Preliminary draft
Please do not cite without permission

INCOME INEQUALITY AND THE DECISION TO DROP OUT OF HIGH SCHOOL

Melissa S. Kearney
Department of Economics
University of Maryland
and NBER
kearney@econ.bsos.umd.edu

Phillip B. Levine
Department of Economics
Wellesley College
and NBER
plevine@wellesley.edu

May 5, 2013

Acknowledgements: The authors thank Nora Gordon, Judy Hellerstein, Sue Dynarski, and Lesley Turner for helpful conversations. We also acknowledge helpful comments from seminar participants at American University School of Public Policy and Notre Dame. Any views expressed are those of the authors alone.

I. INTRODUCTION

Numerous studies have documented that individuals who come from economically disadvantaged backgrounds are more likely to drop out of high school and engage in other "risky behaviors," such as teen childbearing or juvenile crime. But even among disadvantaged youth, there are a range of observed outcomes and behaviors. To state the obvious, some low-SES youth choose to invest in their future by staying in school, delaying childbearing, and avoiding criminal behavior, while others do not. One possible explanation for this is that broader economic conditions which disadvantaged youth face have an impact on the decisions that they make. In response to an adverse economic environment, these so-called "risky" decisions may reflect a decision to "drop out" of the mainstream climb to socioeconomic success. In this paper we specifically focus on exploring the role that income inequality plays in affecting the decision to drop out of high school among disadvantaged adolescents.

Cross-sectional comparisons are striking; places with higher levels of income inequality tend to have higher rates of socially adverse outcomes.[1] Figure 1 displays this relationship across states, showing that greater inequality is related to lower graduation rates. The correlation between higher income inequality and elevated rates of high school non-completion is particularly surprising from a Beckerian perspective of human capital investment. Within such a framework, to the extent that income inequality reflects increased returns to higher education, places with greater levels of inequality should, all else equal, have higher rates of high school completion. We propose a conceptual framework nested within a standard Becker model of human capital investment, which explicitly allows for a "desperational" effect of income inequality. Specifically, we propose that greater levels of income inequality could lead a low-SES individual to have a lower expected person-specific return to investment in human capital.

[1] Wilkinson and Pickett (2009).

In this framework, income inequality has offsetting effects, and its relationship to the decision to drop out of high school becomes ambiguous.

Only around three-quarters of students who begin 9th grade in the U.S. graduate high school within four years.[2] This rate is very low by international standards.[3] And in some states, the graduation rate is even considerably lower than that. Just over half of those who start high school in Nevada graduate in that time period, as compared to close to 90 percent of students in Wisconsin. Some of this geographic variation reflects differences in background characteristics of students. It is well understood that economically disadvantaged youth have higher rates of high school non-completion than more economically advantaged youth. Murnane (2012) reports that, among students in the lowest SES quartile who were in the eighth grade in 1988, 64 percent graduated compared to compared to 95 percent among students in that cohort whose families were in the top quartile. In this paper we ask whether income inequality leads to even lower rates of high school graduation among economically disadvantaged youth.

In our empirical work we aim to distinguish a role for income inequality as distinct from individual economic disadvantage and other aggregate economic conditions, such as the median level of income or poverty rates. We use individual level data pooled from multiple sources to investigate the relationship between aggregate income inequality and individual level high school completion rates, controlling for individual background and other relevant aggregate characteristics. The key to our strategy is to determine whether any effect of inequality is concentrated among those most likely to be adversely affected by it, namely those at the bottom

[2] In this paper, we use the term high school graduation to mean received a regular diploma (no GED). Dropping out of high school also indicates no GED was received. Those who obtain a GED are treated as a separate category. The reported statistic is from Snyder and Dillow (2012).

[3] In 2009, the high school completion rate in the United States ranked 21st among 28 OECD countries (OECD, 2011). The relatively higher high school dropout rate in the United States might be considered all the more puzzling given the very high rate of return to education in the United States relative to other industrialized countries.

of the income distribution. The basic empirical question we investigate is whether the rate of high school completion among economically disadvantaged youth varies with the level of state-level income inequality. Higher-SES adolescents essentially serve as a comparison group.

We will explore the extent to which any observable effect of income inequality on the high school completion rates of low-SES individuals is mediated through channels such as public school spending and residential segregation. But whereas other studies have focused on these channels, we allow for income inequality to have a direct effect on the drop out decision. Our motivating conceptualization is that income inequality might lead to a heightened sense of economic marginalization such that an adolescent at the bottom of the income distribution does not see much value in investing in his/her human capital. This could be due to adverse neighborhood or school conditions driven by elevated rates of income inequality, but it need not. Our framework essentially offers a possible interpretation for a direct effect beyond the often considered channels of family, neighborhood, and schools.

One unique feature of our empirical analysis is our focus on long-standing, persistent variation in income inequality, not within-state variation in income inequality over time. This is deliberate. We are interested in understanding how the economic conditions of a place affect the decision to drop out or remain enrolled in high school. Cross-sectional differences in the level of income inequality are sizable and we are ultimately interested in knowing whether inequality differences across places plays a role in driving cross-sectional variation in high school dropout rates. While it is a standard approach in empirical economics to conduct panel analyses controlling for state and year fixed effects and exploiting variation in the explanatory variable of interest within a state, that is not the approach we take in this paper. In our empirical analysis, the conditional main effect of inequality is absorbed by the state fixed effect. The explanatory

variable of interest is the interaction term of income inequality and an indicator of SES status for an individual.[4]

We find that economically disadvantaged youth in more unequal places are less likely to complete high school, as compared to their counterparts in less unequal places. Among less disadvantaged youth, there are essentially no differences in high school completion rates between those in more versus less unequal places. We also find that higher rates of return to a high school degree, as measured by the ratio of wages for high school graduates to wages for high school dropouts, are positively related to the rate of high school completion among low-SES youths, which is consistent with a standard human capital model of investment. In models that control for those wage returns, the data still indicate an independent negative effect of the income gap between the middle and bottom of the income distribution on high school completion rates. These findings are consistent with our simple economic model, albeit not formal tests of its validity.

II. Relevant Literature

Various theories exist for how income inequality, as distinct from absolute income, might affect individual-level behavior. One possible channel through which income inequality can affect the social outcomes is through increased levels of residential and institutional segregation. For the poor, greater residential segregation can affect social and labor market networks, the presence of high achieving role models, and the establishment of peer groups and norms. The influential work of Wilson (1987) emphasizes the role of "social isolation" in driving rates of

4 As we describe below, our main measure of income inequality is lower-tail income inequality, defined as the ratio of household income at the 50th percentile to the 10th percentile of the distribution. We choose this measure of inequality to be our baseline measure with the thought that the economic and cultural disparities resulting from this gap are more relevant to the lives of the poor, than say, the gap between those at the 90th percentile and the median. We will separately consider alternative measures of income inequality, including wage gaps.

urban joblessness and non-marital childbearing. He hypothesizes that the lack of exposure to mainstream middle class role models plays an important role. Case and Katz (1991) provide an early example of empirical research investigating how the characteristics of one's neighbors affect an individual's outcomes.

There also exists a body of work considering the political economy implications of income inequality for the public financing of public goods. The political economy theory is ambiguous. One strand of thought suggests that income inequality can lead to increased fractionalization, with the rich becoming more politically powerful and less willing to transfer resources to an increasingly alienated poor population. Alternatively, the rich might become more socially fearful of the poor agitating for social change, and to compensate, vote for more redistribution or public good provision. The median voter model implies that increased inequality will lead to increased public good provision. The idea is that as inequality increases, the median falls relative to the mean, and the preferences of the median voter for more distribution from the rich prevail. Recent empirical evidence on the relationship between income inequality and public revenue for school spending finds support for the prediction of the median voter theorem that revenue for public school spending increases in the level of local income inequality (Boustan et al (2012), Corcoran & Evans (2012), Gordon (2013).)

An influential theory in social science posits a role for *relative deprivation* -- as distinct from absolute deprivation -- in leading to acts of social unrest.[5] Luttmer (2005) conducts an empirical economics investigation of this idea and documents that people are less happy when they live around people who are richer than themselves. A somewhat related alternative theory is

[5] The 2009 book by Wilkinson and Pickett, "The Spirit Level: Why More Equal Societies Almost Always Do Better" has been quite influential in the non-academic arena. The book promotes the idea that levels of income inequality across countries are related to negative social outcomes, including lower levels of education, but the evidence presented in that book is purely correlational and based on aggregate-level data.

that one's location in the income distribution matters in shaping one's identity construct, which affects one's decisions. Watson and McLanahan (2011) present evidence that relative income matters for the marriage decision of low-income men. They interpret their model within the framework of an identity construct.

Duncan and Murnane (2011) refer to the "ecological perspective" on the role of income inequality in determining educational outcomes. This perspective recognizes that income inequality affects families, neighborhoods, and local labor markets. Each of these elements in turn directly affects children's educational attainment in a variety of ways, but each also potentially affects school functioning, which then directly affects child educational attainment. These authors have edited a volume titled *Whither Opportunity* (subsequently referred to as the *Whither* volume) that includes a collection of essays focused on the theme of how widening income inequality over recent decades has affected the achievement, educational attainments, and labor-market experiences of low- and high-income children.

