in parentheses)

	Dependent variable = Log hourly wage		Dependent variable = Probability of being employed[1]	
	Whites N=255	Nonwhites N=170	Whites N=1466	Nonwhites N=917
Intercept	2.02 (0.05)	1.91 (0.05)	0.80 (0.03)	0.76 (0.03)
GED dummy	0.14 (0.05)	0.02 (0.07)	0.08 (0.02)	0.03 (0.04)
Female dummy	-0.27 (0.05)	-0.15 (0.05)	-0.11 (0.02)	-0.10 (0.03)
Region dummies	Yes	Yes	Yes	Yes

1. There are more observations in the employment probability regressions because hourly wage questions are only asked of the outgoing rotation group in the October CPS.
2. A linear probability model was used to obtain the estimates in these columns.

Table 2. GED score groups formed from combinations of minimum and mean scores. (Lightly shaded cells = variation in GED-status, dark shaded cells = all possess GED, no shading = none have GED).

Minimum Score	Mean Score	
	<45	>=45
20-34	Scoregrp 1	
35-39	Scoregrp 2	Scoregrp 4
40-44	Scoregrp 3	Scoregrp 5
45-46		Scoregrp 6
47-48		Scoregrp 7
49-50		Scoregrp 8
51-52		Scoregrp 9
53+		Scoregrp 10

Table 3. Univariate descriptive statistics of the underlying individual-level data.

Row		Connecticut data	Florida data	New York data	GEDTS data
	N =	4,902	36,894	58,128	132,886
	Percent...				
1	age 16 at test	0	1.2	0.5	1.0
2	age 17 at test	0.3	5.9	4.5	6.6
3	age 18 at test	5.3	13.1	10.0	14.3
4	age 19 at test	12.8	14.7	11.7	13.1
5	age 20 at test	12.5	10.4	10.2	8.6
6	age 21 at test	7.9	6.8	6.9	5.5
7	age 22+ at test	61.3	47.8	56.3	50.9
8	male	48.1	52.1	——[1]	48.2
9	female	51.9	47.9	——	51.8
10	white	65.1	68.6	——	58.1
11	black	16.2	13.2	——	9.1
12	Hispanic	16.6	16.1	——	30.2
13	other	2.1	2.2	——	2.6
14	min. test = math	44.2	38.9	41.7	50.6
15	min. test = writing	37.0	37.6	26.4	22.6
16	min. test = science	8.7	10.6	14.0	10.9
17	min. test = social studies	5.7	8.2	13.5	9.8
18	min. test = reading	16.8	18.2	21.9	17.2
19	with GED	76.5	82.4	66.0	68.3
20	with "good SSN"	98	98	90.0	76
21	that validated at SSA	90	95	84	90-95

1. The New York individual-level data had many missing values. We discuss this problem and the solution in the Data Appendix.

Table 4. Stratifiers used to construct the cells in the aggregated data set.

Stratifier	Description
Year tested	7 categories in FL, NY, and CT data (1984-1990) and 1 category in the GEDTS data (1990)
State group	16 categories
GED score group	10 categories
Math tertile category[1]	3 categories in NY data and NA in all other data
Imputed missing writing score[2]	2 categories (yes/no) in State groups 6,7,9-14 and not applicable in other states
Demographic group	4 categories (white male & female, nonwhite male & female)
Age group	2 categories (16-21 on test date, over 21 on test date)

1. For use in another project.
2. Due to administrative record-keeping decisions at the GED Testing Service, some writing scores were lost in the GEDTS data. For these records we imputed the missing writing scores using a quartic function of the other 4 test scores, and fitting separate models by race and gender. In the Data Appendix we discuss this process and offer evidence that the imputation process is acceptably accurate.

Table 5. Summary statistics and treatment/comparison-group designation by state of young dropouts who attempted the GED in 1990.

State group[1]	*Col. 1* N	*Cols. 2, 3, & 4* Treatment or Comparison[2] in...			*Cols. 5 & 6* Percent of sample who are...		*Cols. 7 & 8* Mean '95 earnings	
		Exp3	Exp3*	Exp4	White	Male	No GED	GED
AZ	3,457	—	C	T	65%	57%	7176	9336
CT[3]	1,692	—	—	—	71%	51%	8779	10413
FL	17,905	C	—	C	74%	56%	7767	9694
KY	1,949	—	C	T	83%	58%	6797	8149
NY	19,134	C	—	C	53%	56%	6777	8764
TN	1,668	—	C	T	94%	50%	8479	9079
TX	5,027	T	T	—	60%	51%	7755	9286
VA	1,657	—	C	T	78%	59%	7905	9662
Misc [4]	2,852	—	C	T	79%	54%	7559	8813
Atlantic Sts.	3,237	—	C	T	75%	55%	8184	10327
New England	1,801	—	C	T	91%	52%	8273	9727
North Central	2,313	—	C	T	85%	55%	9031	10364
Low pass Sts.[5]	4,138	T	T	—	76%	55%	7899	8732
Misc[6]	4,202	—	C	T	80%	55%	7750	9032
WA and CA	4,713	—	C	—	64%	57%	8017	9351
Hi pass Sts.[7]	4,255	—	C	—	67%	55%	6258	8669
Total[8]	80,000		—	—	69%	55%	——	——

1. States AZ through VA had enough observations to serve as their own "state group." Other states had to be grouped into either geographic or other logical groupings due to small numbers of observations. See the Data Appendix for a listing of the states in each group.
2. Exp 3 refers to the experiment that uses score group 3 as the affected-group, and Exp 4 refers to the experiment that uses score group 4 as the experiment.
3. Connecticut, a medium standard state, is only used in a direct comparison with New York. See Table 8.
4. Alaska, Hawaii, Idaho, Montana, Alabama, Iowa, Kansas, and Nevada.
5. This group is composed of Mississippi, Louisiana, and Nebraska, and together with Texas they comprise the group of states with the lowest passing standard, a mean of 45 or a minimum of 40.
6. This group is composed of Colorado, Illinois, Pennsylvania, and Minnesota, the only states where there were statistically significant differences between individuals with and without imputed writing scores. (See above and the Data Appendix for a discussion of imputed writing scores in the GEDTS data.)
7. This group is composed of Arkansas, Delaware, Maryland, Oklahoma, South Dakota, and Utah, and along with Washington and California, these states have the same higher standards as do New York and Florida.
8. Note that with about 36,000 observations combined, New York and Florida heavily influence the statistics in this row.

