

**Ability Tracking and Class Mobility in High School Mathematics:
The Case of Low Achievers**

Bradley T. Shapiro

Thesis submitted to the faculty of the Virginia Polytechnic
Institute & State University in partial fulfillment of the
requirements for the degree of

Master of Science
In
Mathematics

William Floyd, Chair
T. Nicolaus Tideman
Anderson Norton

May 1, 2009
Blacksburg, VA

Keywords: Ability Tracking, Secondary School Mathematics,
Economics of Education

**Ability Tracking and Class Mobility in High School Mathematics:
The Case of Low Achievers**

Bradley T. Shapiro

Abstract

The goal of this paper is to evaluate commonly held criticisms of the practice of ability tracking in high school mathematics. To do so, I employ data from the *National Education Longitudinal Study of 1988* and follow-ups to model classroom selection and education production. This paper will focus only on the causes and effects of tracking on students who were tracked as low-ability in eighth grade. From this, we can see how many students, if any, switched out of the low-ability track by tenth grade and how various switches have affected their test scores in mathematics. I find that students exercise mobility between ability-tracks as late as tenth grade and that ability-track placement is largely determined by test scores. In addition, I find evidence that there would be minimal, if any, test score improvement among low-ability students if they were all moved to a class of heterogeneous ability.

Acknowledgements

I would first like to thank my advisor in the economics department, Nicolaus Tideman, for his patience and guidance in helping me to develop my own research. His passion for research has inspired me to further pursue economic research at the doctoral level. I'd like to thank William Floyd not only for his wonderful instruction of Real Analysis, but for his encouragement of whatever research path was the most interesting to me. I would also like to thank Anderson Norton for his time in serving on my committee. Finally, I must thank Elaine, my parents, Mary and Jeffrey, and my sister, Allison, for their love and support.

Contents

i.	Title Page	
ii.	Abstract	
iii.	Acknowledgements	
iv.	Contents	
v.	List of Figures	
1.	Introduction	(1)
1.1	Why Do We Care?	(3)
2.	Developing a Model	(4)
2.1	The Tracking Debate - History and Literature	(7)
2.2	Question 1: How is Track Placement Determined at the High School Level?	(13)
2.3	Question 2: What Effect Does Tracking Have On Test Score Achievement?	(14)
3.	The <i>National Education Longitudinal Study of 1988</i>	(15)
3.1	Selection of the Control Variables: The Multinomial Logit Model	(18)
3.2	Selection of the Control Variables: The Education Production Function	(21)
4.	Execution of the Models	(22)
4.1	Track Mobility and the Multinomial Logit	(22)
4.2	Assumptions of the Multinomial Logit	(23)
4.3	Mechanics of the Multinomial Logit	(24)
4.4	Results of the Multinomial Logit	(26)
4.5	Test Score Effects and the Education Production Function	(30)
4.6	Assumptions of Least Squares Estimation	(31)
4.7	Properties of Estimators and the Gauss-Markov Theorem	(32)
4.8	Heteroskedasticity	(33)
4.9	Results of Approach I - Robust Standard Errors	(41)
4.10	Results of Approach II - Weighted Least Squares	(45)
5.	Where Else Could We Go From Here?	(47)
5.1	A Look Into Matching	(47)
5.2	Proposed Model Idea	(49)
5.3	Addressing the Model Questions	(50)
5.4	Broader Impacts of the Proposal	(51)
6.	Conclusions	(51)
7.	References	(55)

List of Figures

1. Figure 1: Improvement vs. Initial Test Score	(6)
2. Table 1: Eighth Grade Test Scores By Tenth Grade Track	(20)
3. Table 2: Race By Tenth Grade Track	(20)
4. Table 3: What are the Determinants of Track Mobility	(27)
5. Scatter Plot 1: Residuals vs. Base Year Scores	(34)
6. Scatter Plot 2: Residuals Squared vs. Base Year Scores	(35)
7. Example Plot 1: Un-weighted Least Squares Line	(39)
8. Example Plot 2: Weighted Least Squares Line	(40)
9. Table 4: Parameter Estimates with Robust Standard Errors	(42)
10. Table 5: Estimation of Residuals Squared	(45)
11. Table 6: Weighted Least Squares Estimation	(46)

1. Introduction

The practice of ability tracking in education systems is not a new phenomenon. Many school systems throughout the world use programs that place students into classes according to ability to appropriately focus the material to the specific group. In theory, if all of the bright students were in a class together, more material could be covered. Conversely, in a classroom of heterogeneous ability, it is possible that the teacher would be forced to slow down the presentation of material to allow the slower students to keep up.

There are many different kinds of ability tracking at different levels of education. Some schools use intelligence tests in the early years of schooling to determine which track a student will follow. Other schools use a combination of performance in the previous year, teacher recommendations and beginning-of-year tests to determine the placement of students into classrooms. As students advance in age, it is possible that schools allow students to choose their own classes.

In recent years, the usefulness of ability tracking in primary and secondary schools in the United States has come into question (Oakes 1992, Slavin 1990). Critics of the practice cite a plethora of problems including unfair track placements, a lack of achievement gains among high achievers attributed to

tracking, a lack of mobility between tracks, and unequal outcomes for lower and higher track students. From these criticisms, many conclude that the only logical course of action is to abandon the tracking system in favor of exclusively heterogeneous classes.