Most of the essays in the *Whither* volume focus on characterizing various features of the income gap in achievement. The chapter by Bailey and Dynarski (2011) documents a growing income-based gap in college completion. The chapter by Reardon (2011) documents that the achievement gap (as measured by standardized reading and math test scores) between children from high- and low-income families – defined as families at the 90th percentile of the family income distribution and a family at the 10th percentile, respectively -- is roughly 30 to 40 percent larger among children born in 2001 than among those born twenty-five years earlier. Reardon investigates whether the growing income gap in educational test scores is attributable to widening inequality itself. On this point, his empirical analysis does *not* provide confirming

evidence. Reardon finds that the increase in the gap is driven by the increased attainment of families above the median income level.

A key distinction between our paper and the Reardon (2011) chapter is that we are focused on documenting the relationship between income inequality and the outcomes of those on the bottom, as opposed to the income gap in outcomes or how the experiences of those at the bottom and the top of the distribution differ. Mayer (2001) is interested in the same underlying question as we are – namely, does state level inequality affect individual level educational outcomes. But, she focuses on within-state changes in inequality over time, in contrast to our focus on cross-sectional variation in income inequality. Mayer's approach is to use the 1993 PSID to estimate the relationship between an individual's educational attainment and state inequality. Her empirical approach exploits over-time variation in income inequality (as measured by the GINI coefficient) at the state level. Mayer's main specification suggests that higher levels of income inequality lead high income individuals to obtain more education and lead low-income individuals to have lower levels of educational attainment. However, when state fixed effects are included in the regression model, the analyses does not yield statistically significant effects.

III. MOTIVATING FRAMEWORK

The design of our empirical analysis is motivated by the hypothesis that inequality can negatively affect the perceived returns to investment in education from the perspective of an economically disadvantaged adolescent. The notion we have in mind is that a greater gap between the bottom and the middle of the income distribution might lead to a heightened sense of economic marginalization such that an adolescent at the bottom of the income distribution does not see much value in investing in his/her human capital. This could be due to adverse

neighborhood or school conditions driven by elevated rates of income inequality, but it need not. Our framework essentially offers a possible interpretation for a direct effect beyond the often considered channels of family, neighborhood, and schools. And more importantly, it offers an explanation within the standard human capital framework of decision-making for why greater inequality – which might reflect in part a greater return to human capital investment – does not necessarily lead to greater rates of educational attainment for certain segments of the population.

We offer here an extremely stylized framework. An individual chooses to drop out of school in the current period if the following condition is met:

$$u_o^d + E(V^d) > u_o^e + E(V^e) \quad (3)$$

where u^d is current period utility if the student drops out and u^e is current period utility if s/he remains enrolled. V is the present discounted sum of future period utility.

If $u^d < u^e$, it is never optimal to drop out. But if $u^d > u^e$, which would be the case if the student experiences substantial utility costs from remaining in school (e.g. psychic costs), then that current period utility boost needs to be compared to the potential option value lost. Dropping out of school negatively affects expected future utility by leading to lower levels of consumption in the future. For simplicity, we characterize utility in future periods as taking high and low values, U^{high} and U^{low}, respectively. We assume that dropping out reduces the likelihood of achieving U^{high}. We define U^{low} to be the level achieved by a student who does drop out. The present discounted value of the future utility stream is thus deterministic and captured by V^{low}. If the adolescent remains enrolled, there is some positive probability p that s/he will achieve the "high" utility position in future periods.

We can therefore write the condition to drop out of school as follows:

$$u_o^d + V^{low} > u_o^e + pV^{high} + (1-p)V^{low} \quad (4)$$

This condition indicates that the change in lifetime utility from delayed childbearing comes from two opposite-signed sources: (1) the loss of current period enjoyment of a baby and (2) a positive probability of achieving the high- utility state in the future. Rearranging terms, we see that a student will choose to remain enrolled if and only if:

$$pV^{high} + (1-p)V^{low} > V^{low} + (u_o^d - u_o^e) \quad (5)$$

Of course, the student does not perfectly observe p, as in Manski (1993).[6] Instead, the student bases the decision on his perception of p, in particular, on his perception of his individual-specific p. Let us call this subjective probability of one's individual likelihood of success conditional on investment q, and rewrite the condition for deciding not to drop out as follows:

$$qV^{high} + (1-q)V^{low} > V^{low} + (u_o^d - u_o^e) \quad (6)$$

If an adolescent perceives that s/he has a sizable chance at achieving economic success -- and thereby capturing V^{high} -- by investing in education, the comparison is more likely to favor the choice "stay enrolled." On the other hand, if the student perceives that even if s/he stays enrolled, his/her person-specific chances of economic success are sufficiently unlikely -- in other words, if q is very low -- then the comparison is more likely to favor dropping out in the current period.

[6] Jensen (2010) highlights this point and suggests that the perception of returns might be particularly inaccurate in developing country settings, which could potentially explain (in part) why rates of educational attainment remain low despite high measured returns. He conducts an experiment among 8th grade boys in the Dominican Republic whereby students at randomly selected schools are given information about the measured returns to completing school. He finds that male students at schools who receive this information complete an average of 0.20-0.35 additional years of school over the subsequent four years as compared to male students in the comparison schools.

We speculate that for an adolescent at the bottom of the income distribution, a greater gap between one's position and the middle of the distribution might have a "desperational" effect on one's subjective *q*. If the middle class is sufficiently far from one's own experience, than the student's perceived chances of getting there -- even if he/she does stay in school -- are sufficiently low, There are many mechanisms that could drive these negative perceptions, such as social isolation of the kind described by Wilson (1987), low quality schools or peer networks, or other psychological effects. We do not purport to exhaustively test for various mechanisms in this paper. Our main goal with the empirical analyses of the current paper is to determine whether there does appear to be an effect of income inequality on drop-out rates, conditional on rates of disadvantage and other obviously relevant features of the state environment.

In Kearney and Levine (forthcoming) we proposed a similar model to characterize the decision of a young, unmarried woman to have a baby in the current period or to delay childbearing. In that paper we offer some empirical support for the proposition that low-SES adolescents growing up in relatively more unequal places actually do have a lower chance of achieving higher income in later life. To test that idea, we examine data from the restricted-use NLSY79 Geocode data.[7] The regression results show that children who grow up in low SES households and who live in a state with high lower-tail inequality are estimated to have permanent incomes that are over 30 percent lower than similar children in low lower-tail inequality states. High and low inequality states are distinguished by a one point increase in the

[7] We distinguish youth respondents by their parents' educational attainment and define "permanent income" to be the average of all inflation-adjusted values of family income observed 15 or more years after the original 1979 survey, when youth respondents are in their late 20s or older. The sample used in that exercise includes all 8,226 respondents who lived with at least one of their parents at age 14 and who provided any income values in the 1994 survey or beyond. We assign the level of inequality to each respondent based on the respondents' 1979 state of residence.

50/10 ratio. If perceptions of economic opportunity are gauged on actual outcomes, then these findings are consistent with our proposition.

This framework has important implications for how to conduct our empirical analysis in terms of the appropriate level of geography. The way we are thinking about the possible effects of income inequality implies that the appropriate unit is a fairly broad area, such as a state or an MSA. These would allow for the effects of any type of residential or institutional segregation that might occur as a result of widened income inequality and affect perceptions of success. If we were motivated by relative deprivation theories, we would instead want to define income inequality much more locally. In the current version of this paper, we conduct our analysis at the state level. In future work we will look to more localized areas in the interest of comparing estimated effects.

IV. EMPIRICAL STRATEGY & DATA

As we previewed earlier, the goal of our econometric analysis is to determine whether teens from disadvantaged backgrounds who live in areas that exhibit relatively high rates of long-term income inequality experience higher rates of dropping out of high school. In essence, this question sets up a quasi-experiment where less disadvantaged teens form a pseudo-control group. These teens live in areas that differ by their level of persistent income inequality as well and they serve as a means of determining whether attributes of these environments generate differential rates of high school completion. What we seek to determine is whether those differences are exacerbated among the group of disadvantaged teens. There are weaknesses to this identification strategy, as we describe and address subsequently, but this logic forms the basis of our identification strategy.[8]

More formally, the econometric model that we estimate takes the form:

[8] This empirically strategy is conceptually identical to that which we used in Kearney and Levine (forthcoming).

$$Outcome_{is} = \beta_0 + \beta_1 (I_s \cdot LS_{is}) + \beta_2 (I_s \cdot MS_{is}) + \beta_3 LS_{is} + \beta_4 MS_{is} + \beta_5 X_{is} + \beta_5 E_s + \gamma_s + \varepsilon_{is} \quad (1)$$

where the outcome is some measure of educational attainment (high school dropout, GED, or high school graduate), I is our measure of inequality, LS and MS are indicators of low and middle SES, respectively, and the interaction terms are the main regressors of interest. Their coefficients represent the differential response of low and middle SES teens to inequality relative to higher SES teens. The subscripts i and s index individuals and states, respectively; γ_s represents state fixed effects. The vector **X** consists of additional personal demographic characteristics – race/ethnicity and an indicator for living with a single parent at age 14. The vector **E** captures environmental factors including relevant public policies and labor market conditions in the state-year.[9] By including all of these individual and state level controls in the model, our estimated effect of inequality for low-SES women is net of effects driven by policies that might be correlated with inequality.