Table 6. DD estimates of the impact of the GED on 1995 earnings of dropouts who tested in 1990. (standard errors in parentheses)

		Experiment 4			Experiment 3			Experiment 3*		
		State Passing Standard is:		Low-Hi Standard Contrast	State Passing Standard is:		Low-Hi Standard Contrast	State Passing Standard is:		Low-Hi Standard Contrast
		Low	Hi		Low	Hi		Low	Hi	
Panel A: Whites										
Test Score is:	Low	9628 (361)	7849 (565)	1779 (670)	9362 (400)	7843 (312)	1509 (507)	9362 (400)	8616 (219)	746 (456)
	Hi	9981 (80)	9676 (65)	305 (103)	9143 (135)	9165 (63)	-23 (149)	9143 (135)	9304 (135)	-162 (150)
Difference-in-Differences for whites				1473** (678)			1531** (529)			907* (481)
Panel B: Nonwhites										
Test Score is:	Low	6436 (549)	8687 (690)	-2252 (882)	7005 (347)	7367 (347)	-363 (495)	7005 (347)	6858 (290)	147 (452)
	Hi	7560 (184)	8454 (96)	-894 (207)	7782 (214)	8375 (93)	-593 (233)	7782 (214)	7568 (133)	214 (252)
Difference-in-Differences for nonwhites				-1357 (906)			231 (548)			-67 (518)

** = significant at the 0.05 level, * = significant at the 0.10 level

Experiment 4: Test Score Low: score group=4; Test Score Hi: score groups=5–10
Passing Standard Low: 35 minimum score and 45 mean score; Passing Standard Hi: 40 minimum score and 45 mean score
Low Passing Standard states: All states except for TX, LA, MS, NE, FL, NY, CA, WA, and CT; Hi Passing Standard states: NY and FL

Experiment 3: Test Score Low: score group=4; Test Score Hi: score groups=5–10
Passing Standard Low: 35 minimum score and 45 mean score; Passing Standard Hi: 40 minimum score and 45 mean score
Low Passing Standard states: All states except for TX, LA, MS, NE, FL, NY, and CT; Hi Passing Standard states: NY and FL

Experiment 3*: Test Score Low: score group=4; Test Score Hi: score groups=5–10
Passing Standard Low: 35 minimum score and 45 mean score; Passing Standard Hi: 40 minimum score and 45 mean score
Low Passing Standard states: All states except for TX, LA, MS, NE, FL, NY, and CT; Hi Passing Standard states: NY and FL

Table 7. DD estimates of the effect of the GED using the log of mean earnings. (standard errors[1] in parentheses)

		Experiment 4			Experiment 3			Experiment 3*		
		State Passing Standard is:			State Passing Standard is:			State Passing Standard is:		
		Low	Hi	Low–Hi Standard Contrast	Low	Hi	Low–Hi Standard Contrast	Low	Hi	Low–Hi Standard Contrast
Panel A: Whites										
Test Score is:	Low	9.094 (0.04)	8.909 (0.08)	0.185 (0.086)	9.072 (0.04)	8.926 (0.04)	0.146 (0.060)	9.072 (0.04)	9.002 (0.03)	0.71 (0.052)
	Hi	9.172 (0.01)	9.158 (0.01)	0.013 (0.011)	9.068 (0.02)	9.103 (0.01)	-0.035 (149)	9.068 (0.02)	9.098 (0.01)	-0.030 (0.017)
Difference-in-Differences for whites				0.172** (0.087)			0.181** (0.062)			0.100* (0.054)
Panel B: Nonwhites										
Test Score is:	Low	8.660 (0.09)	9.021 (0.08)	-0.361 (0.120)	8.843 (0.05)	8.887 (0.05)	-0.043 (0.069)	8.843 (0.05)	8.785 (0.04)	0.058 (0.065)
	Hi	8.896 (0.03)	9.033 (0.01)	-0.136 (0.027)	8.914 (0.28)	9.022 (0.01)	-0.108 (0.030)	8.914 (0.03)	8.890 (0.02)	0.024 (0.033)
Difference-in-Differences for nonwhites				-0.225* (0.124)			0.065 (0.075)			-0.035 (0.073)

** = significant at the 0.05 level, * = significant at the 0.10 level

1. Standard errors obtained by a first order Taylor series approximation.

For all experiments the Test Score Low/Hi, Passing Standard Low/Hi, and states by passing standards are the same as in Table 6.

Table 8. Dual-state DD estimates for whites. (standard errors in parentheses)

		CT vs NY[1]			TX vs FL			TX vs AZ		
		State Passing Standard is:			State Passing Standard is:			State Passing Standard is:		
		Low	Hi	Low-Hi Standard Contrast	Low	Hi	Low-Hi Standard Contrast	Low	Hi	Low-Hi Standard Contrast
Test Score is:	Low	11332 (966)	7610 (779)	3722 (1241)	10169 (656)	7743 (417)	2425 (777)	10169 (656)	8427 (854)	1742 (1077)
	Hi	11559 (320)	9205 (104)	2353 (337)	9587 (189)	9389 (77)	198 (204)	9587 (189)	9209 (220)	387 (291)
Difference-in-Differences for whites				1368 (1286)			2227** (804)			1364 (1117)

** = significant at the 0.05 level

1. Note that if we assumed that Connecticut and New York had similar labor markets for young dropouts, then the simple T-C difference would estimate the effect of the GED on earnings. This estimate would be $3722, with a standard error of 1241. However, the large differences between THi and CHi in this experiment indicate that these states do not have similar labor markets, and that the DD estimator is needed to account for this fact.

CT vs NY: Test Score Low: score group=4; Test Score Hi: score groups=5–10
Passing Standard Low: 35 minimum score and 45 mean score; Passing Standard Hi: 40 minimum score and 45 mean score
Low Passing Standard state: Connecticut; Hi Passing Standard state: New York

TX vs FL: Test Score Low: score group=3; Test Score Hi: score groups=5–10
Passing Standard Low: 40 minimum score or 45 mean score; Passing Standard Hi: 40 minimum score and 45 mean score
Low Passing Standard state: Texas; Hi Passing Standard state: Florida

TX vs AZ: Test Score Low: score group=3; Test Score Hi: score groups=5–10
Passing Standard Low: 40 minimum score or 45 mean score; Passing Standard Hi: 35 minimum score and 45 mean score
Low Passing Standard state: Texas; Hi Passing Standard state: Arizona

Table 9. Specification test summaries.

Threat to Identification	Most Plausible Bias	Method of Assessment	Results
Comparison group is not an appropriate group as determined by pre-treatment earnings comparisons. (In particular, it may be that the treatment group has better pre-treatment outcomes, and so we would expect them to have better post-treatment outcomes, even with a zero treatment effect.	↑	Compare the pre-treatment earnings of the treatment and comparison groups. (Use DD estimates to account for state differences.)	No differences in pre-treatment outcomes for Experiments 4 and 3*. Pre-treatment differences are found in Experiment 3 that favor the comparison group, and hence, would bias *against* finding a treatment effect even if there was one. Bottom line: Pre-treatment contrasts are either zero, or they favor the comparison group over the treatment group.
Individuals endogenize the passing standards, and thus, we do not have appropriate comparison groups.	↑	Use plausible assumptions to adjust the DD estimates downward to account for this type of bias.	Our upper bound estimates drop from a 19 percent treatment effect to a 15 percent treatment effect. Bottom line: Even in the face of this type of endogeneity, plausible assumptions leave our upper bound estimates little changed.
The levels model is mis-specified. The correct model is in log earnings, and we are in a part of the data where answers in levels and logs would be substantially different.	↑ ↓	1) Examine the 1st moment and 2nd moments of the key groups that make up the DD estimates. 2) Obtain DD estimates using the log of mean earnings as a first order approximation of the mean of log earnings.	The first and second moments of key groups are very close for whites (less so for nonwhites), and DD estimates using the log of mean earnings are very close to DD estimates in the levels. Bottom line: It appears that this type of mis-specification is not a problem, particularly with the estimates for whites.
There are missing interactions (e.g., state–skill or state–GED) in the model.	↑	First, establish that the most plausible direction of all biases would be upward, then compare the estimates from experiments that should be affected differently in the presence of such interactions.	The point estimates in Experiments 4 and 3 are less than 4 percent apart. Bottom line: There is little evidence that omitted interactions are driving our estimates.