It is instructive to ask how students are placed and to inquire into the fairness of that process. Are students stuck in these tracks for their entire educational careers, unable to change? It seems plausible that after several years in a particular track that a student may no longer be able to make a jump into a more advanced setting. Are non-academic factors such as race, economic background, or extracurricular activities affecting placements? Even if such considerations are not affecting placements, are they strong predictors of track? Available data could shed some light on all of these questions.

Additionally, while track placement is indeed an interesting process, the results of those placements are of primary concern. Are the students in the low ability tracks showing smaller test score improvements than their colleagues in the high ability tracks? If so, is tracking the cause of the lower scores, or is there simply something else correlated both with test score improvements and track placement such as intelligence, motivation or attention span? These are questions that, with cleverly designed models, we may be able to learn

more about. As I will discuss in the pages to come, separating innate differences in ability from the effect of track placement is rather difficult, as intelligence, motivation and attention span are to a large extent unmeasurable, are potentially dynamic and surely correlate with achievement.

1.1 Why Do We Care?

We care about these questions for a few reasons. First, if students are unable to move between tracks, students who are simply a little slow to develop might completely miss out on the opportunity to be exposed to advanced material later in their educational careers. This could lead to decreases in co-curricular and extra-curricular opportunities, eventual educational attainment and job market earnings. Next, if students are being placed in *ability* tracks on the basis of race or socio-economic status, there would be obvious reasons for concern.

Another reason that we care about these questions is that there is indeed a movement to de-track America's schools. Such a reform would be costly in time and money and would be difficult to reverse. Before such comprehensive reforms are put into place, it would be wise to consider the validity of the claims against tracking.

In publicly providing education, we want students at all levels of ability to be achieving to their fullest potential,

and the system in which they are placed may play a large role in their eventual success in life. If simple reforms can make a large difference in the quality of education for traditionally ‘left behind’ students, then it is worthwhile to explore their merits and costs.

As most of the criticisms of tracking express concerns about the students tracked to low-ability classes, this paper will focus only on the causes and effects of tracking on students who were tracked as low-ability in an initial time period: eighth grade in our particular data. From this, we can see how many students, if any, switched out of the low-ability track by tenth grade and how various switches have affected their test scores in mathematics.

2. Developing a Model

Developing an empirical model can be a daunting task. Often, insight from economic theory can help to specify such models. In addition to theory, the literature also aids the process of specifying a model. In the topic at hand, we are faced with two distinct questions:

1. How is track placement determined at the high school level?
2. How does track placement affect test score improvements?

As it turns out, the appropriate tools for these two questions are different, but they are related in several ways. I address the first question using a discrete choice model to estimate the

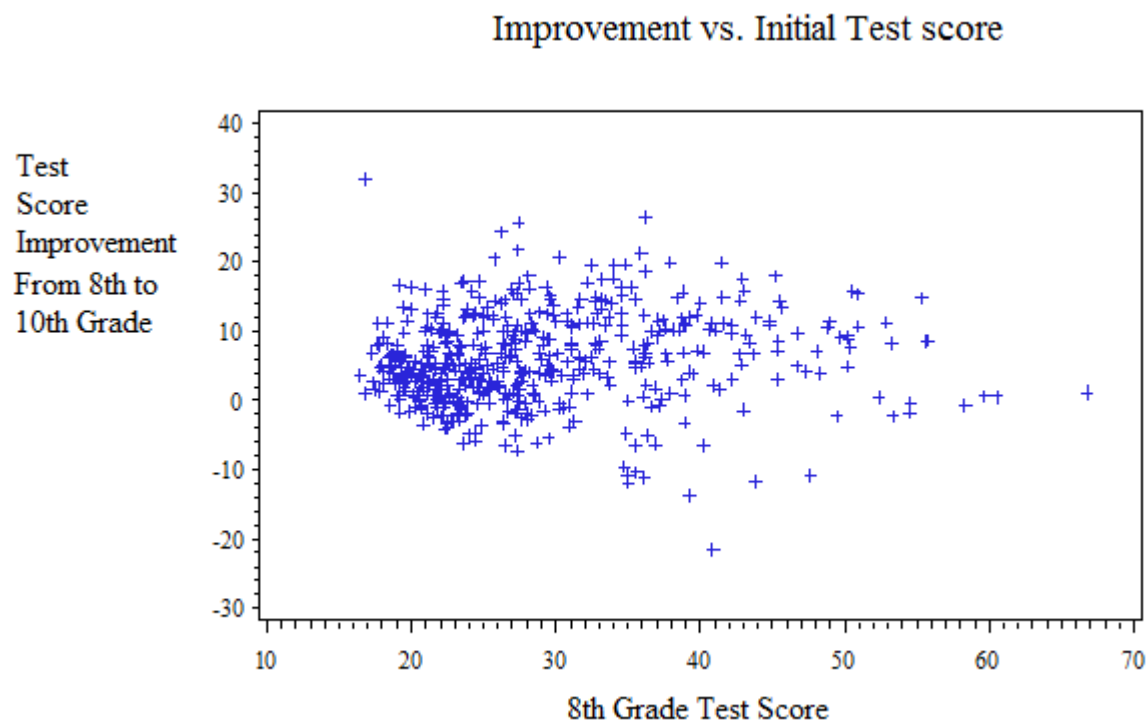
probability of 10th grade placement in a particular track based on a variety of factors. I address the second question using an education production function approach.