The main shortcoming with this empirical strategy is that any omitted, state-specific factor that is fixed over time and correlated with long-term measures of income inequality may generate biased results. Problems will arise if they are, indeed, the factors that generate differences in educational attainment across SES groups. We have no definitive approach to resolve this problem, but we do implement a method designed to determine whether potentially likely alternatives are playing this role. To do so, we estimate regression models of the form:

$$\begin{aligned} Outcome_{is} = {} & \beta_0 + \beta_1 (I_s \cdot LS_{is}) + \beta_2 (I_s \cdot MS_{is}) + \beta_3 (A_s \cdot LS_{is}) + \beta_4 (A_s \cdot MS_{is}) \\ & + \beta_5 LS_{is} + \beta_6 MS_{is} + \beta_7 X_{is} + \beta_8 E_s + \gamma_s + \varepsilon_{is} \end{aligned} \quad (2)$$

[9] These variables include: the unemployment rate, an indicator for a welfare family cap, the maximum welfare benefit for a family of three, an indicator for SCHIP implementation, an indicator for whether the state Medicaid program covers abortion, an indicator for whether state abortion regulations include parental notification or mandatory delay periods, and whether the state Medicaid program includes expansion policies for family planning services (see Kearney and Levine, 2009a for a discussion of these expansion policies).

In essence, our approach involves including potential alternative state factors (A_s) that could plausibly introduce bias and examine whether the results change when we include them in the same manner in which we have included the inequality/SES interactions. If the coefficients on the interaction terms of primary interest change when we add the additional interactions between SES and these alternatives, then it would suggest the results generated from equation (1) are biased. It is impossible to rule out this form of bias unless we try including every possible alternative, but if what we believe are important alternatives have no impact, then we can be more confident in a causal interpretation of our findings.

When we implement this approach, we consider four categories of these other state factors that are designed to examine four alternative sets of hypotheses. The first is about measurement. As we describe subsequently, we use the 50/10 ratio as our measure of inequality. But perhaps that is not the most relevant features of the income distribution. The alternatives we explore include the 90/50 ratio and the 10^{th} and 50^{th} percentiles of the income distribution. The second alternative set of factors we include are measures of the returns to education. This is important because it enables us to identify the incentive effect of higher returns (as in a standard Becker model) separately from any offsetting "desperational effect" of the type we propose. Third, we consider a set of alternatives that could be considered mediating factors. What is the mechanism by which increased inequality alters educational attainment? If we include these mechanisms in the model as we express in Equation (2), we should see a change in the estimated impact of the 50/10 ratio. Finally, we include a set of potential confounding factors that would be more typically addressed when thinking about problems of omitted variable bias.

To estimate these models, we use four sources of individual-level data. Two of the sources are available from the National Center for Education Statistics (the National Educational

Longitudinal Survey – NELS, and the Educational Longitudinal Survey – ELS) and the other two are different cohorts (1979 and 1997) of the National Longitudinal Survey of Youth (NLSY79 and NLSY97).[10] Each of these datasets has the distinct advantage of including detailed measures of educational attainment, including the ability to separately identify those who receive a degree through passing a general educational development (GED) test and those who receive a traditional high school degree. Their combination also generates a sample of tens of thousands of teens who are moving through (or just recently completed) their high school years. The NLSY79 originally surveyed 12,686 respondents born between 1957 and 1964 (age 14-22 in 1979). NELS surveyed 14,915 8th graders in 1988 who were also surveyed in 1994, when we can determine whether they completed high school. NLSY97 surveyed 8,984 respondents born between 1980 and 1984 (age 12-18 in 1997). ELS surveyed 15,300 10th graders in the spring of 2002 who were also surveyed in 2006, when high school completion can be measured. In combination, a maximum of 51,885 respondents are available. In reality, mainly because of missing state identifiers, missing information regarding SES (defined below as level of maternal education), and sample attrition (in the NLSY surveys) we have available 38,590 teens for our analysis. Limited time variability is available when we combine these datasets, but our analysis relies on long-term geographic variability anyway, as we described earlier.

A critical feature of these data for the purposes of estimating the models described earlier is that we need a measure of their socioeconomic status. The measure that is available in each of these datasets that we are able to use is mother's level of education. We distinguish students according to whether their mother dropped out of high school, graduated high school, or attended

[10] Although public use versions of each of these datasets exist, to obtain state identifiers that are necessary for our analysis restricted use agreements are required. This means that we are not able to share our data with other researchers, although we are happy to provide our programs so that those who are able to obtain their own agreement can follow our steps. Another comparable NCES survey, High School and Beyond, is not suitable for this project because one cannot obtain state identifiers even under a restricted use agreement.

college (regardless of their graduation status). Although maternal education does not perfectly predict economic status, it is strongly correlated with SES and we take advantage of that in our analysis.

Although the availability of all four of these datasets provides a unique opportunity to generate a large sample of high school students and follow them through the completion (or not) of their degree, their combination also presents challenges. In particular, identifying a consistently selected sample and outcome measures is somewhat complicated. Sample selection is an issue because individuals entered the samples at different ages/grades. The NELS, for instance, initially surveyed 8th graders whereas the ELS initially surveyed 10th graders. Survival to 10th grade, though, represents a degree of success that changes the composition of the sample since more poorly performing students may never make it to 10th grade. We address issues like these in the attached data appendix. In terms of outcomes, we have chosen to focus on a consistent measure of educational attainment defined by high school completion status by age 20. In each of these datasets, we are able to determine whether a student completed high school and received a traditional diploma, whether the student received a GED, or whether the student never obtained a high school degree via either route. These three indicators of educational attainment by age 20 represent our outcome measures.

Our main measure of inequality is lower-tail income inequality, which we empirically capture with the gap between the 50th percentile and the 10th percentile of total household income. The 50/10 ratio is our preferred measure of inequality because it focuses on what those at the bottom of the income distribution could conceivably achieve.[11] We calculate these

[11]The increase in income inequality in the United States since the 1980s is characterized by consistently rising upper-tail income inequality, meaning that the gap between the 90th percentile and the 50th percentile has continued to expand (c.f. Autor, Katz, Kearney, 2008). Our speculation is that those at the bottom of the distribution are less

measures by state and survey year using microdata from the 1980, 1990, and 2000 Censuses along with the 2006-2008 American Community Surveys on household income. These data are available from IPUMS-USA (Ruggles, 2010). We then take the long-term average over all years for a state. We do so because we are trying to capture something about the permanent or semi-permanent economic and cultural landscape in the place where an adolescent lives, as opposed to short-term fluctuations.[12] We focus on (semi-)permanent characteristics of states because we are interested in the way income inequality affects individual's decisions and behaviors and have in mind a model where it does so by affecting their experiences, perceptions, and aspirations. We are not thinking about responses to short-term or temporary conditions. If a state experiences a temporary decrease in lower-tail income inequality, it is unlikely that neighborhoods will change sufficiently quickly and sufficiently visibly that either economic opportunities or the perceptions thereof will be altered.

V. RESULTS

To highlight the identification strategy that we use, we initially present the results of a descriptive analysis of educational outcomes for teens by their socioeconomic status and the level of income inequality that exists in their state. Figures 2 and 3 present the results of this descriptive analysis. We classify states into those in the top and bottom quartiles of inequality as measured by the 50/10 ratio, and the middle two quartiles. The bars in these figures represent the percentage of students who did not complete high school (Figure 2), obtained a GED (Figure 3), or received a traditional high school diploma (Figure 4) by the age of 20. Students are separated

focused on those at the very top, which would make the 50/10 ratio a better measure than, say, the 90/10 ratio. In our past work on teen pregnancy, we found that the 50/10 ratio was the most empirically relevant measure.

[12]It is also the case that there is much more cross-sectional variation in lower-tail income inequality across states as compared to within a state over time. Across states the standard deviation in the 50/10 ratio is 0.42. Within states, the average standard deviation in the 50/10 ratio across the four observed values (3 Census years plus the aggregated ACS years) is 0.17.

into categories according to their mother's educational attainment to proxy for the SES, along with the level of inequality that exists in their state.

The framework of our strategy is to treat higher SES students as a comparison group for lower SES students. Each of these figures groups SES categories so that the pattern in educational outcomes by inequality status within SES category is readily apparent. For instance, in Figure 2, we see that not quite 5 percent of students from higher SES families drop out of high school regardless of the level of income inequality that they face in their state. No obvious pattern is evident among the middle SES students in different inequality categories either. On the hand, among low SES students, higher inequality is associated with higher rates of dropping out of high school. The magnitude of the difference is sizeable. Low SES students in high inequality states are about 5 percentage points more likely to drop out of high school than low SES students in low inequality states. Although there are some problems with this interpretation, as we described earlier and address subsequently, if higher SES students represent a suitable control group, we can attribute causation to this relationship.

Figures 3 and 4 display the same analysis for GED receipt and high school completion. No obvious pattern in GED receipt appears across the three SES categories at different levels of inequality. If GED receipt is unaffected, then the impact of inequality on high school graduation must be the opposite of what we observed in Figure 2 regarding the likelihood of dropping out of high school. This pattern is exactly what we see in Figure 4. Although there is a slight pattern of greater graduation rates among high SES students, that pattern is considerably more pronounced among low SES students. Again, the difference in high school graduation among low SES students in high and low inequality states is about 5 percentage points, the opposite of what we observed regarding dropping out of high school. These results suggest that GED receipt is not the

relevant margin on which students respond as inequality changes. It is about the decision to drop out or graduate that matters.