Table 10. DD estimates for young white dropouts using different high scoring groups as the "high comparison group." (standard errors in parentheses)[1]

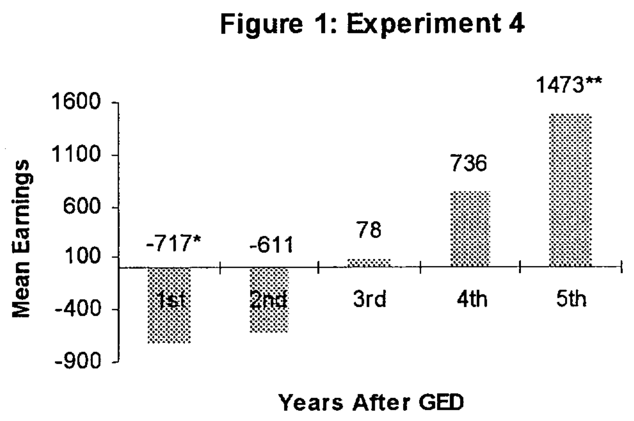
Score group used to construct the high comparison group	Experiment 4	Experiment 3	Experiment 3*[2]
	Treatment N=653 Comparison N=260	Treatment N=471 Comparison N=732	Treatment N=471 Comparison N=171
Score group 5	1732** (691)	1379** (566)	439 (540)
Score group 6	1195* (714)	960 (619)	398 (596)
Score group 7	1592** (724)	1654** (630)	810 (610)
Score group 8	1460** (735)	2062** (657)	1477** (637)
Score group 9	1263* (760)	2691** (698)	2148** (692)
Score group 10	1284* (735)	1551** (690)	988 (684)

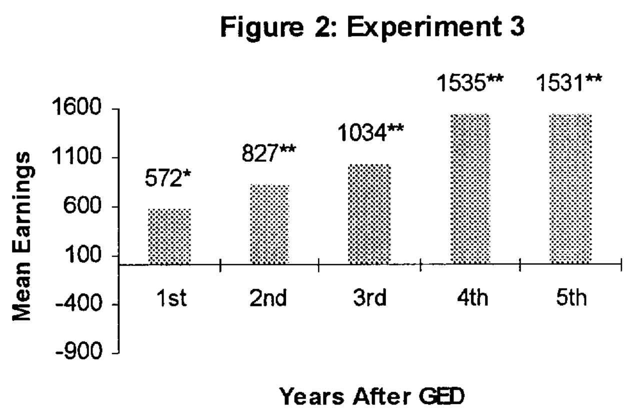
** = significant at the 0.05 level, * = significant at the 0.10 level.

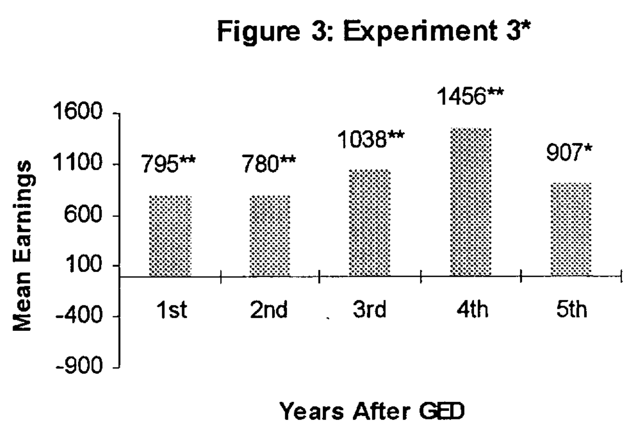
1. "Treatment N" and "Comparison N" in this table refer to the number of treatment and comparison observations in the affected-score group.

2. Experiment 3* uses states in the GEDTS data as comparison states, and thus these estimates would be lower bound estimates of the true GED treatment effect.

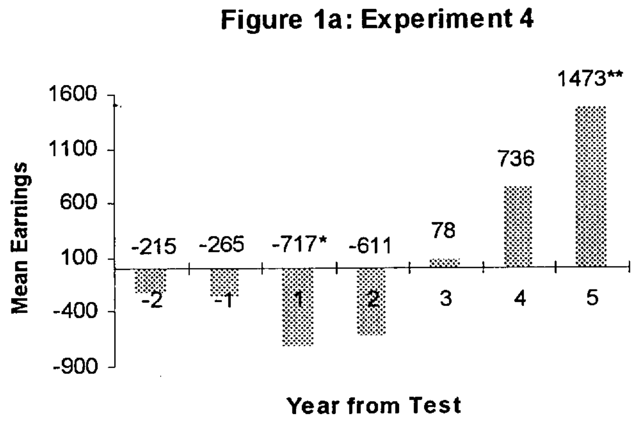
Figures 1-3. DD estimates for young whites in the 1st through 5th years after GED receipt.
(*=significant at the 0.10 level, **=significant at the 0.05 level)

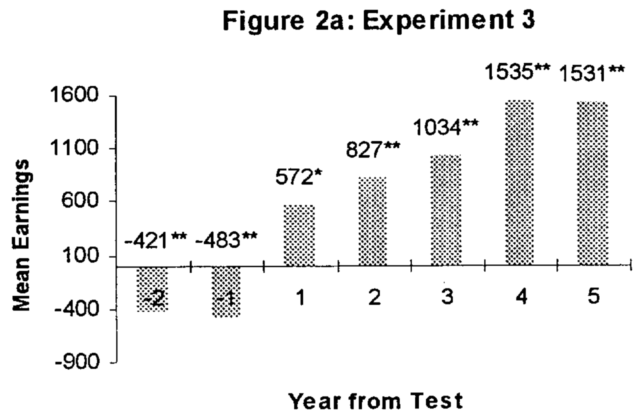


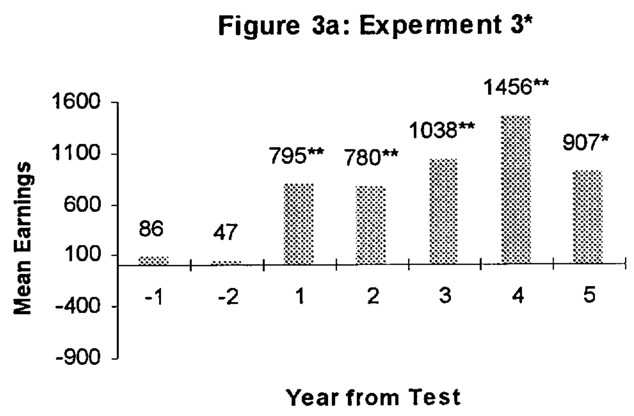




Figures 1a-3a. Pre-treatment and 1st through 5th year DD estimates for young white dropouts.
(*=significant at the 0.10 level, **=significant at the 0.05 level)







Appendices

A. Data Appendix

A.1 Data Sources

As discussed in the text, the underlying individual-level data for this project comes from four different sources: the state departments of education in Connecticut, Florida, and New York and the GED Testing Service. Each of the data sets from these states is constructed differently, and we discuss each in turn.