Unfortunately, there is very little guidance from economic theory with regard to designing education production functions (Hanushek 1986). There have been attempts that are described below to accurately weigh the effects of a variety of factors on test scores. Intuitively, it seems obvious that factors such as class size and teacher experience should be included, among many others. What proves most difficult is deciding the functional form of the relationship between test score improvement and past test scores. Past literature suggests that there is something more complicated than a linear relationship, but there is relatively little detail pertaining to the particular form.

Below, mathematics test score improvement from eighth to tenth grade is plotted against eighth grade mathematics test scores. We might expect that those close to the minimum score in eighth grade have nowhere to go but up and those at the top of the distribution have the opposite effect, but what happens in the middle? Also, base year test score may serve as a proxy for intelligence, so we might expect marginally higher base year scores to be associated with marginally better improvements. My specification came after many trial and error attempts to

ascertain the relationship between the two variables in addition to the contributions from the literature discussed below.

Figure 1



One major criticism of the production function approach is that most such studies use data that are aggregated over a district, state or even the country, while one of the most defining characteristics of the student body in the United States is its heterogeneity. In this paper, I use student level data from the National Education Longitudinal Study of 1988 to estimate the effects of changing from a tracked to an untracked math class between eighth and tenth grade on students initially

in classes that the teachers identified as having “lower levels of performance.”

This tracking study, like many others, analyzes high school achievement. Tracking at the high school level must be seen in a different light than tracking at the primary school level. Separation by ability in high school may be largely a matter of student choice. A teacher-reported “high ability” course may be an Advanced Placement course or an International Baccalaureate option. While classes of this nature may require certain prerequisites, they may be chosen primarily by students and parents. Any effort to detrack a high school may require the removal of these elective courses in favor of courses accessible to all students in the ability distribution, or the placing of all students in such accelerated classes.

In addition, it must be noted that if students are able to exercise mobility at the tenth grade level, being tracked in a low-ability class earlier must not have precluded them from partaking in more difficult classes later on in their educational careers.

2.1 The Tracking Debate - History and Literature

Oakes (1992) provides harsh criticisms of ability-tracking practices. Her most critical claims are that tracking placements are both permanent and unfair, that tracking hurts low-ability students while not helping high-ability students,

and that outcomes are not equal for the different tracks due to limited availability of opportunities. Ability tracking comes in many different forms, making it infeasible to address all forms in a single study. Therefore, tracking will be addressed in this paper by asking, "Is there a case for planned heterogeneity in ability levels in mathematics classrooms in particular?" In addition, this paper will only address tracking in mathematics at the high school level. It is possible that students respond differently to tracking at different ages and in different subjects; those possibilities are not addressed here. However, addressing high school tracking may be particularly informative because it could provide insight into the question of whether or not students are locked into tracks at an early age.

Some conclusive studies on classroom composition examine tracking in its most basic form, special education. Hanushek, Rivkin & Kain (1998) conclude that special education programs increase achievement of both regular students and special education students. Of course, these results must be accepted cautiously since not all special education students take the same tests as regular students. However, both groups improved when they were separated.

Argys, Rees, & Brewer (1996) conclude that Oakes' claim that detracking does not hurt high-ability students is false.

They come to this conclusion by controlling for both track-assignment and classroom characteristics. They estimate that the average student, if moved from a low-ability class to a high-ability class would increase her mathematics test score by 8.6 percent. In addition, they estimate that a move in the opposite direction would lead to a decrease in score by 8.4 percent. They estimate that complete detracking would cause a 2.0 percent decline in overall average test scores. These results imply that there are clear winners and losers in detracking. Such results are potentially susceptible to misspecification bias, and the present paper reexamines this question. According to Argys, Rees, & Brewer (1996), if a school system's goals include avoiding changes that make some students worse off, detracking is not the answer. If the school system wants a solution that improves the performance of the student who is performing at the lowest level without regards to impacts on other groups, detracking ought to help. However, the work reported in the present paper suggests that they may not have captured the whole picture.

Figlio & Page (2000) consider the possibility of Tiebout voting - the idea that families vote with their feet. They choose to live in districts that will provide satisfactory educations for their children. Since tracking seems to help high-ability students, a large-scale detracking effort by a

particular school would cause a movement of high-ability students away from that school. In other words, with mobility of the student body taken into account, the benefits obtained from the presence of the high-ability students might be lost. Indeed, Figlio & Page find that both gifted *and* remedial programs attract high-income and medium-income families, making the student body more economically diverse. Summers & Wolfe (1975) find that low-ability students benefit greatly from having high-ability students in the school. This may be because the parents of high-ability students contribute to the classroom atmosphere (Hoover-Dempsey, 1987), because the peer effect of high-achieving students with them in untracked classes such as physical education is motivation for low-ability students or because the correlation between high ability and high income leads to more resources being spent on the low-ability students. Thus it is likely to be in the interest of low-ability students to keep high-ability students around. Therefore, Figlio & Page (2000) conclude that tracking actually helps low-ability students because without it the high-ability students would likely leave.