These findings from our descriptive analysis are affirmed when we estimate the regression models described in equation 1. In essence, these regressions are analogous to the data reported in Figures 2 through 4 with the exception that the 50/10 ratio is treated continuously rather than in categories and additional explanatory variables are included. As such, it is perhaps not surprising that the estimation results closely mimic those obtained in the graphical analysis we just presented.

Table 1 presents those results for all students in the sample and then separately for boys and girls. Columns 1 through 3 are identical except they focus on our three different measures of educational outcomes (high school dropout – Column 1; GED – Column 2; and high school graduation – Column 3). The percentage of students in each category is displayed just above the regression results to help aid in interpretation. When we focus on dropping out of high school for all students (the top panel of the table), we see that a one point increase in the 50/10 ratio increases the likelihood of dropping out by 5.1 percentage points and 3.2 percentage points for students from low and middle SES families, respectively. Because the high and low inequality states in the earlier figures have 50/10 ratios that differ by about one point, the impact of inequality on low SES students, in particular, is virtually identical to what we observed earlier. One conclusion we can draw from this is that the additional explanatory variables that we have included in these models are not a relevant source of omitted variable bias. As before, we see little impact on GED receipt and an estimated impact on high school graduation that is almost exactly the negative of the impact on dropping out of high school.

The remainder of the table reports the results of estimating identical models separately for boys and for girls. The point estimates indicate that the impact on dropping out for low SES students in high inequality locations is larger for boys than it is for girls, but we do not have adequate precision to determine that this difference is statistically significant. Table 2 reports the results of a similar exercise, distinguishing students by race/ethnicity. The notable finding in this table is that the largest impact on dropping out of high school is observed among black, non-Hispanic, low SES students in high inequality areas. For these students moving from a low to a high inequality state (increase the 50/10 ratio by one) would reduce the dropout rate by 9.4 percentage points. This is a large estimate based on the average dropout rate for this group of around 17.4 percent.

In the next set of tables, we estimate models of the form of Equation 2 that are designed to examine the extent to which other state-specific factors may matter and alter our interpretation of a causal impact of income inequality. In these tables, we focus solely on the outcome of dropping out of high school. We do this because we are unable to identify any impact on GED receipt, which makes it superfluous to estimate models of dropping out and graduating high school. Table 3 addresses what are the right components of the income distribution that affect educational attainment. The alternatives we consider are the 90/50 ratio, and, separately, the 10^{th} and 50^{th} percentiles of the income distribution. Each of the alternative measures of the income distribution capture different attributes. The 90/50 ratio represents income inequality at the top of the income distribution. This is part of the distribution that has grown over time. We have argued that the 50/10 ratio is a better measure of inequality for the low SES population because it may more realistically indicate what is available to them if they were able to move up the ladder, but this is an empirical question. We also include the 10^{th} and 50^{th} percentiles of the income

distribution separately to understand whether our findings based on their ratio are actually attributable to one of the two components separately.

As described earlier, the approach we take here is to include the 50/10*SES interactions along with interactions between SES and these other measures. We can directly interpret the coefficients on those interactions and we can also observe whether substantive changes occur in the coefficients on the 50/10*SES interactions. The results based on this approach are reported in Table 3. These estimates provide no substantive reason for changing our earlier conclusions that the 50/10 ratio is the appropriate measure of income inequality to consider. If anything, including the 90/50 ratio strengthens the relationship between the 50/10 ratio among low SES students and dropping out of high school. Interactions with the other measures are generally statistically insignificant.[13]

The purpose of Table 4 is to more directly consider how inequality may play a distinct offsetting role from the incentive effect that is central to a standard human capital investment model focused on returns to education. Recall from our earlier discussion that to the extent greater inequality is capturing a greater return to investment in human capital, the Beckerian framework predicts that all else equal, students should invest more, which in this case, means drop out less often. But as we have seen, our analyses are finding that greater levels of inequality lead low-SES adolescents to drop out more often. In Column (2) we estimate a regression model that includes separate interaction terms for low-SES status and (a) lower-tail inequality and (b)

[13] In column 2 we see that the estimated coefficient on the interaction between the 90/50 ratio and an indicator for being the child of a high school graduate mother is negative and marginally statistically significant. One potential explanation for this is that when upper tail inequality increase, high-SES adolescents (captured with an indicator for having a mother with college education) are more likely to complete high school, which makes this middle group, by comparison, appear less likely to complete high school, even if their own behavior (in levels) does not change.

the wage premium for high school graduates relative to high school dropouts. The high school graduate wage premium is calculated from the same Census and ACS data that we used to estimate measures of inequality.

The results of this specification indicate two separate effects. Consistent with the standard prediction of the human capital model, increases in the high school wage premium lead to a lower rate of drop out among low SES adolescents. Importantly, though, even with this additional interaction term in the model, the point estimate on the interaction term between low-SES status and lower tail inequality is virtually unchanged from the initial specification. The data indicate a positive effect of income inequality on the likelihood that a disadvantaged youth drops out of school, conditional on the high school wage premium.

In Column (3) we conduct an analogous exercise, focusing on the college/high school wage premium. This specification is less straightforward to interpret. On the one hand, an important part of the return to completing high school is the option value of attending college and being in a position to receive the college wage premium. If that effect dominates, than an increase in the college wage premium should, all else equal, lead to lower dropout rates (according to the human capital model). On the other hand, for students who do not plan to attend college or see college as a viable possibility, when the college wage pulls away from the high school wage, lowering either the actual or perceived return to high school without college, then those students might be more inclined to drop out. Indeed, in Column (3), the estimated coefficients on the interaction terms with the college wage premium are not statistically significant. However, we continue to see the statistically significant point estimate of around 0.05 on the interaction term of the 50/10 ratio and low-SES status.

Table 5 continues this methodological approach to explore the extent to which we can identify mediating factors that may play a role in altering educational outcomes in the presence of greater inequality. Increased income inequality may have an impact on inequality in educational production or the nature of the residential environment in which the students live. We explore these possibilities by consider the following alternative state characteristics – an index of racial segregation, per capita educational expenditures, and pupil-teacher ratios.[14] To the extent that any of these factors, when interacted with SES, have a statistically significant effect and/or alter the estimated impact of the SES*50/10 ratio interactions, one could conclude that they are important mediating factors. The results reported in Table 5 provide no such evidence of this sort of effect. None of the coefficients on the interactions with these factors are statistically significant and their inclusion has a negligible impact on the SES*inequality interactions.

It is important to interpret these last set of results correctly. We do not mean to definitively conclude from them that educational production or segregation is not a mediating factor. It is quite conceivable that we have measured them sufficiently crudely that we are unable to capture their impact. One alternative interpretation, though, is that income inequality has a direct impact on educational attainment through perceptions of economic success that are not transmitted through other, previously identified channels. Yet it would also be inappropriate to

[14]We thank Liz Cascio for generously sharing the historical data she compiled on per pupil expenditures and per pupil teacher ratios. The racial segregation index was obtained from the University of Michigan Population Studies Center website:
http://www.psc.isr.umich.edu/dis/census/segregation.html, accessed on April 25, 2012.
The source is William H. Frey, Brookings Institution and University of Michigan Social Science Data Analysis Network's analysis of 2005-9 American Community Survey and 2000 Census Decennial Census tract data. The index is constructed as a Dissimilarity Indices that measure the degree to which the minority group is distributed differently than whites across census tracts. They range from 0 (complete integration) to 100 (complete segregation) where the value indicates the percentage of the minority group that needs to move to be distributed exactly like whites.

conclude from these results that we have proven that either. It remains what we believe to be an intriguing hypothesis that deserves further attention.

In the last set of "horse race" specifications, Table 6 presents the results of including one additional set of interactions with other state-specific factors that could simply represent confounding factors. These include the percentage of the state's population that is minority, the poverty rate in the state, and the state's incarceration rate. The goal here is determine whether there is some other state-specific factor that is capturing the social environment of the state that may be related to inequality and driving the differential high school dropout rates. Based on the results reported in Table 6, there is no evidence that this is the case. Interactions between any of these factors and socioeconomic status are universally insignificant and their inclusion in the regression model has no substantive impact on the estimated effect of the interactions between lower-tail inequality and individual SES status.

VI. DISCUSSION

This paper has presented empirical evidence that economically disadvantaged youth in more unequal places are less likely to complete high school, as compared to their counterparts in less unequal places. Among less disadvantaged youth, there are essentially no differences in high school completion rates between those in more versus less unequal places. The analysis also indicates that higher rates of return to a high school degree, as measured by the ratio of wages for high school graduates to wages for high school dropouts, are positively related to the rate of high school completion among low-SES youths, which is consistent with a standard human capital model of investment. In models that control for these wage returns, the data indicate a distinct negative effect of the income gap between the middle and bottom of the income distribution on high school completion rates. We interpret these results within the context of a simple framework that proposes a "desperational" role for income inequality that offsets the incentive

effects for those at the bottom of the income distribution. The results of our empirical analysis are consistent with our simple economic model, albeit not formal tests of its validity.

We ultimately have little to say about the precise mechanisms by which income inequality leads to the "desperational" effect that we document. We have estimated specifications that attempt to determine whether the effect of inequality is being driven by school expenditure levels, and the data do not support that hypothesis. We have also considered alternative characteristics, including a measure of residential racial segregation. More research is warranted into the mechanisms underlying the observed empirical relationship between lower-tail income inequality and high school completion rates.