Data from the GED Testing Service

The original data provided to us from the GED Testing Service (GEDTS) contain basic demographic information and the scores on the 5 tests in the GED battery for a sample of individuals who attempted the GED in 1990. Individuals from all but two states are represented in the data, and we use the data on individuals from 42 different states.[67] These data were originally collected by the GEDTS as part of an ongoing, several-year project to study responses to proposed new GED test items.[68]

The data that GEDTS forwarded to us for our project has considerably fewer observations in every state than the number of test-takers indicated in GEDTS reports on the 1990 tests (GED Testing Service, 1991). Several steps in the collection and data storage process that began at individual GED test sites and ended at GEDTS could have resulted in lost data. First, it is important to understand that the provision of data in the GEDTS test-item-response study was usually accomplished when individual test sites[69] would forward the original hard copy answer sheets to

---

[67] The passing standards for several states made them unsuitable for use in this project.

[68] We are in appreciation to Alice Marable for information regarding the collection, processing, and storage of the GEDTS data (Marable 1997).

[69] In 1990 most states did not have (and still do not have) centralized scoring of the GED exams. In most cases they were either scored on site, or sent to local centralized locations where they were scored. Pass/fail results and basic demographic information for individuals, but not usually the individual test scores, were then sent to the state education agency. The state education agency then reported the aggregate year-end results to GEDTS for use in their annual *Statistical Report*.

GEDTS.[70] Actions that could have resulted in lost data in that process include the following:

1. according to GEDTS officials, a number of sites did not ever send data;
2. sites that were using old-style Scantron half-sheet readers were exempt from sending data, since the GEDTS had no way to handle such data;
3. for data that was on diskettes, Social Security Numbers (SSNs) were usually not a part of the record that was sent to GEDTS;[71]
4. when diskettes were sent, some diskettes were not useable due to errors in creating the data files at the GED center or state education agency; and,
5. idiosyncratic administrative behavior at the state or center level caused the loss of some data.[72]

Finally, in the process of merging separate tests together so that the complete 1990 record for an individual would be attached to that individual, GEDTS programmers discarded observations that lacked scores on all 5 tests, an action that would make the number of observations in our data even more discordant with the number of test-takers in the GEDTS 1990 Statistical Report. As a result of these data collection, storage, transfer, and merging processes, the data transferred to us represent only a sample of 1990 test-takers from 48 states. Based upon our review of the processes that produced the GEDTS data, we have no compelling reason to believe that within-state sample selection bias is a problem with these data.

It is important to understand that as a result of the merge process at GEDTS, that the GEDTS sample only includes individuals who attempted, either in one sitting or in several sittings, all 5 tests in the GED battery in 1990. Thus, as discussed in the text, these data are only a one-year "snapshot" of GED-takers, and some portion of the individuals who are coded as having not passed the battery on the basis of the 1990 scores, actually have a GED as a result of higher scores from

---

[70] The sites that used machine scoring sent diskettes containing the information.

[71] In this case, the data are not lost for the original GED purposes, but such observations were deleted from our study.

[72] For example, when one state switched from local to state scoring in this time period there was considerable confusion in that state regarding the handling of answer sheets for individuals who only took a part of the GED battery. As a result, for a period of time, data for these "partial" testers was not forwarded to the GEDTS.

previous years that are not in the data or higher scores from subsequent re-attempts in post-1990 years.

One other point about the GEDTS data needs discussion. About 50 percent of the observations in the data forwarded to us had missing writing scores. Inquiries to programmers at GEDTS indicated that the writing scores were missing as a result of administrative record-keeping decisions at GEDTS.[73] To retain observations and prevent sample selection bias that might otherwise occur if we eliminated records with missing writing scores, we used a quartic function of the 4 test scores available to predict the writing score, and we fit separate models for males and females, whites and nonwhites. To analyze how accurately we could place individuals into score groups based on the predicted writing scores, we compared the score-group placement of individuals with true writing scores vs. predicted writing scores in the half of the data where the writing scores were present. The main concern is that the prediction process incorrectly placed relatively high scoring individual in the treatment affected-groups (score groups 3 and 4). If this were to happen in a systematic manner, then the DD estimates would be biased upward, since the treatment group would be contaminated with individuals who actually belonged in higher score groups. Our analysis using those with all 5 test scores showed that we tended to overestimate the writing score, and thus inaccurately place low scoring individuals in the treatment affected-group. That analysis showed that 90 percent of those predicted to be in score group 3 (the treatment affected-group in Experiment 3) were accurately placed. Of the 10 percent who were inaccurately placed, their actual

---

[73] Since the essay part of the writing test generally took longer to score in the field, records on the other 4 tests were often forwarded to GEDTS while the essays were being scored. Upon completion the essay scores were then forwarded to GEDTS. As the main data storage facilities at GEDTS became periodically overburdened, data that had been processed for the original test-item-response study was periodically taken off of the main storage devices and stored on tapes, even if the writing scores had not yet arrived and been merged in. GEDTS programmers did not try to re-match late arriving writing scores with data that had been removed from the main system. The result was that many observations had all test scores except for the writing score.

An additional reason for missing writing scores was that many essay results were sent to the GEDTS from a different location (the essay scoring site as opposed to the testing center or state agency), and in many cases whole batches were not sent. Furthermore the essay portion of the test battery did not have a scannable Social Security number grid. Thus, the only matching rubric possible to recombine essay scores with other tests in the battery was a combination of the code that identifies a unique test "booklet" and the test center identification number. In many cases, however, the essay scoring centers sent essays in batches without indicating centers, making it impossible to match/merge.

scores placed them in score groups below the affected-group. Based on these numbers, we concluded that the prediction process was doing a reasonably accurate job of correctly placing individuals, and that any bias induced by the process would result in conservative estimates of any observed treatment effects.

Data from the Connecticut State Department of Education

Data from this source include basic demographics and GED test-scores on all individuals who attempted the GED battery in the years 1984-1990. This is the most complete data set we have on GED-attempters because Connecticut retains the complete GED test history on individuals. Thus, we have every score from every attempt for every individual, plus the dates on which the test was attempted.

Data from the Florida State Department of Education

Data from this source also include basic demographics and GED test-scores on all individuals who attempted the GED battery in the years 1984-1990. As a result of the manner in which GED records are maintained in Florida, these data only contain the test scores from the first test ever taken and the accumulated best test scores. Thus, if a person only attempted the GED once or twice, then we have their complete record. However, if they attempted the battery more than twice, then their final test results only reflect the accumulated best scores from all of their attempts. Furthermore, we cannot identify in the data the number of repeat attempts; we only know if individuals attempted the battery once or more than once.[74] For the purposes of this study, the Florida data are sufficient, since we are only concerned with individuals' scores as they relate to receipt of the GED, and the relevant scores for that purpose are the accumulated best scores.