Hallinan (1994, 1996) explains that many objective and subjective criteria are taken into consideration when determining track placement. Such considerations as scheduling, teacher assessments of potential, previous track and test scores

are parts of the track determination process. In addition, track mobility provides an opportunity to correct for inappropriate initial track placements. Hallinan explains that while practice deviates from the theory of tracking, many of these can be fixed at a relatively low cost compared to the cost of eliminating tracking altogether. One of the major problems she describes is a lack of homogeneity in classes. Without homogeneity in classes, the potential for directing instruction at a specific ability level is diminished and all that remains of each track is a potentially harmful label. In order to maintain homogeneity, track mobility is crucial. This paper finds indeed that students who outperform their classroom peers are able to move to higher tracks.

Zimmer (2003) tackles the question of homogeneity in tracked classes. He models education production taking into account both track placement and class average test score. His findings are consistent with Figlio's claim that tracking is advantageous to all groups if the classes are truly homogeneous in ability. That is, high ability students perform better in higher tracks as the average test score increases. In addition, students in homogeneous low ability tracks outperform their untracked counterparts when the average class achievement is the same. This suggests that there may be some de-facto equilibrium tracking even in 'untracked' settings.

Argys, Rees & Brewer (1996) look deeper into the issue of the determinants of track placements. They use a sample of students of all abilities, and estimate the probability of being placed into various tracks based on different factors, using a multinomial logit model. They find that the largest determinants of track placement are previous track placement and test scores in the previous period. This study estimates a similar model, but differs in the sample used and in the interpretation of the results. As I use only data from students initially placed in low-ability tracks in the sample, I can isolate the probability of switching tracks as opposed to the probability of being in a particular track. In other words, I am attempting to compare apples with apples.

Finally, Oakes decries the fact that economic outcomes for high-ability and low-ability students are different. Figlio and Page (2000) state that “those studies [that suggest unequal outcomes] have not adequately addressed the possibility that track placement and tracking programs may be endogenous with respect to student outcomes.” In other words, we should expect that some students will have a higher potential for learning and therefore learn more, and it should not be surprising if those who have learned more have higher wages. These students could possess unmeasurable characteristics such as motivation, attention-span and raw intelligence. While some of these

unmeasurables could be dynamic and affected by tracking, surely some of them are natural. Still, Oakes does present a reasonable concern when she says that many opportunities (e.g. high quality teachers) afforded to high-track students may not be afforded to low-track students. It is possible that this problem could be fixed at a relatively low cost at the school level, with a change in administrative attitude.

Indeed, many high schools allow students to select their courses. At the elementary school level, however, such choice is much less common, and some level of parental participation in placements might be preferable to exclusively administrative placements. In addition, the teaching pedagogy may have room for improvement in the lower tracks. Such improvements could be made by different teacher assignment practices, merit incentives, or some other method. The possibilities for improved content and pedagogy in tracks are not considered in this paper, but they are certainly worthwhile avenues for future research.

2.2 Question 1: How is Track Placement Determined at the High School Level?

Using the available data, I start with the group of students who were, according to their teachers, placed in a class of lower ability for eighth grade in mathematics. From there, I would like to estimate the probability of one student

switching into a class of high, average or varying ability by tenth grade. Questions of this type are not new, and I follow the example from Argys, Rees and Brewer (1996), who determine the probability of being tracked high, average, varying or low in eighth grade, that incorporates a variety of factors using a multinomial logit model. However, in contrast to the Argys model, I focus only on students initially in low-ability classes. This way, the predicted probabilities may be interpreted as the probability of switching from a low track to some other track. I find that higher test scores are associated with higher probabilities of switching to each of high, mixed and average tracks. In addition, I find that a variety of other factors are also associated with track changes. However, I do not find that race and socioeconomic status are significantly related to track changes. I do find that a substantial proportion of students change tracks. Thus the data from the *National Education Longitudinal Study of 1988* refute claims that students get “locked in” to certain tracks without the opportunity to change.

2.3 Question 2: What Effect Does Tracking Have on Test Score Achievement?

I would like to determine whether a switch to heterogeneous classes would have a positive effect on the test scores of low achieving students. To do this, I estimate an education

production function for the entire sample of initially low achieving students in eighth grade, using the tracks to which students are assigned in tenth grade as regressors. The production function is a standard regression function, estimated by least squares. After using several methods to allow for unequal variance of the error terms (heteroskedasticity), I find that there is evidence that the effects of switching to a heterogeneous class are approximately zero.

3. The National Education Longitudinal Study of 1988 Data

The National Education Longitudinal Study of 1988 (NELS) is a nationally representative survey that follows students from eighth grade into the workforce. For each student, surveys are completed during eighth grade, tenth grade and twelfth grade by the student, two of the student's teachers, the student's parent and a school administrator. No other available study has included so much detail from so many sources at the student level as the *NELS*.

Each student surveyed took achievement tests at the end of the eighth grade and tenth grade in reading, mathematics, science and history. The scores represent an estimated number of items the student answered correctly. Only the math and reading test results are used in this paper.