REFERENCES

Altonji, Joseph G. and Richard K. Mansfield. 2011. "The Role of Family, School, and Community Characteristics in Inequality in Education and Labor-Market Outcomes." In Greg J. Duncan and Richard J. Murnane, eds*., Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, New York: Russell Sage Foundation.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." Review of Economics and Statistics 90(2), 300-323.

Bailey, Martha J. and Susan M. Dynarski. 2011. "Inequality in Postsecondary Education," in Ed. Greg J. Duncan and Richard J. Murnane, eds., Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances, New York: Russell Sage Foundation: Chapter 6.

Boustan, Leah, Fernando Ferreira, Hernan Winkler, and Eric Zolt. 2012. "The Effect of Rising Income Inequality on Taxation and Public Expenditures: Evidence from US Municipalities and School Districts, 1970-2000." Forthcoming, Review of Economics and Statistics.

Cascio, Elizabeth, and Sarah Reber. 2012. "The War on Poverty's K-12 Education Battle: The History and Legacy of Title I." NBER working paper…

Corcoran, Sean and William N. Evans. 2010. "Income inequality, the median voter, and the support for public education." Cambridge, MA: NBER working paper 16097.

Dee, Thomas S. and Brian A. Jacob, 2006. "Do High School Exit Exams Influence Educational Attainment or Labor Market Performance?" NBER working paper 12199, National Bureau of Economic Research.

Dobbie, William and Roland Fryer. 2012. "Are High Quality Schools Enough to Reduce Social Disparities? Evidence from the Harlem Children's Zone," Harvard University mimeo, December.

Duncan, Greg J., and Richard J. Murnane. 2011. "Introduction: The American Dream, Then and Now." In Greg J. Duncan and Richard J. Murnane, eds., Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances. New York: Russell Sage Foundation.

Gordon, Nora. 2013. "High School Graduation in the Context of Changing Elementary and Secondary Education Policy and Income Inequality: The Last Half Century," Georgetown University mimeo, April.

Guryan, Jonathan, Eric Hurst, and Melissa Kearney. 2008. "Parental Education and Parental Time with Children," Journal of Economic Perspectives 22(3): 23-46.

Heinrich, Carolyn and Harry J. Holzer, "Improving Education and Employment for Disadvantaged Young Men: Proven and Promising Strategies," in the Annals of the American Academy of Political and Social Science 2011 635: 163-191.

Heisz, Andrew, Phil Oreopoulos, and Till Von Wachter (forthcoming), "Short- and Long-Term Career Effects of Graduating in a Recession," American Economic Journal: Applied Economics.

Jensen, Robert. 2010. "The (Perceived) Returns to Education and the Demand for Schooling," Quarterly Journal of Economics, 125(2), p. 515-548.

Kearney, Melissa and Phil Levine (forthcoming). "Income Inequality and Early Non-Marital Childbearing," *Journal of Human Resources*.

Lareau, Annette. 2003. Unequal Childhoods: Class, Race, and Family Life." Berkeley: University of California Press.

Lillard, Dean R. & DeCicca, Philip P., 2001. "Higher standards, more dropouts? Evidence within and across time," Economics of Education Review 20(5): 459-473.

Luttmer, E. F. P. 2005. "Neighbors as Negatives: Relative Earnings and Well-Being." Quarterly Journal of Economics 120(3): 963-1002.

Manski, Charles F.. 1993. "Adolescent Econometricians: How Do Youth Infer the Returns to Education?" in Studies of Supply and Demand in Higher Education, Charles T. Clotfelter and Michael Rothschild, eds.

Mayer, Susan. 2001. "How did the Increase in Economic Inequality between 1970 and 1990 Affect Children's Educational Attainment" American Journal of Sociology 107(1):1-32.

OECD (2011). *Education at a Glance, 2011: OECD Indicators*. Paris: OECD Publishing.

Reardon, Sean. 2011. "The Widening Academic Achievement Gap between the Rich and the Poor: New Evidence and Possible Explanations," in Ed. Greg J. Duncan and Richard J. Murnane, Whither Opportunity, New York: Russell Sage: Chapter 5.

Snyder, Thomas D., and Dillow, Sally A. (2012). Digest of Education Statistics 2011 (NCES 2012-001). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Watson, Tara. 2009. "Inequality and the Measurement of Residential Segregation by Income." Review of Income and Wealth 55: 820-844.

Watson, Tara, and Sara McLanahan. 2011. "Marriage Meets the Joneses: Relative Income, Identity, and Marital Status." Journal of Human Resources 46: 568-586.

Wilkinson, Richard and Kate Pickett. 2009. The Spirit Level: Why Greater Equality Makes Societies Stronger. London: Bloomsbury Press.

Wilson, William J. 1987. The Truly Disadvantaged: The Inner City, the Underclass and Public Policy. Chicago: University of Chicago Press.

Table 1: Impact of Long-Term Inequality on Educational Attainment by Age 20, by Socioeconomic Status

	High School Dropout (1)	GED Receipt (2)	High School Graduate (3)
		All	
Percent in Category	11.8	5.2	83.0
50/10 Ratio*	0.051	0.003	-0.054
Mom HS Dropout	(0.019)	(0.015)	(0.018)
50/10 Ratio*	0.032	-0.001	-0.032
Mom HS Graduate	(0.011)	(0.008)	(0.013)
		Boys	
Percent in Category	13.0	5.8	81.2
50/10 Ratio*	0.062	-0.012	-0.050
Mom HS Dropout	(0.022)	(0.020)	(0.021)
50/10 Ratio*	0.040	0.001	-0.041
Mom HS Graduate	(0.016)	(0.011)	(0.018)
		Girls	
Percent in Category	10.7	4.6	84.7
50/10 Ratio*	0.039	0.015	-0.053
Mom HS Dropout	(0.024)	(0.015)	(0.026)
50/10 Ratio*	0.023	-0.003	-0.020
Mom HS Graduate	(0.012)	(0.013)	(0.017)

Notes: reported standard errors are adjusted for clustering at the state level. Additional explanatory variables in each regression include maternal educational attainment, race/ethnicity, an indicator variable for living with a single parent at age 14, the state unemployment rate at age 16, state education policies (compulsory schooling age and indicators for high school exit exam requirements), state welfare policies (family cap and maximum AFDC/TANF benefit for a family of 3), state abortion policies (Medicaid funding, parental notification/consent, and mandatory delay laws), and an indicator variable for SCHIP implementation, along with state and cohort fixed effects. The total sample size is 38,590, with 18,815 boys, 19,775 girls.

Table 2: Impact of Long-Term Inequality on Educational Attainment by Age 20, by Socioeconomic Status

	High School Dropout (1)	GED Receipt (2)	High School Graduate (3)
		White, Non-Hispanic	
Percent in Category	9.8	4.9	85.4
50/10 Ratio*	0.041	-0.001	-0.041
Mom HS Dropout	(0.033)	(0.017)	(0.031)
50/10 Ratio*	0.000	-0.006	0.006
Mom HS Graduate	(0.011)	(0.011)	(0.017)
		Black, Non-Hispanic	
Percent in Category	17.4	6.7	76.0
50/10 Ratio*	0.094	-0.003	-0.091
Mom HS Dropout	(0.034)	(0.026)	(0.021)
50/10 Ratio*	0.112	-0.028	-0.084
Mom HS Graduate	(0.020)	(0.012)	(0.022)
		Hispanic	
Percent in Category	19.1	5.5	75.4
50/10 Ratio*	-0.019	0.062	-0.043
Mom HS Dropout	(0.031)	(0.033)	(0.042)
50/10 Ratio*	0.005	0.097	-0.102
Mom HS Graduate	(0.043)	(0.035)	(0.049)

Notes: See notes to Table 1. The total sample sizes are 23,432 non-Hispanics, 6,322 black, non-Hispanics, 5,816 Hispanics.

Table 3: Impact of Alternative State Economic Conditions on the Likelihood of Dropping Out of High School, by Socioeconomic Status
(standard errors in parentheses)

	50/10 ratio (1)	90/50 ratio (2)	10th Percentile of Income (in $10,000s) (3)	50th Percentile of Income (in $10,000s) (4)
Correlation between 50/10 ratio and characteristic:		0.72	0-.55	-0.13
50/10 Ratio*	0.051	0.083	0.052	0.051
Mom HS Dropout	(0.019)	(0.026)	(0.022)	(0.019)
50/10 Ratio*	0.032	0.019	0.030	0.031
Mom HS Graduate	(0.011)	(0.014)	(0.014)	(0.012)
State Characteristic*	---	-0.116	0.003	-0.003
Mom HS Dropout	---	(0.067)	(0.031)	(0.007)
State Characteristic*	---	0.040	-0.004	-0.002
Mom HS Graduate	---	(0.035)	(0.018)	(0.004)

Notes: see notes to Table 1.

Table 4: Impact of Educational Wage Premiums on the Likelihood of Dropping Out of High School, by Socioeconomic Status (standard errors in parentheses)

	50/10 ratio (1)	HS Grad to HS Dropout Wage Premium (2)	College Grad to HS Grad Wage Premium (3)
Correlation between 50/10 ratio and characteristic:		0.17	0.41
50/10 Ratio*	0.051	0.054	0.053
Mom HS Dropout	(0.019)	(0.018)	(0.019)
50/10 Ratio*	0.032	0.030	0.028
Mom HS Graduate	(0.011)	(0.012)	(0.012)
State Characteristic*	---	-0.134	-0.019
Mom HS Dropout	---	(0.064)	(0.065)
State Characteristic*	---	0.043	0.032
Mom HS Graduate	---	(0.049)	(0.030)

Notes: see notes to Table 1.