Data from the New York State Department of Education

Data from this source also include basic demographics and GED test-scores on all individuals who attempted the GED battery in the years 1984-1990. The GED record-keeping system in New

[74] In the Connecticut data where we have more complete information, 79 percent of the 1990 test-takers had attempted the battery only once, 15 percent had attempted it twice, 4 percent had attempted it three times, and only 2 percent had attempted the battery four or more times.

York is similar to that of Florida. However, the data forwarded to us from New York only contain the accumulated best scores of GED test-takers in that state.

As a result of New York collection and record maintenance procedures, 53 percent of the data on 1990 test-takers in this state were either missing data on gender or race or both. Fortunately, these data did contain information on month and year of birth. Using month and year of birth to validate SSNs (see the discussion on SSN validation later in this appendix), SSA programmers were able to attach gender and racial information to observations missing values on these variables in the New York data. However, because much of the New York data had to be matched on month and year of birth, instead of demographic variables, the validation rate in the New York data is slightly lower than in Connecticut and Florida, as indicated in Table 3 in the body of the paper.

A.2 SSN Validation

Table 3, in the body of the paper, shows that SSA programmers successfully matched and validated a high percentage of records. Except in the case of New York, which is explained above, matches were validated using SSA data on sex, race, and year of birth. A proposed match was assumed to be valid if the GED and SSA sex codes matched and the GED and SSA years of birth differed by no more than 2 years in absolute value and if the GED and SSA race codes did not disagree on white vs. black designation.

A.3 Social Security Earnings Coverage and Taxable Maximum

As detailed by Angrist:

> OASDI and HI (Medicare) are contributory programs in which covered workers pay a tax and report earnings to the SSA via employers or the Internal Revenue Service. The SSA keeps track of taxable earnings on the Summary Earnings Record (SER)...Under the Federal Insurance Contributions Act (FICA) of the Self-employment Contributions Act (SECA), earnings are taxed for Social Security purposes up to the Social Security taxable maximum, and are therefore recorded on the SER only up to this maximum. The taxable maximum has been raised every year. By 1980, over 85 percent of covered male workers had earnings below the maximum. A separate and higher taxable maximum was instituted for HI in 1991. All taxable earnings, whether reported for the purposes of OASDI or HI, and whether reported as part of a mandatory or voluntary coverage provision, should appear on the SER.
>
> About 95 percent of all jobs in the U.S. were covered by the OASDI program as of 1991.

> Exceptions fall into five major categories: (1) Federal civilian employees hired before 1984, (2) railroad workers, (3) some employees of State and local governments already covered under a retirement system, 94) household and farm workers with low earnings, and (5) persons with no wage and salary earnings and very low earnings from self-employment. Members of the uniformed services have been covered since 1956, and receive noncontributory wage credits to improve their insured status. Recent important changes include the 1983 coverage of most federal employees and employees of nonprofit organizations, HI coverage of many state and local employees in 1986, OASDI coverage of State and local employees without an employer retirement plan in 1990, and coverage of reserve soldiers in 1987. (Angrist 1997)

Taxable maximum values for the years 1991-1995 are:

1991—$53,400

1992—$55,500

1993—$57,600

1994—$60,600

1995—$61,200

Given the generally low annual earnings of dropouts, we did not anticipate that earnings values top-coded at the taxable maximum would be a problem in our data, and this turned out to be the case. In the GEDTS, Connecticut, and Florida data,[75] the percent of young dropouts with any top-coded values are 0.1 percent, 0.2 percent, and 0.2 percent respectively.

A.4 SSA Confidentiality Edit

The SSA confidentiality regulations that affected the data released to us censored cells that had fewer than 3 individuals or cells where the standard deviation of earnings was zero. This censoring masked the 1995 mean earnings in 164 of the 2,138 total cells released to us on individuals who took the GED battery in 1990.

[75] We do not have top-coding information on New York.

B. Statistical Appendix

Weighting in the DD Estimator

As discussed in the body of the paper, whenever we obtain an estimate with the DD estimator over a sample that is not homogeneous in gender and race, or when more than one state makes up either the treatment or comparison group, then we have to balance demographic attributes and/or state distribution across the comparison groups. We do this by weighting the data. To illustrate the problem, consider Figure 1A and the simple example where (1) we are conducting Experiment 4 using score group 8 as the "high comparison" group, (2) we are using only Connecticut as a treatment state and only New York as a comparison state, and (3) we are forming the estimate using a sample of, say, white females.

Figure 1. Groups making up the DD estimator (shading: $D_i = 1$).

Affected-Group Score group 7	**Hi Comparisons** Score group 8
T	**T_Hi**
C	**C_Hi**

In this case there would be no heterogeneity of gender or race to balance across the different cells, and there would be no worry that different distributions across T and T_Hi states or across C and C_Hi states could affect the results; white females from Connecticut make up both the T and T_Hi cells, and white females from New York make up the C and C_Hi cells. The simple DD estimator is sufficient in this situation.

Consider, however, the problem of using both New York and Florida as comparison states. In this case, if white females were distributed differently across New York and Florida in the C group than in the C_Hi group, and if there were different labor markets for white females in New York and Florida, this could lead to spurious findings. To account for this possibility, we could weight the earnings to account for the different distributions across states in the C and C_Hi groups. The same logic applies when more than one state is used in the T and the T_Hi groups.

Following this line of reasoning, when we use a sample that is not homogeneous in gender and race to form the estimate, we have to weight all four of the groups so that they are balanced across these attributes. For these reasons, the DD estimator we use compares weighted mean earnings using weights that accomplish the following:

1) the T_Hi, C, and C_Hi groups are weighted to reflect the same demographic distribution as group T,

2) the T_Hi group is weighted to reflect the same distribution across states as group T, and

3) the C_Hi group is weighted to reflect the same distribution across states as group C.

Finally, the weights are normalized so that,

$$\sum_{j=1}^{J} w_j * n_j = \sum_{j=1}^{J} n_j$$

where w_j is the weight applied to the mean earnings in the j_{th} cell in the data.

C. Additional Results

Producing "adjusted" estimates to account for possible bias arising from endogenous decision-making in the face of higher standards

Individual-level data from New York offer us a good opportunity to observe the behavior of GED candidates who are faced with a different passing standard. In September 1985, the GED passing standard was raised in New York from one that required a minimum score of 35 and a mean of 45 to the current standard, which requires a minimum score of 40 and a mean of 45. Table A1 gives the distribution of GED test-scores in New York in the year preceding the change in standards and the year after the change.

Table A1 here.

The table shows a sharp drop between the pre- and the post-change periods in the percentage of test-takers whose scores would place them in score group 4—the scoring range just high enough to obtain the GED in the pre-change era and the scoring range just below the cutoff in the post-change era. We postulate that this drop is likely some combination of an "intimidation" effect and score-group sorting.

The intimidation effect would occur if some individuals who are "true 4s"—individuals whose true underlying skills would allow them to obtain scores that would, *at best*, place them in score group 4— are intimidated by the higher standard, and so they never attempt the test when faced with the higher standard.[76] Such behavior would obviously result in a drop in the percentage of individuals observed in score group 4.