Argys, Rees and Brewer find that the main determinant of tenth grade track is eighth grade track. However, in this data set “track” is a characteristic of each class that is reported by the teacher. A high ability track, as reported by the teacher, might be an Advanced Placement course or an International Baccalaureate course. Since most high school students are able to choose the classes they take, “track” may be more a matter of student choice than an administrative decision. While tracks of a student’s earlier years may determine the student’s preparation for the high school courses, only the effects of the high school track itself, which may be chosen by the student, are examined in this paper and in other papers that use the *NELS* data set. More specifically, this paper looks into the results for those students who switched from a low ability track to other tracks or to an untracked setting, to determine if track mobility is possible and/or helpful. Changes between eighth grade and tenth grade are considered because most students in the United States begin high school in ninth grade. This presents the possibility that a student has re-evaluated his or her future in deciding which courses to take or even which school to attend.

For each student, two teachers were interviewed to provide classroom level characteristics. The sample in this paper is restricted to those students who had a mathematics teacher

interviewed. The teachers answered whether the class was “lower levels of achievement”, “average levels of achievement”, “higher levels of achievement” or “varying levels of achievement.”

While such data may not be optimal, as teachers may have subjective views as to how classes should be classified, these data are likely to be more reliable than student self-reported track placement. Students who had teacher reports of “lower levels of achievement” in the classroom at the eighth grade and different teacher reports of classroom ability at the tenth grade were categorized as switching from Low to Average, Low to High or Low to Untracked. Of the 3959 students that had both eighth grade and tenth grade mathematics teachers respond to the question of track placement, 561 students were categorized in eighth grade as being in a math class with “lower levels of achievement.” Of those 561 students, 321 of them had tenth grade teachers reporting that the student was in a track other than “lower levels of achievement.” 191 students were reported to have switched from a “lower levels of achievement” class to an “average levels of achievement” class. 51 were reported to have switched to a “higher levels of achievement” class, and 79 switched to an untracked setting. The number of students switching is statistically bigger than zero at any conventional level, contradicting Oakes’ claim that students are not able to move among tracks.

The survey data provide a large variety of information, giving many possible control variables at the student, classroom and school level for both the multinomial logit model and education production function. For example, teachers were surveyed on the ability level of their classes, class sizes and their own experience levels. Administrators were surveyed on school characteristics such as the number of students in the school taking remedial math courses. Information was also gathered on student socioeconomic status, gender and race. Those characteristics most likely to be associated with test score achievement are included in an education production function below.

3.1 Selection of the Control Variables: The Multinomial Logit Model

The main variable of interest in the track placement model is eighth grade test score. This is because if the tracking system is functioning flexibly and efficiently, students who are misplaced in a lower track and subsequently perform well should be able to switch tracks. Also, if initial test scores are a signal of ability and motivation, it is possible that this could bias the results of the education production function. For example, if someone with higher ability switches to a higher track, we might expect that student to show greater improvement.

As Hallinan (1994) points out, non-academic and motivational factors may play a role in determining tracks. Scheduling issues with extra-curricular activities may also play a role. I have included variables that may signal student motivation, such as the student's estimation of how far she will go in school and the teacher's estimation of how frequently the student completes homework. In addition, I have included a dummy for whether or not the school has band available as a proxy for extra-curricular activities present in the school. Base year class sizes are included as a control, and race and socioeconomic status are included to test the claim that tracks are chosen on the basis of discriminatory behavior.

Table 1 below presents the mean eighth grade test score by tenth grade track for the 561 students who were in low tracks in the eighth grade. Interestingly, students who switched out of the low track have eighth grade math scores significantly higher than their colleagues who remained in the low track. As might be expected, those switching to a high track have the highest eighth grade scores. Those switching to heterogeneous classes and to average ability classes have mean eighth grade scores that are very close. Given those numbers, we might expect that students who switch to other tracks are the brighter students and will show the greatest improvements.

Table 1 Eighth Grade Test Scores By Tenth Grade Track					
10th Grade Track Placement	N	Mean	Std Dev	Minimum	Maximum
Higher Levels	51	37.71	12.17	17.43	66.81
Average Levels	191	30.15	8.566	17.84	55.34
Lower Levels	240	25.58	6.772	16.38	54.40
Varying Levels	79	30.32	9.178	17.66	60.52

Table 2 shows the number and the percentage of students of each race in tenth grade track. Nothing stands out as particularly troubling about these numbers, but they do suggest that we must take any race inferences from our model with a grain of salt, as the sample sizes are quite small. That is, we should probably not try to draw too much inference from the fact that two out of eight Native American students switched to a high track.

Table 2 Race By 10th Grade Placement						
10th Grade Track Placement	White	African American	Hispanic	Asian American	Native American	Race Unknown
Total	359	85	76	28	8	5
Higher Levels	33 (9.2%)	5 (5.9%)	8 (10.5%)	3 (10.7%)	2 (25%)	0
Average Levels	122 (34.0%)	31 (36.5%)	25 (32.9%)	8 (28.6%)	3 (37.5%)	2
Lower Levels	146 (40.7%)	42 (49.4%)	35 (46.1%)	13 (46.4%)	1 (12.5%)	3
Varying Levels	58 (16.1%)	7 (8.2%)	8 (10.5%)	4 (14.3%)	2 (25%)	0

3.2 Selection of the Control Variables: The Education Production Function

As the literature has pointed out, there are many determinants to test score improvement. Following the literature as well as intuition, a rich set of control variables that may be correlated with improvement in test scores have been included. While there have been ongoing debates about the effects of class size achievement (Krueger 2000, Hanushek 1999), it seems natural that class size could have some effect on test scores, so it is included. It is worth noting that a remedial class is not the same as a small class. Indeed, Boozer and Rouse (2001) find that class size effects become more pronounced when remedial education is taken into account. Since we are controlling for track placement in this paper, the class size coefficient should not absorb the effects of the class ability level. Following the suggestion of Figlio (2000), a variable is included for the percentage of students in the school who are in remedial mathematics. If his hypothesis is true, the coefficient attached to this variable should be significant and negative. Controls for race, socioeconomic status, gender and teacher experience have also been included in the model.