Table 5: Impact of Alternative State Characteristics on the Likelihood of Dropping Out of High School, by Socioeconomic Status
(standard errors in parentheses)

	50/10 ratio (1)	Segregation Index (3)	Per Capita Educational Expenditures (x 1,000) (6)	Pupil Teacher Ratio (x10) (7)
Correlation between 50/10 ratio and characteristic:		0.09	0.25	-0.29
50/10 Ratio*	0.051	0.049	0.043	0.041
Mom HS Dropout	(0.019)	(0.021)	(0.018)	(0.018)
50/10 Ratio*	0.032	0.033	0.031	0.029
Mom HS Graduate	(0.011)	(0.010)	(0.009)	(0.011)
State Characteristic*	---	0.0007	-0.0006	-0.0002
Mom HS Dropout	---	(0.0007)	(0.0025)	(0.019)
State Characteristic*	---	-0.0006	-0.0038	0.018
Mom HS Graduate	---	(0.0004)	(0.0013)	(0.014)

Notes: see notes to Table 1.

Table 6: Impact of Alternative State Characteristics on the Likelihood of Dropping Out of High School, by Socioeconomic Status
(standard errors in parentheses)

	50/10 ratio (1)	Percent Minority (2)	Poverty Rate (3)	Incarceration Rate (x1,000) (4)
Correlation between 50/10 ratio and characteristic:		0.37	0.55	0.38
50/10 Ratio*	0.051	0.060	0.066	0.045
Mom HS Dropout	(0.019)	(0.020)	(0.023)	(0.019)
50/10 Ratio*	0.032	0.026	0.027	0.019
Mom HS Graduate	(0.011)	(0.011)	(0.014)	(0.010)
State Characteristic*	---	-0.0006	-0.0031	-0.032
Mom HS Dropout	---	(0.0004)	(0.0018)	(0.067)
State Characteristic*	---	0.0003	0.0008	0.056
Mom HS Graduate	---	(0.0002)	(0.0011)	(0.036)

Notes: see notes to Table 1.

Figure 1: Relationship between Inequality and 4-Year High School Graduation Rate

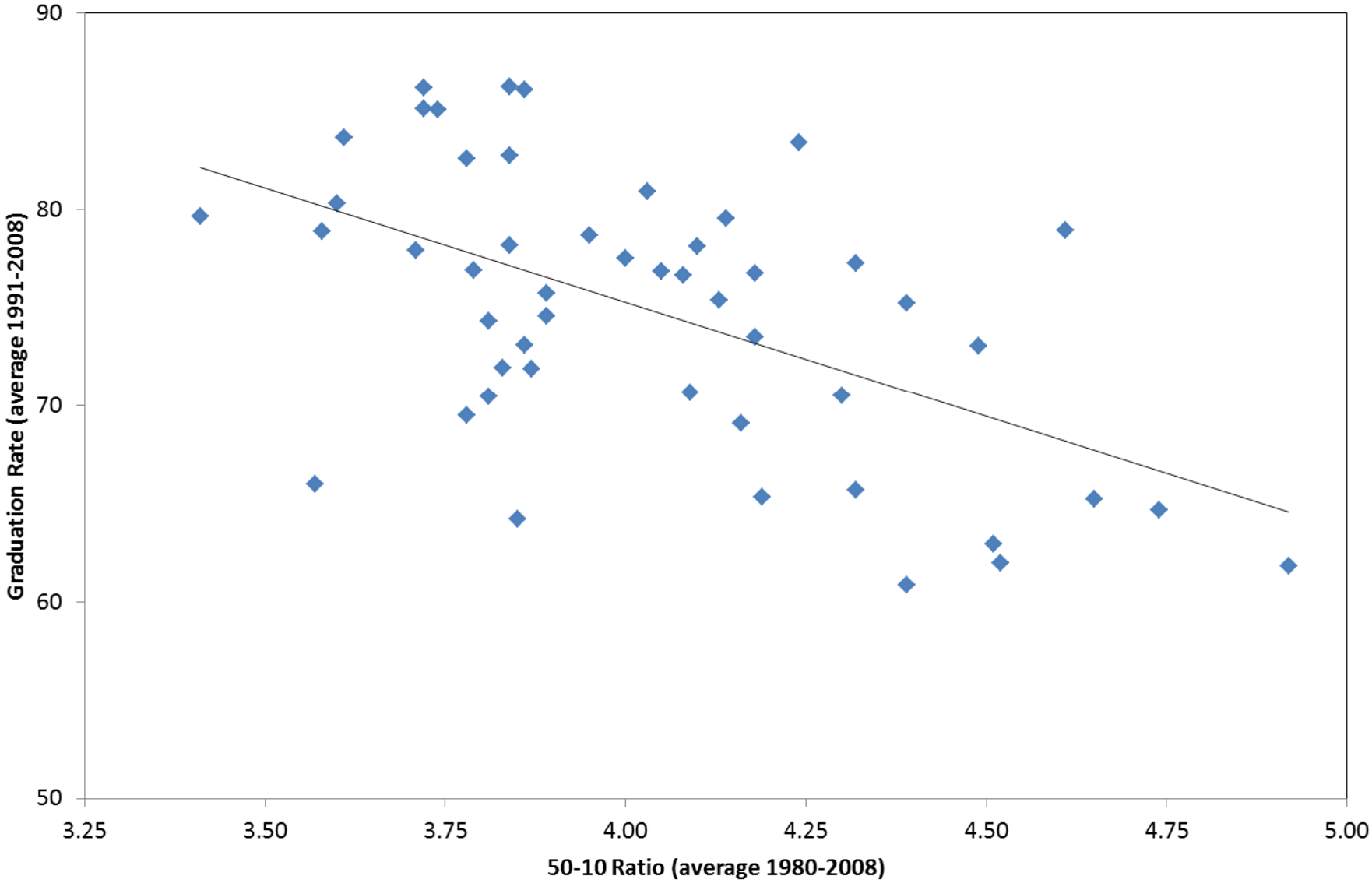


Figure 2: High School Dropout Rate by Mother's Level of Education and State Level of Income Inequality

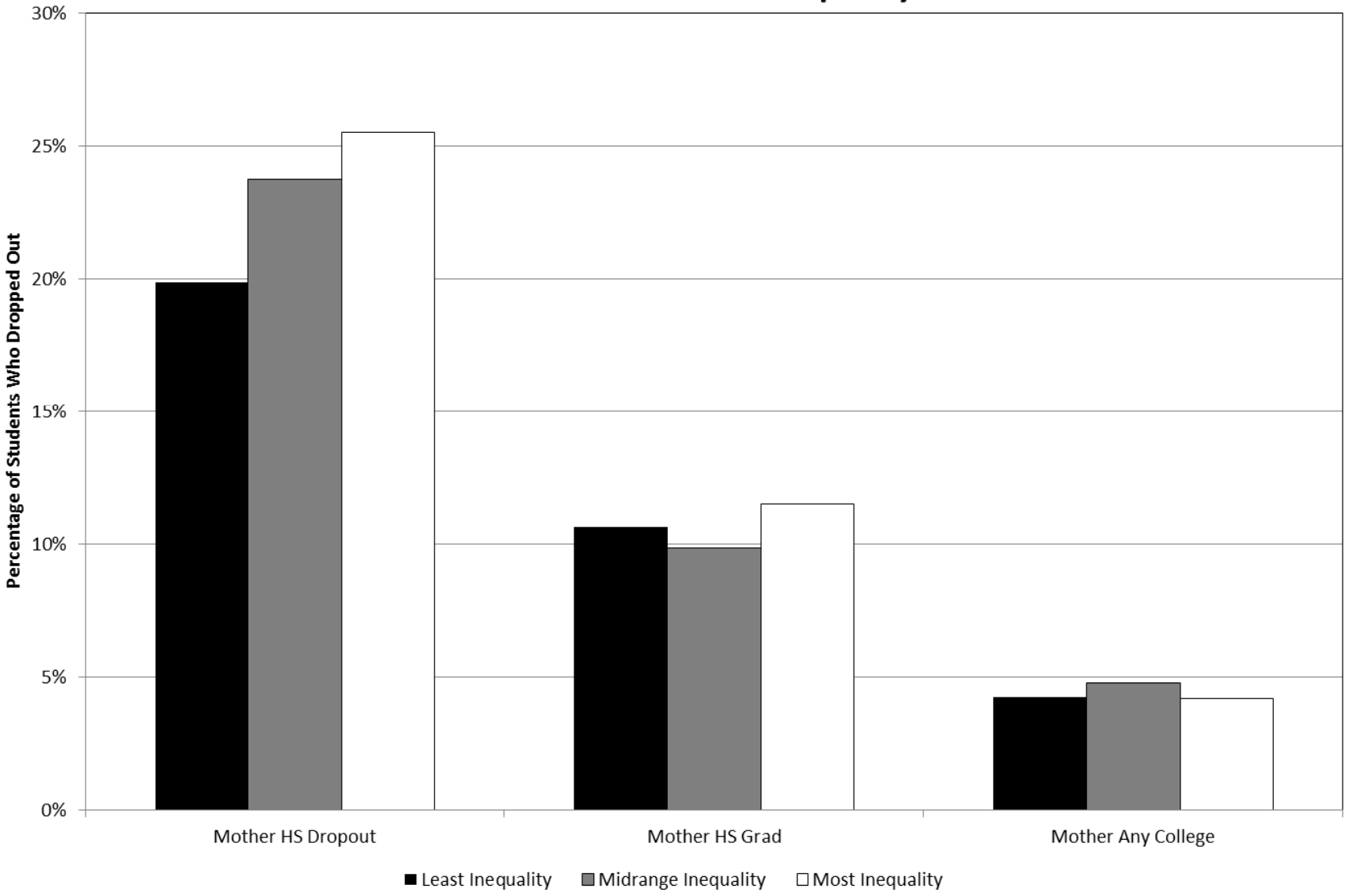


Figure 3: Rate of GED Completion by Mother's Level of Education and State Level of Income Inequality