Score-group sorting around the standard change could occur in the following manner: when faced with the old, lower standard, some portion of individuals who were true 5s or 6s scored in the just-good-enough score group 4 range, and received their GED certificates; presented with

[76] This effect is likely exacerbated by the prevalence of GED practice tests that are used in GED-preparation programs and that are available through commercial publishers. Thus, many "true 4s" may have discovered on the practice test that they could not score high enough to acquire the credential under the new standards, and thus, elected to not attempt the battery.

the higher standard, however, these same individuals would be observed in score group 5 or 6, since it would require attaining scores at least high enough to place them in score group 5 to pass the battery.

It is important to note in Table A1 that score group 3, which is also a non-passing score group close to the GED threshold, exhibits much more stability around the change in standards. The more stable behavior associated with score group 3, especially relative to score group 4, is likely a result of the definitions of the two score groups. For example, it would require either raising several scores on the GED battery by a moderate amount each or increasing one score on the battery by a relatively large amount to move an individual from score group 3 (non-passing) to score group 5 (passing). On the other hand, it potentially only requires raising one test by a few points to move from score group 4 (non-passing) to score group 5 (passing). For this reason, to the extent that individuals do tend to sort themselves when faced with different GED passing standards, this may be less of a concern in Experiment 3 than in Experiment 4.

This supposition is supported by the figures in Table A2, which gives the distribution of observations across score groups for New York, Florida, Connecticut, those states in the GEDTS data where a GED is awarded in score groups 4 and above, and in those states in the GEDTS data where a GED is awarded in score groups 3 and above. Of particular interest are comparisons in the distributions across score-groups 3, 4, and 5 for New York, Florida, and Connecticut. The fact that the percentage of observations in Connecticut in score-group 3 is roughly the same as in New York and Florida (and not substantially smaller), even though this score-group is immediately below the cut-off in Connecticut, suggests that there is little mobility associated with score-group 3, regardless of the state passing standard.[77]

Table A2 here.

We are still concerned, however, with individual-level endogeneity, especially in Experiment

[77] The distributions around the GED cut-off in the GEDTS data, as compared to New York, Florida, and Connecticut, reflect, in part, the contamination issue in these data already discussed. That is, we expect that the percentage in the score group just below the cut off in the GEDTS states to be slightly overstated, since some of these individuals received their GED in subsequent years.

4. The potential biases associated with this type of endogeneity can be explored by making some assumptions regarding the manner in which young dropouts may sort themselves when faced with different passing standards. First, we assume that dropouts with relatively high skills use those skills to earn their true score on the GED, and thus individuals in the higher scoring groups, 6 through 10, sort themselves into their true scoring group. Second, we assume that lower skilled individuals, those whose scores place them below the GED threshold in any particular state, tend to score as high as they can, and thus they sort themselves into their true scoring group. Third, we assume that some individuals with medium levels of skills—true 4s, 5s, and 6s—may bring just enough effort to the GED battery to pass, and so we observe some of these individuals in lower score groups than their true skills warrant. Figure A1 displays the sorting that would result from these assumptions as dropouts faced the three different passing standards represented in our data.

Figure A1 here

If this type of sorting occurs, then our basic assumption regarding the use of the natural experiments is violated. That is, if Figure A1 is an accurate depiction of what happens when dropouts are faced with different GED passing standards, then no longer is GED-status in the affected score-group exogenously determined. Practically, this means that if we contrast the mean earnings of those in the Medium-standard (treatment) states who are in score group 4 with the mean earnings of those in the Hi-standard (comparison) states who are in score group 4, our estimate of the treatment effect of the GED would be biased upward as a result of the contamination of the treatment group with true 5s and 6s.

We can adjust our estimates to account for endogeneity bias by bringing two assumptions to the data. To obtain adjusted estimates we assume that:

(1) the observed proportion of individuals in the affected-group in the comparison-group states represents what the true skill distribution in the affected-group in the treatment states would be in the absence of individual sorting, and

(2) the mean earnings of higher skilled individuals who score less than their true abilities are the same as the observed mean earnings of similarly skilled individuals who sort themselves into their true score groups.

Adjusted estimates based on these assumptions are in Table A3. Since we assume that the treatment group in each of the experiments could be contaminated by true 5s or true 6s, we employ three different assumptions in Table A3 regarding the mix of 5s and 6s. First, we assume no contamination, which simply redisplays our original results from Table 6 for ease of comparison. Next, we assume a 50-50 mix of 5s and 6s, and then to get very conservative, lower-bound estimate, we assume only 6s contaminate the treatment group.

Table A3 here.

We begin with Experiment 4 in the table. If we assume that the contamination in the treatment group comes from a 50-50 mix of true 5s and 6s, then our original DD estimate of \$1473 is biased upward by about 20% (\$307). The \$1166 adjusted estimate in this case represents a 15% annual earnings premium attributable to GED-acquisition. Using the more conservative estimate of considerable contamination, we get an adjusted DD estimate of \$612.[78] We believe, however, that the most plausible assumption is that there is some mix of true 5s and 6s in the treatment affected-group, and thus \$1166 is the best lower bound DD estimate for Experiment 4.

The most conservative adjusted DD estimate for whites in Experiment 3, on the other hand, is little different from our original estimated treatment effect. That is, even if we assume that all of the 281 high scorers that potentially contaminate the treatment group in this experiment are true 6s, then the original DD estimate is only 11 percent larger than the conservative lower bound estimate of \$1374. And, the original estimate from Experiment 3* is only 3 percent larger than the most conservative adjusted estimate of \$881.

These exercises in obtaining alternative DD estimates lead us to the following conclusions. First, we believe that individual sorting is less of an issue in experiments 3 and 3* than in Experiment 4. This is because the definitions of the two different affected-groups leave it relatively more difficult for individuals to adjust their scores upward from score group 3 than from score group 4, and there is evidence of this in Table A2. As a result, we believe that our conservative estimate of \$1374 for Experiment 3 in Table A3, represents a plausible lower bound on the estimate for whites, accounting for endogeneity bias. This estimate of the effect of the

[78] We do not make the additional assumptions that would be needed to construct confidence intervals around our adjusted point estimates in Table A3.

GED on earnings represents a 17 percent increase in earnings for young white dropouts with the credential relative to the comparison group earnings.

Overall, the estimates for whites in Table A3 that have been adjusted for potential endogenous decisions by young dropouts support the general findings of this paper. Namely, for young white dropouts, acquisition of a GED appears to be associated with a substantial increase in earnings that is attributable to the effect of the GED on the earnings of young dropouts.

Assessing the returns to skills for whites and nonwhites.

To explore the possibility that nonwhite are relegated to jobs where skills matter little, we conduct a simple exercise using the Florida data.[79] In these data, we combine individuals into four skill categories. The middle two categories (call them 2 and 3) are individuals in the score groups just below and just above the GED threshold in the state. The lowest skill category (1) is comprised of all individuals in the score groups below category 2, and the highest skill category (4) is comprised of all individuals in the score groups above category 3. Differences in 1995 mean earnings between categories in this setup reflect measures of returns to skills or skills plus any GED-effect as demonstrated in Figure A2 below.