While specifying the functional form of the previous test-score effect is quite difficult, after many attempts, a

reasonable specification emerged. There was evidence of a non-linear relationship between previous test scores and improvements in test scores, and the best fit I found was a quadratic relationship. The possibilities of logarithmic and hyperbolic relationships were explored, but they complicated the intuition of the model while not fitting the data any better. The estimation of the rest of the model was robust to changes in the test score specification. Eighth grade reading test scores are included in the model, as they could signal other components of ability.

4. Execution of the Models

4.1 Track Mobility and the Multinomial Logit

The first question at hand is that of ability track mobility. To further examine the question, we look at changes in track as utility-maximizing decisions. The utility gained as a result of switching from a low track to a different track can be described as

$$U(T_{ia}) = f(Y_{i8}, X_{i8}, X_{i10}) \tag{1}$$

In equation (1), T stands for the 10th grade track of student i in initial track a . Y_{i8} represents the mathematics test score of student i in eighth grade. X_{i8} is a vector of control variables reported by the student, teacher, school and parent level for eighth grade. X_{i10} is a similar vector of control variables for

the tenth grade. We will use a multinomial logit model to describe the probabilities of switching to various tracks.

4.2 Assumptions of the Multinomial Logit Model

As with any empirical model, the multinomial logit starts with some basic assumptions (Cramer, 1991).

- (1) The data are case specific. That is, each independent variable has only a single value for each case ;
- (2) The dependent variable is not perfectly predicted from the independent variables ;
- (3) The independence of irrelevant alternatives ;
- (4) The alternatives do not have an inherent order ; and
- (5) The error terms of the utility functions (equation 1) are independently and identically Gumbel distributed.

Now, the first two assumptions are easily confirmed with the data. The third merits some discussion. What it means is that the introduction of an irrelevant alternative will not affect the choice of a particular alternative. That is, if we are choosing whether to walk, bike or bus to school and wish to model it with a multinomial logit, the introduction of different types of buses (blue or pink, for example) should have no effect on whether I choose to bike or walk. Is this a good assumption? It is difficult to tell, so it should be kept in mind when making inferences.

Assumption four is easily confirmed, as the tracks are not used as sequencing devices. That is, I can assign each track different numbers without consequence to inference. Assumption five is highly unintuitive, but with the number of observations we have, the difference between the Gumbel distribution and the Normal distribution are quite small, indeed small enough to be ignored. If we chose to assume the errors to be normally distributed, we would need to use a multinomial probit model, which is costly both in terms of intuition and computation.

4.3 Mechanics of the Multinomial Logit Model

The model I employ differs from that of Argys, Rees and Brewer in that the dependent variable is modeled only for the subset of the student population that was initially in track a , in this case, the “lower levels of ability” track. In effect, this paper is looking for the conditional probability of placement in track b given that the student was initially in track a . This allows a more direct analysis of the test score effects of switching from one particular track to another.

A student enrolled in track a in 8th grade will choose to enroll in track b in 10th grade if and only if

$$U(b) > U(c) \quad (2)$$

for all available tracks $c \neq b$. A student chooses to remain in track a if

$$U(a) \geq U(c) \quad (3)$$

for all available tracks $c \neq a$. If the errors of equation (1) are Gumbel distributed, the probability that a student changes from track a to another track can be represented by a multinomial logit model. The probability of switching from track a to track b is:

$$Pr(T_{10} = b \mid T_8 = a) = \frac{e^{U(b)}}{1 + \sum_j e^{U(j)}} , \quad (4)$$

where $U(j)$ is the utility gained from being in track j . The probability of remaining in track a is considered the reference outcome and is represented by:

$$Pr(T_{10} = a \mid T_8 = a) = \frac{1}{1 + \sum_j e^{U(j)}} . \quad (5)$$

The multinomial logistic regression is used to determine the probability of an event with more than two discrete possible outcomes that do not have a natural ordering. In this case, we use the different possibilities for mobility as our dependent variable. It is clear from equations (4) and (5) that the summation of these numbers over all alternatives is equal to one, so it may indeed be viewed as a probability. Coefficients are estimated using maximum likelihood.

4.4 Results of the Multinomial Logit

The results from the estimation of the multinomial logit model, equation (1), are seen below in Table 3.

As expected, the eighth grade test score of a student in a lower achieving math course is a strong indicator of a move to average, untracked and higher achieving classes. An increase in score by 0.1 standard deviations is associated with an increase in the log odds of switching to an average track mathematics course of 0.062. The log odds of switching to a high achieving mathematics course are increased by 0.131 with a 0.1 standard deviation increase in score. Log odds of switching to untracked change by 0.068 with a test score increase of 0.1 standard deviations. This suggests that students switching to untracked settings are the ones that are scoring better initially, which would tend to bias test score effects upwards of a switch to an untracked class.