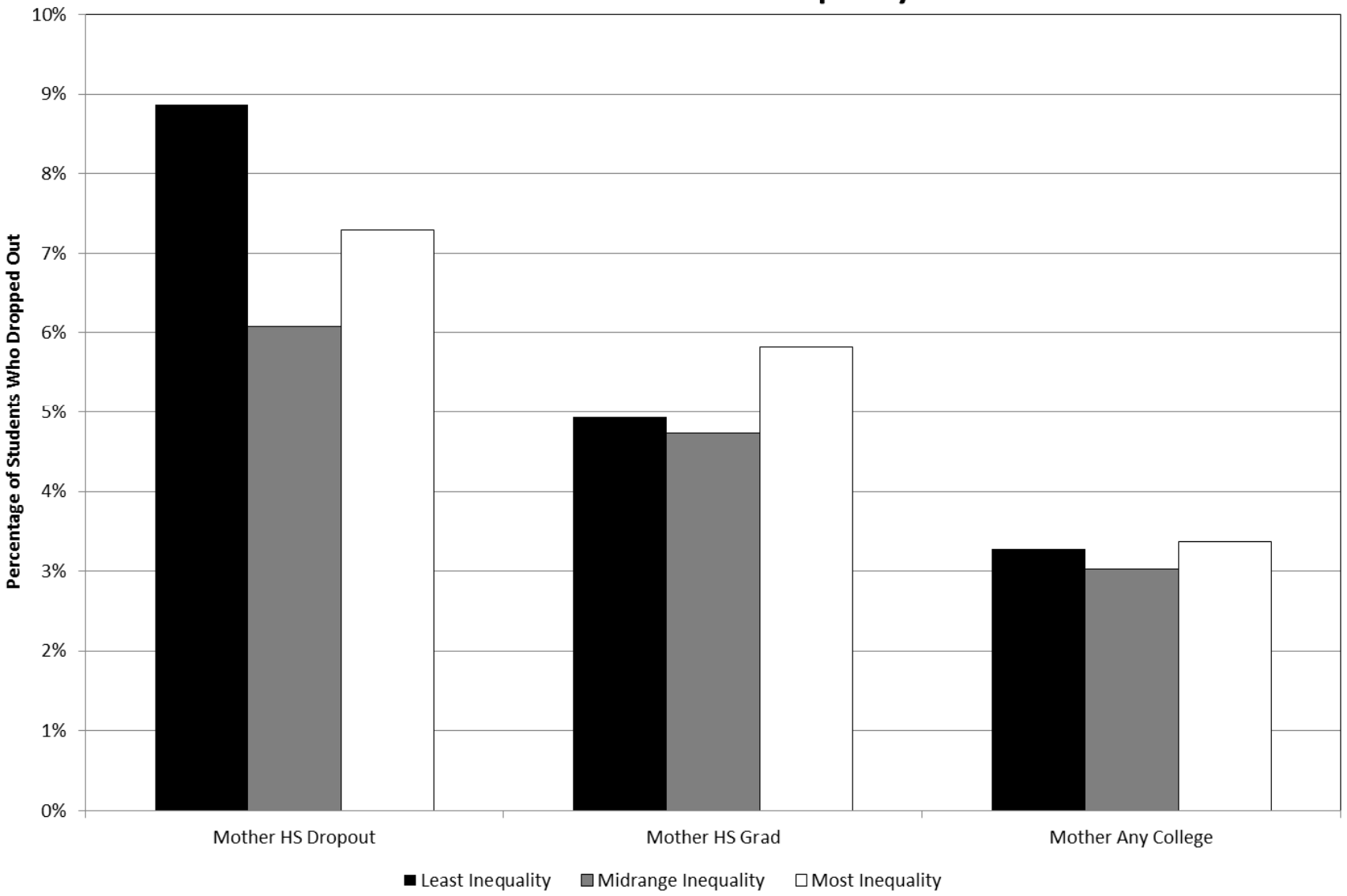
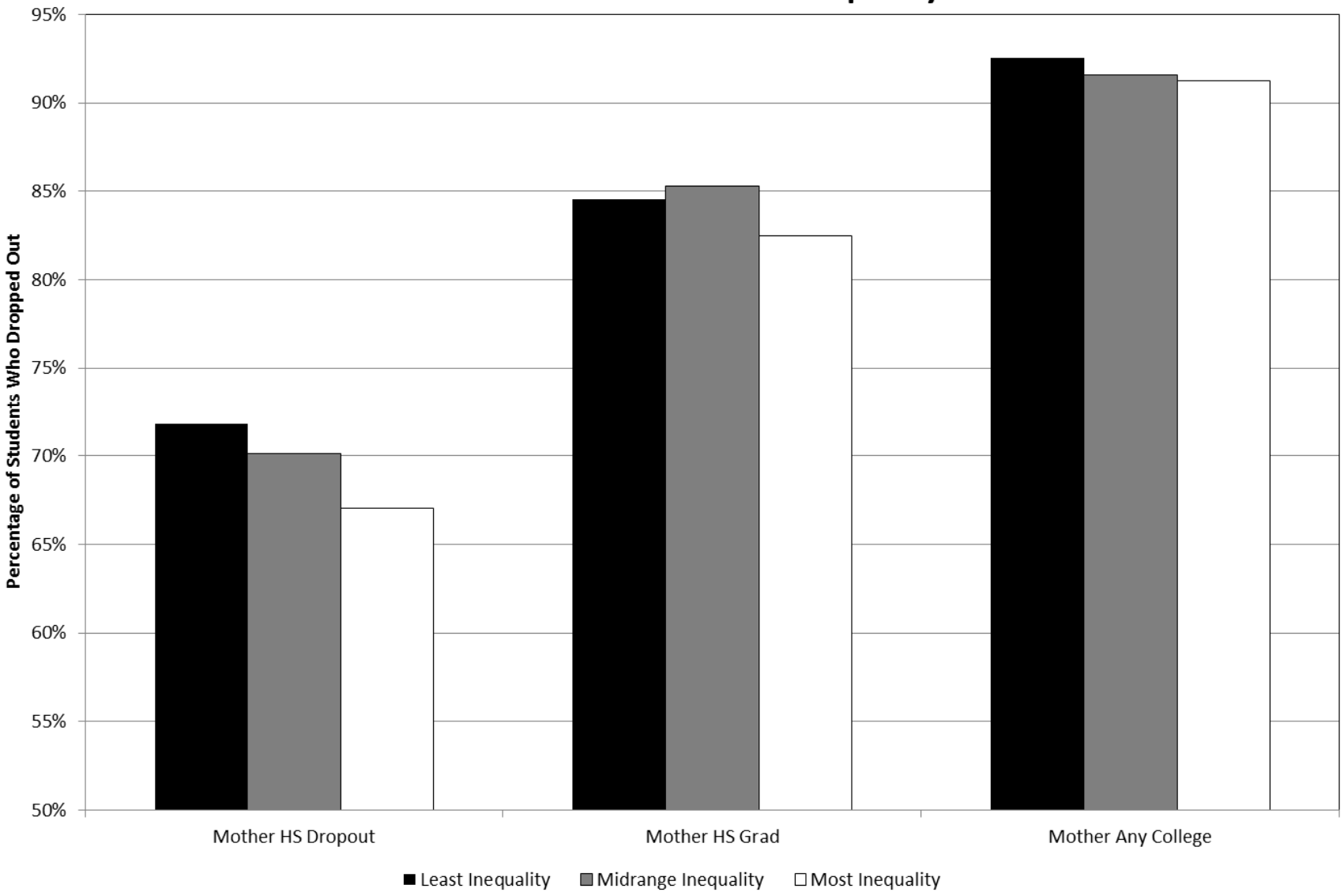


Figure 4: High School Graduation Rate by Mother's Level of Education and State Level of Income Inequality



Data Appendix: Measuring Educational Attainment in NCES, NLSY, and CCD data

National Longitudinal Survey of Youth, 1979 (NLSY79)

This data source originally surveyed 12,686 respondents born between 1957 and 1964, who were between the ages of 14 and 22 on the first survey date in 1979. The sample was not nationally representative, but sample weights are available to provide national representative estimates. Retention rates have been very high in these data, reducing the likelihood of attrition bias, particularly over relatively short periods. Respondents were re-interviewed every year through 1994 and then every other year after that. Because the NLSY is not a school-based survey, the universe of respondents is not restricted to those currently enrolled in a certain grade, as in the NCES data sources. On the other hand, some respondents are older than mandatory schooling ages on the initial survey and are reporting their ultimate educational attainment and the timing of its completion retrospectively, introducing the possibility of recall bias.