Figure A2 here

Since no dropout in categories 1 or 2 had a GED, differences in earnings between these categories reflect returns to skills for dropouts with low-levels of skills; since all dropouts in categories 3 and 4 possess a GED, differences in earnings between these categories reflect returns to skills for dropouts with relatively high-levels of skills (as measured by the GED exams); and, since dropouts in category 3 have a GED, while those in category 2 do not, differences in mean earnings between these categories reflect some combination of returns to skills plus returns to the

[79] We report only the Florida evidence, but we find similar results in analyses using all of the other state groups.

GED.[80] We are particularly interested in the 2–1 and the 4–3 measures for nonwhites as a way of examining whether or not there are payoffs to skills for nonwhites.

Table A4 displays the mean earnings comparisons for males and females across the three relevant differences: 2–1, 3–2, and 4–3. The results from this table are consistent, striking, and, since we have replicated these findings using the other state groups in the data, they refute the hypothesis that there are no returns to skills for nonwhite dropouts.

Table A4 here

The relevant comparisons in this table occur along the two pairs of white-nonwhite rows. The shaded cells represent some combination of returns to skills plus returns to GED-acquisition, and the results in the shaded cells for males in Florida fit the now-familiar story that nonwhites experience no returns to a GED five years after receipt (–$198), while whites do ($1797). We are primarily interested, however, in the non-shaded comparisons, and the results here show clear white-nonwhite differences, but they favor nonwhite males. Measures of both returns to low-level skills (2–1) and higher-level skills[81] (4–3) show that skill differences, both at the bottom and the top end of the skill distribution, translate into larger earnings differences for nonwhite males than for white males in Florida. The results in Table A4 strongly suggest that the lack of GED treatment effects for nonwhite dropouts does not stem from a lack of payoff to skills.[82]

GED treatment effects: Effects on wages and hours-worked of the already-employed vs. Effects on decreasing persistent non-employment

---

[80] It should be remembered, however, that skill groups 2 and 3 are very close together, so that the 3–2 difference is primarily driven by returns to a GED.

[81] While we designate this differences as a measure of returns to "high-level" skills, it is important to keep in mind that this simply means returns to the basic skills of individuals who score in the upper ranges of the GED exams.

[82] These higher returns for nonwhites are consistent with the findings of Neal and Johnson (1996). Using the *NLSY*, they find that black males and black and Hispanic females have higher returns to the AFQT than do whites.

Using other data available to us, we are able to get some idea of how much of the effect of the GED on the earnings of whites that we measure in Table DD is due to a decrease in number of individuals with zero earnings in the year (individuals who are "persistently unemployed") versus how much is due to an increase in the wages and/or hours-worked of those already working.[83] Unfortunately, these additional data do not allow us to disaggregate dropouts by age-at-test, since they contain dropouts of all ages, and thus, the results from this exercise can only suggest what we might actually find for young dropouts. Using these data, however, and the same experiments as discussed throughout this paper to identify treatment effects, we find that the GED decreases persistent non-employment among white females by 8 to 10 percentage points depending on the experiment, and that the credential does not change persistent non-employment rates among white males. These employment effects for white females are substantial since the comparison group of white females in these experiments has a persistent non-employment rate of 37 percent, and they explain about one-half of the earnings increases attributable to GED-acquisition for young white females.[84] Thus, to the extent that what we would find among young white dropouts is the same as the effects we estimate over dropouts of all ages, the GED affects the mean earnings of young white males through increases in wages and/or hours worked of those already working, and it affects the mean earnings of young white females by some combination of increases in wages and/or hours worked of those already working and decreases in the number of white females with zero earnings in the year.

[83] The available data for this analysis only allow for these coarse employment vs. wage effects.

[84] We determine the maximum possible effect of decreased persistent non-employment for females in the following manner. The treatment effect of the GED on the earnings of young white females represent increased earnings of 14 percent (in Experiment 3) and 22 percent (in Experiment 4). If we assumed that there were no increases in wages or hours worked for those with positive earnings and that the newly employed had earnings comparable to those with positive earnings, then the effects from decreased persistent non-employment (-8 percentage points and -10 percentage points) could explain around one-half of the earnings gains.

Table A1. Distributions across GED score groups in New York before and after a change in standards in 1985.

Score group	Proportion pre-change	Proportion post-change
1	0.075	0.086
2	0.103	0.124
3	0.048	0.058
4	0.053	0.012
5	0.343	0.350
6	0.123	0.116
7	0.081	0.079
8	0.060	0.060
9	0.041	0.044
10	0.072	0.071
Total	1.00	1.00

Table A2. Distributions across score groups by state or passing standard for 1990 test-takers. (shaded cells = GED-granting score groups)

	Proportion in each score group for ...				
Score group	New York	Florida	Connecticut	GEDTS -med. standard states	GEDTS - low standard states
1	0.073	0.038	0.059	0.096	0.128
2	0.093	0.048	0.071	0.122	0.141
3	0.047	0.028	0.030	0.070	0.095
4	0.016	0.009	0.069	0.037	0.022
5	0.340	0.344	0.306	0.251	0.231
6	0.126	0.135	0.125	0.121	0.121
7	0.100	0.112	0.085	0.099	0.094
8	0.069	0.094	0.082	0.077	0.065
9	0.053	0.077	0.064	0.054	0.051
10	0.085	0.115	0.111	0.073	0.052
Total	1	1	1	1	1

Table A3. Different DD estimates under assumptions of individual score-group sorting.

	Adjusted numbers of observations					Adjusted DD Ests. Under Different Assumptions	
	p_c[1]	N_{treat}[2]	N_T[3]	$p_c(N_{treat})$[4]	$[N_T - p_c(N_{treat})]$[5]	Treatment group contamination assumption of ...	White DD Est.
Experiment 4							
whites	0.011	18332	653	202	451	no contamination.[6]	1473
						the contamination is 50% true 5s and 50% true 6s.	1166
						all contamination is from true 6s.	612
Experiment 3							
whites	0.031	6137	471	190	281	no contamination.[6]	1531
						contamination is 50% true 5s and 50% true 6s.	1715
						all contamination is from true 6s.	1374
Experiment 3*							
whites	0.062	6137	471	380	91	no contamination.[6]	907
						contamination is 50% true 5s and 50% true 6s.	936
						all contamination is from true 6s.	881

1. p_c = the observed proportion of individuals from comparison states in the comparison bubble group.
2. N_{treat} = the total number of observations in the treatment states, across all score groups.
3. N_T = the number of observations in the treatment bubble group.
4. The assumed number of treatment group individuals with the correct true score.
5. The assumed number of treatment group individuals with true scores that would place them in score group 5 or 6; these are the "contaminating" individuals.
6. We include our original DD estimates in this table (the estimates we obtain under an assumption of no contamination) for ease of comparison.
7. The mean earnings of score group 5 for the treatment states in this experiment are actually slightly lower than the mean earnings of the treatment bubble group here, and so a calculation using the earnings of true 5s would result in a *higher* adjusted estimate.