Table 3 What are the Determinants of Track Mobility? Estimates of Equation 1 Likelihood Ratio = 150.6 Standard Errors in Parentheses			
Variable	Average	High	Untracked
Intercept	-2.777 (0.813)	-8.497 (1.491)	-2.402 (0.980)
8th Grade Test Score	0.062*** (0.016)	0.131*** (0.021)	0.068*** (0.019)
8th Grade Class Size	0.063*** (0.016)	0.081*** (0.027)	0.053** (0.021)
Gender (1 = Female)	0.301 (0.219)	0.525 (0.368)	0.531* (0.284)
Socioeconomic Status	0.005 (0.164)	-0.013 (0.276)	-0.019 (0.219)
Band is Available in School (1 = Yes)	-1.216** (0.544)	0.013 (0.969)	-1.448** (0.644)
How Far Student Thinks she will Go in School (Increasing in Years)	0.198** (0.091)	0.305* (0.168)	-0.068 (0.115)
Student Often Does Not Turn in Homework (1 = Yes)	-0.642*** (0.252)	-2.467*** (0.777)	-0.403 (0.323)
Native American	1.421 (1.120)	2.738** (1.374)	1.720 (1.274)
African American	0.090 (0.310)	-0.039 (0.632)	-0.534 (0.473)
Hispanic	-0.243 (0.332)	0.013 (0.551)	-0.543 (0.472)
Asian American	-0.396 (0.539)	0.192 (0.775)	-0.083 (0.639)
*** Significant at the 1% Level ** Significant at the 5% Level *Significant at the 10% Level			

Having Band available as an extracurricular activity is negatively associated with switching to average or untracked settings. This suggests that the existence of band may cause

scheduling conflicts that could restrict a student's ability to switch tracks. However, the band option seems unlikely to prevent a student from moving to a higher track. Students may think that forgoing band or other extracurricular activities for which band proxies is worth the benefit of being among the higher achieving students.

Class size in the eighth grade class shows a positive and significant relationship to all track changes. This means that the larger the eighth grade class was for a particular student, the larger the probability is that he or she switched. The interpretation of this is not clear. It could mean that students were unhappy with larger class sizes, so they elected to move to another track. However, higher tracks tend to have higher average class sizes. Another idea is that the students in the larger classes that performed better would be more likely to be funneled out of the classes because the teachers and administrators had an incentive to keep class sizes low in the low-ability classes.

The eighth grade teacher reporting that a student rarely turns in homework is negatively associated with all track changes. This is not terribly surprising, as a student who lacks engagement in class may very well not be willing to take the initiative to make a change of course.

The student's expectation of how far she will advance in school is positively and significantly associated with a switch to average track. It shows a weak positive correspondence to switching to a high track and a weak negative correspondence to a switch to an average track. This variable is likely a proxy for motivation rather than raw intelligence. By sheer motivation, a student could probably take the more advanced material in the average class, but without the raw intelligence, success in the high track is less likely.

Being female shows a weak, insignificant association with changing to an average or high track, but a significant association with a switch to an untracked class. Socioeconomic status and race show no significance at any conventional p-value, providing no evidence to the theory that track mobility is reserved only for the privileged. Defacto segregation may be a cause of this, as schools are often homogeneous in race and socio-economic status. If this is the case, then it appears that schools with students with mostly low socio-economic status or that are mostly homogeneous in race still allow to change tracks. As expected by Hallinan (1994), variables other than simple achievement go into the decision to exercise upward track mobility. Contrary to the beliefs of many tracking critics, none of the race variables show statistical significance to the decision to change course designations. While race has been

shown (Argys, Rees & Brewer 1996) to be a determinant to track *placement*, it appears not to be significantly associated with track *mobility*. The same holds true for socio-economic status.

4.5 Test Score Effects and the Education Production Function

In this paper, I use a standard education production function to describe achievement at the tenth grade level:

$$(Y_{ia10} - Y_{i18}) = \beta_1 Y_{i18} + \beta_2 Y_{i18}^2 + \beta_3 X_i + \beta_4 T_{ia} + \varepsilon_i \quad (7)$$

In equation (7), $(Y_{ia10} - Y_{i18})$ represents the improvement in mathematics test score of student i from 8th to 10th grade who had an eighth grade track placement of l and a tenth grade track placement of a . The variable Y_{i18} indicates the initial test score of student i in eighth grade. Following Zimmer (2003) and Argys, Rees and Brewer (1996), this term (and its square) are intended to capture all of the effects of education inputs at the eighth grade level and below. Thus, the remaining terms should capture just the effects of these inputs for the tenth grade year. The independent variable X_i is a vector of control variables for the student, including eighth grade reading test score, socio-economic status, class size, race and teacher experience. The vector T_{ia} is composed of dummies indicating a change to track a by student i initially in track l . I estimate a production function only for those students initially in a track categorized as “lower achieving” by the teacher. With this methodology, I can compare the achievement of those who

have switched away from a low track to those who have remained in a low track. It is by this process that I am able to see whether switching to a different track is associated with test score improvements for students initially in lower-ability classrooms. Given the information from Table 1, it is possible and likely that track switches of students are endogenous with test scores. That is, students who score well above the average in a low ability track are likely to move to a higher track. Thus parameters attached to the vector T_{ia} are likely to be biased upward.