Statistics on educational attainment presented in Mishel and Roy (2006) are roughly consistent with our findings. We focus on educational attainment by age 20 and they focus on educational outcomes in 1984 for those respondents who are between 20 and 22 in that year. We find that 17.9 percent dropped out of high school and 5 percent have a GED. Comparable statistics in their analysis are 16.8 percent and 4.3 percent. These differences are minor, and they are a reasonable approximation considering the minor sample restrictions that they placed on their data and the slightly different universe.

Choosing appropriate sample weights using the NLSY data is complicated by the nature of the data on educational attainment. Ideally, if we had complete information in every survey

year for every respondent, we would assign a panel weight for the year the respondent turned age 20 to obtain representative estimates for the 1957 through 1964 birth cohort at that age. Of course, some attrition occurs, and to complicate matters further, some individuals miss some surveys, but then are added back to the sample. Since we are obtaining data retrospectively on educational attainment, those individuals can be included in the analysis. It is not clear what weighting variable to use for those individuals. In the end, we rely on the fact that sample attrition is very low in these data and our analysis only tracks individuals for a small number of years after the survey began, so we chose to use sample weights from 1979. We have experimented with alternative weights, but our estimates are very similar regardless of our decision here.

National Education Longitudinal Survey (NELS)

NELS initially surveyed 8th graders in the spring of 1988, when most of them were 14 years old. They were re-interviewed in 1990, 1992, 1994, and 2000. In total, 14,915 respondents were interviewed initially in 1988 and again in the 1994 round, which represents the point at which we measure educational outcomes. Survey responses regarding educational attainment were recorded in each of these survey years and a subsample of these responses were checked against transcript records indicating their accuracy. The survey excluded 5.4 percent of selected students in the base year "because of physical or mental disabilities, or because of limited English language proficiency" (Ingels and Quinn, 1996). Of those students excluded, 38 percent were not enrolled in school four years later and almost half (42.4 percent) of the remainder had fallen behind grade level at that point (Ingels and Quinn, 1996).The sample was "freshened" in subsequent surveys so that representative estimates could be drawn for the sophomore class in 1990 and the senior class in 1992.

We have chosen to focus on those surveyed in the base year, despite the fact that this decision introduces an upward bias on the educational attainment of this group. The reason we have done so is because using the "refreshed" samples introduces an additional form of upward bias on educational outcomes. Those students who have made it to their sophomore or senior years are a positively selected group of students. For instance, in the NLSY79, 6 percent of respondents never make it to 10^{th} grade by age 20. Indeed, even starting a sample in 8^{th} grade introduces an upward bias in educational attainment since 1 percent of respondents reported that by the age of 20, they never reached the 8^{th} grade.

For these reasons, it is not surprising that educational outcomes in the NELS are superior to those observed in the NLSY79. In the NELS, we estimate that 9.8 percent of students drop out and 5.3 percent of students obtain a GED by age 20 compared to 17.9 and 5.0 percent, respectively, in the NLSY79 data. These differences are partly attributable to differences in sample design and partly attributable to actual changes. Our estimates are comparable to those in Hurst, et al. (2004), who find that 12 percent of students drop out and 6 percent of students obtained a GED by 1994 (when most respondents are age 20). Minor coding differences can account for this small discrepancy.

Sample weights in the NELS data represent the inverse of the probability of being included in the sample. The weighting variable that we use in this analysis is labeled F3PNLWT. It represents the panel weight that applies to sample members who were originally surveyed in 1988 and then continued in the survey through 1994. Since these respondents will have complete information on the educational outcomes by age 20, we viewed this as the appropriate weight to use.

National Longitudinal Survey of Youth, 1997 (NLSY97)

These data include information on 8,984 respondents who were born between 1980 and 1984, making them 12 to 18 on the first survey date. The sample was not nationally representative, but weights are available to provide nationally representative estimates. Retention rates have been very high in these data, reducing the likelihood of attrition bias, particularly over relatively short periods. Respondents have been re-interviewed every other year since 1997 with the most recent available survey having been completed in 2009. Relative to the NLSY79, these data have the advantage that virtually all students are still in school at the time of the initial survey, so we can more reliably track their exit as they age. Aside from differences in the possibility of recall bias, these data are the most comparable to those from the NLSY79. Indeed, we adopt the same decision-making process in choosing sample weights as those that come from the 1997 survey, just as we chose to use 1979 sample weights in the earlier NLSY cohort.

Using these data, we find that 13.7 percent of respondents dropped out of high school and 7.4 percent obtain a GED by age 20. These statistics compare favorably to those in Mishel and Roy (2006), who find that 12.8 percent of respondents dropped out of high school and 5 percent earned a GED for those respondents who were between the ages of 20 and 22 in 2002. These estimates also reflect a change in educational attainment compared to the NLSY79. Fewer students dropped out of high school than in the 1979 cohort, but more obtained their GED. The number of students who obtained a high school diploma rose from 77.1 to 78.9 percent, representing a modest change. Since the format of these sources of data are the most similar, this comparison is a useful one and the observed differences suggest that traditional high school graduation was roughly constant over the period, although those who did not complete high school appear to have been more likely to obtain a GED.

Education Longitudinal Survey of 2002 (ELS)

This survey included students who were in 10th grade in the spring of 2002. Students were re-surveyed in 2004 and 2006, so that they are around 20 years old in the latest year of available data. There were 15,300 students who responded to both the base year survey and the 2006 survey, when educational outcomes were measured. We can use these data to examine ultimate educational outcomes for these students, but this is something of a selected sample. Many students who drop out of high school never make it to 10th grade in the first place, so they are not included in these data. As indicated earlier in our discussion of the NELS data, 6 percent of respondents in the NLSY79 never made it to 10th grade. In the NLSY97 data, which is a cohort closer to the ELS, 8 percent of respondents never made it to 10th grade. About one-third of those received a GED and virtually all of the others dropped out of high school. The distribution of educational attainment that we report in Appendix Table 1 from ELS data is comparable to that reported in Aud, et al. (2011) from these data.

Sample weights in the ELS data represent the inverse of the probability of being included in the sample. The weighting variable that we use in this analysis is labeled F2BYWT. It represents the panel weight that applies to sample members who were originally surveyed in their sophomore year of high school in 2002 and then were resurveyed in 2006. Since these respondents will have complete information on the educational outcomes by age 20, we viewed this as the appropriate weight to use.

Other Sources of NCES Microdata

Two other sources of microdata are available from NCES that could conceivably be used for this exercise. One survey, the National Longitudinal Survey of the High School Class of 1972, focuses on high school seniors in the 1972 base survey year. Since many high school dropouts

and GED recipients never make it to their senior year (see Appendix Table 2), these data are inappropriate for our analysis. A second survey, High School and Beyond, is potentially more useful. Commencing in 1980, these data tracked students who were either high school sophomores or seniors in that year. Using seniors would be inappropriate for our analysis, as we just described. Sophomores, however, would be a viable option. Although that source of data would be subject to the same limitations as the ELS, described earlier, these data could still provide a valuable addition to our analysis. The binding constraint that prevents us from using these data at all is that we are not able to obtain state-level geographic identifiers for these data.

Appendix Table 1: Educational Attainment Measured in Alternative Longitudinal Data Sources.

	Educational Attainment by Age 20		
	GED	High School Dropout	High School Graduate
NLSY79	5.0	17.9	77.1
NELS (1988)	5.3	9.8	85.0
NLSY97	7.4	13.7	78.9
ELS (2002)	4.2	7.6	88.2

Source: Authors' calculations.

Appendix Table 2: Degree Status by Highest Grade Completed at Age 20

	Below 8th Grade	8th Grade	9th Grade	10th Grade	11th Grade	12th Grade and Higher
			NLSY79			
Percent at Level	1.2	1.8	3.3	4.5	9.8	83.9
Degree Status:						
HS Dropout	99.2	97.6	93.8	95.8	89.8	3.4
GED	0.2	1.3	4.2	4.1	4.3	5.2
HS Graduate	0.6	1.1	2.0	0.1	5.9	91.4
			NLSY97			
Percent at Level	0.6	3.0	4.5	5.5	6.4	80.0
Degree Status:						
HS Dropout	89.7	75.4	63.3	58.8	61.6	1.1
GED	7.5	22.0	33.3	37.6	31.9	1.4
HS Graduate	2.8	2.6	3.5	3.6	6.5	97.5

Sources:

Aud, Susan, Angela Kewal Ramani, and Lauren Frohlich (2011). *America's Youth: Transitions to Adulthood* (NCES 2012-026). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Mishel, Lawrence and Joydeep Roy (2006). *Rethinking High School Graduation Rates and Trends.* Washington, DC: Economic Policy Institute.

Ingels, Steven J., and Peggy Quinn (1996). *Sample Exclusion in NELS:88: Characteristics of Base Year Ineligible Students; Changes in Eligibility Status after Four Years [National Education Longitudinal Study of 1988 Second Follow-Up].* Washington, DC: U.S. Department of Education. National Center for Education Statistics.

Hurst, David, Dana Kelly, and Daniel Princiotta (2004). *Educational Attainment of High School Dropouts 8 Years Later*. Washington, DC: U.S. Department of Education, National Center for Educational Statistics.