Table A4. Differences between 1995 mean earnings for adjacent skill groups in Florida. (standard errors in parentheses)

Group	Race	Returns to... Low-level skills (2 - 1)	Med-level skills + GED (3 - 2)	High-level skills (4 - 3)
Florida males	whites	148 (671)	1797** (593)	-84 (236)
	nonwhites	1673 (1064)	-198 (1012)	1989** (424)
Florida females	whites	427 (549)	1117** (468)	1097** (201)
	nonwhites	1193 (793)	210 (736)	1503** (406)

** = significant at the 0.05 level, * = significant at the 0.10 level.

Figure A1. Score-group sorting around three GED passing standards (shaded cells = score groups where the GED is awarded).

Unobserved sorting of true skill distribution among...	Observed GED score groups 1	2	3	4	5	6	7	8	9	10
low standard states.[1]	1	2	3,4,5,6	4,5,6	5,6	6	7	8	9	10
medium standard states.[2]	1	2	3	4,5,6	5,6	6	7	8	9	10
hi standard states.[3]	1	2	3	4	5,6	6	7	8	9	10

1. States where the GED is awarded in score groups 3 and above.
2. States where the GED is awarded in score groups 4 and above.
3. States where the GED is awarded in score groups 5 and above.

Figure A2.

This difference reflects returns to low-level skills:	This difference reflects returns to medium-level skills plus any GED-effect:	This difference reflects returns to high-level skills:
2 - 1	3 - 2	4 - 3

References

Angrist, Joshua D. "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants." *Econometrica* forthcoming (1997).

Auchter, Joan Chikos, and Gary Skaggs. "The Performance of 1993 Graduating High School Seniors: How Do U. S. High Schools Measure Up?" A paper delivered at the National Council on Measurement in Education, New Orleans, Louisiana, 1994.

Baldwin, Janet. *GED Profiles: Adults in Transition.* Washington, D.C.: GED Testing Service of the American Council on Education, 1990, 1.

Baldwin, Janet. *GED Profiles: Adults in Transition.* Washington, D.C.: American Council on Education, 1992, 5.

Batts, Jacqueline. "Personal communication, June 1997." , 1997.

Besley, Timothy, and Anne Case. "Unnatural Experiments? Estimating the Incidence of Endogenous Policies." *NBER Working Paper #4956* , no. December 1994 (1994).

Boesel, David, Nabeel Alsalam, and Thomas M. Smith. *Educational and Labor Market Outcomes of GED Certification.* Washington, D.C.: Office of Educational Research and Improvement, Dept. of Education, 1997. Working Paper.

Cameron, Stephen V. "Assessing High School Certification for Women Who Dropout." *University of Chicago Working Paper* (1994).

Cameron, Stephen V., and James J. Heckman. "The Nonequivalence of High School Equivalents." *Journal of Labor Economics* 11, no. 1 (1993): 1-47.

Campbell, Donald T. "Reforms as Experiments." *American Psychologist* 24 (1969): 409-429.

Card, David, and Alan B. Krueger. "Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84 (1994): 772-793.

Cave, George, Hans Bos, Fred Doolittle, and Cyril Toussaint. *JOBSTART: Final Report on a Program for Dropouts.* New York: Manpower Demonstration Research Corporation, 1993.

Committee on Ways and Means U.S. House of Representatives. *1993 Green Book: Background Material and Data on Programs Within the Jurisdiction of the Committee on Ways and Means.* Washington, DC: U.S. Government Printing Office, 1993.

Erwin, Cathy. "Personal communication, June 1997." , 1997.

Friedlander, Daniel, and Karin Martinson. "Effects of Mandatory Basic Education for Adult AFDC Recipients." *Educational Evaluation and Policy Analysis* 18, no. 4 (1996): 327-337.

Gareth, Michael S., Zhongren Jing, and Mark Kutner. *The Labor Market Effects of Completing the GED: Asking the Right Questions*. Washington, D.C.: American Institutes for Research, 1995.

GED Testing Service. *GED 1990 Statistical Report*. Washington, D.C.: American Council on Education, 1991a.

GED Testing Service. "GED Examiner's Manual." . Washington, D.C.: American Council on Education, 1991b.

GED Testing Service. *Who Took the GED? GED 1996 Statistical Report*. Washington, D.C.: American Council on Education, 1997.

Greene, William H. *Econometric Analysis*. Second ed. New York: Macmillan Publishing Company, 1993.

Gruber, Jonathan. "The Incidence of Mandated Maternity Benefits." *American Economic Review* 84, no. 3 (1994): 622-641.

Heckman, James J., and Jeffrey A. Smith. "Assessing the Case for Social Experiments." *Journal of Economic Perspectives* 9, no. 2 (1995): 85-110.

Kane, Thomas J., and Cecilia Rouse. "Labor Market Returns to Two-Year and Four-Year College." *American Economic Review* 85, no. 3 (1995): 600-614.

Kemple, James J., Fred Doolittle, and John W. Wallace. *The National JTPA Study: Final Implementation Report*. New York: Manpower Demonstration Research Corporation, 1992.

Lang, Kevin. "Does the Human-Capital/Educational-Sorting Debate Matter for Development Policy?" *The American Economic Review* 84, no. 1 (1994): 353-358.

Lang, Kevin, and David Kropp. "Human Capital Versus Sorting: The Effects of Compulsory Attendance Laws." *The Quarterly Journal of Economics* 101, no. August (1986): 609-624.

Marable, Alice. "Personal communication, June 1997." , 1997.

Meyer, Bruce D. "A Quasi-Experimental Approach to the Effects of Unemployment Insurance." *NBER Working Paper No. 3159, November 1989* (1989).

Meyer, Bruce D. "Natural and Quasi-Experiments in Economics." *Journal of Business and Economic Statistics* 13, no. 2 (1995): 151-161.

Murnane, Richard J., John B. Willett, and Katherine P. Boudett. "Do High School Dropouts Benefit from Obtaining a GED?" *Educational Evaluation and Policy Analysis* 17, no. 2 (1995): 133-147.

Murnane, Richard J., John B. Willett, and Katherine P. Boudett. "Does Acquisition of a GED Lead to More Training, Post-secondary Education, and Military Service for School Dropouts?" *Industrial Labor Relations Review* forthcoming (1997).

Murnane, Richard J., John B. Willett, and John H. Tyler. "What Are the High School Diploma and the GED Certificate Worth in the Labor Market? Evidence for Males from High School and Beyond." *Harvard Graduate School of Education Working Paper* (1996).

National Center for Education Statistics. *Digest of Education Statistics*. Washington, D.C.: U.S. Department of Education, 1996.

Neal, Derek A., and William R. Johnson. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy* 104, no. 5 (1996): 869-895.

Orr, Larry L., Howard S. Bloom, Stephen H. Bell, Fred Doolittle, Winston Lin, and George Cave. *Does Training for the Disadvantaged Work? Evidence from the National JTPA Study*. Washington, D.C.: The Urban Institute Press, 1996.

Spence, Michael. "Job Market Signaling." *Quarterly Journal of Economics* 87 (1973): 355-374.

Stiglitz, Joseph. "The Theory of Screening, Education, and the Distribution of Income." *American Economic Review* 65 (1975): 283-300.

Weiss, Andres. "Human Capital vs. Signalling Explanation of Wages." *Journal of Economic Perspectives* 9, no. 4 (1995): 133-154.