4.6 Assumptions of The Least Squares Estimation of (7)

The assumptions of the regression model used here are as follows (Greene):

- (1) The independent variables are fixed in repeated samples;
- (2) The dependent variable cannot be predicted perfectly from the independent variables;
- (3) The ε_i 's are normally distributed;
- (4) $E[\varepsilon_i] = 0$;
- (5) $\text{Var}[\varepsilon_i^2] = \sigma^2$ for all i (homoskedasticity); and
- (6) $\text{Cov}[\varepsilon_i, \varepsilon_j] = 0$ if $i \neq j$.

As will be discussed below, many of these assumptions are not realistic, and I must employ some correction mechanisms to make sure that any inferences are not misleading. In

particular, assumption (5) is rejected easily. The other assumptions cannot be rejected with our data.

4.7 Properties of Estimators and The Gauss-Markov Theorem

Least squares estimation is the typical choice for this type of function because it is intuitively tractable and computationally simple. The main result driving the use of least squares estimation is the Gauss-Markov Theorem. Before we discuss that theorem, here are definitions of some desirable properties of estimators.

Definition 1: An estimator, \hat{z} of z is said to be *unbiased* if $E[\hat{z}] = z$. In other words, if the mean of the sampling distribution of the estimator is equal to the parameter we are trying to estimate.

Definition 2: Let Z be a class of estimators $\hat{z} \in Z$. The estimator \hat{z} is said to be *relatively efficient* if $\text{Var}[\hat{z}] \leq \text{Var}[z]$ for all $z \in Z$.

These properties are by no means the only desirable properties in estimators, but they are some of the most easily testable. While unbiasedness is often desirable, it is possible that requiring unbiasedness could come with much greater costs in terms of efficiency. The efficiency of estimators is necessary to make inferences. That is, if the variance of the estimator's sampling distribution is very high, we will have a difficult time drawing any conclusions from that estimator.

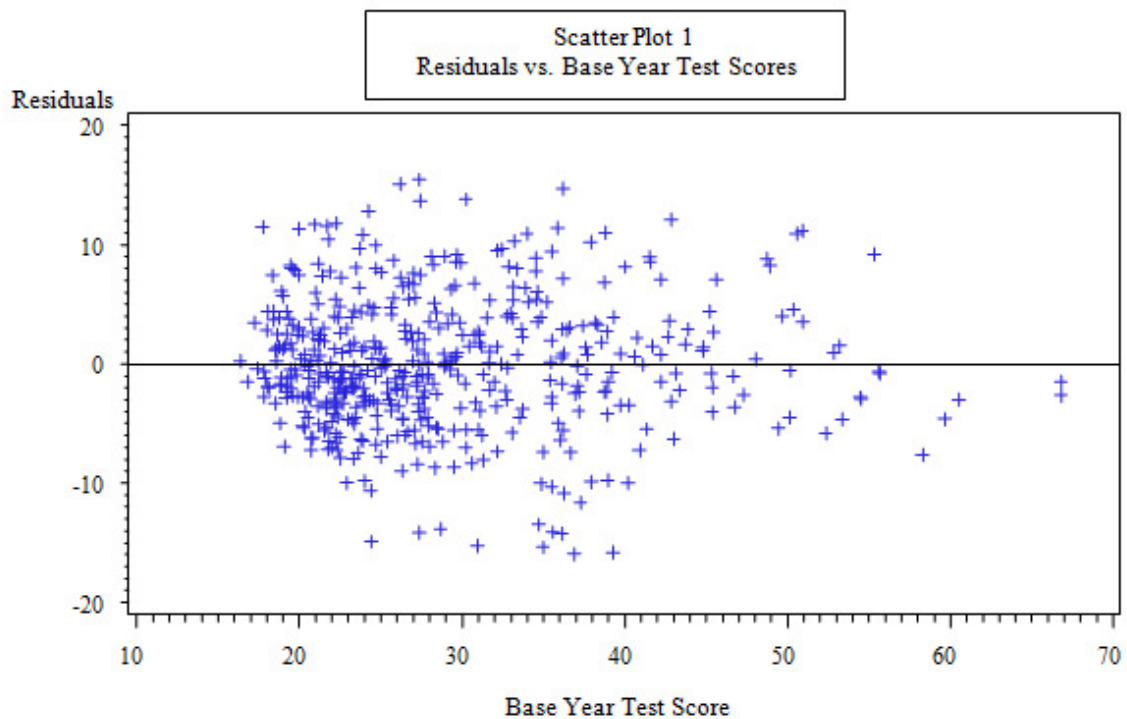
Theorem 1 (Gauss-Markov): In a model where the errors have expected value zero, are uncorrelated and have equal variance, the best linear unbiased estimators (BLUE) may be obtained through least squares estimation.

Best in terms of Gauss-Markov here means relatively efficient when compared with the class of linear and unbiased estimators. As mentioned before, it may not always be desirable to require that estimators be unbiased, but in this case, we will be satisfied with estimators that are relatively efficient in this class.

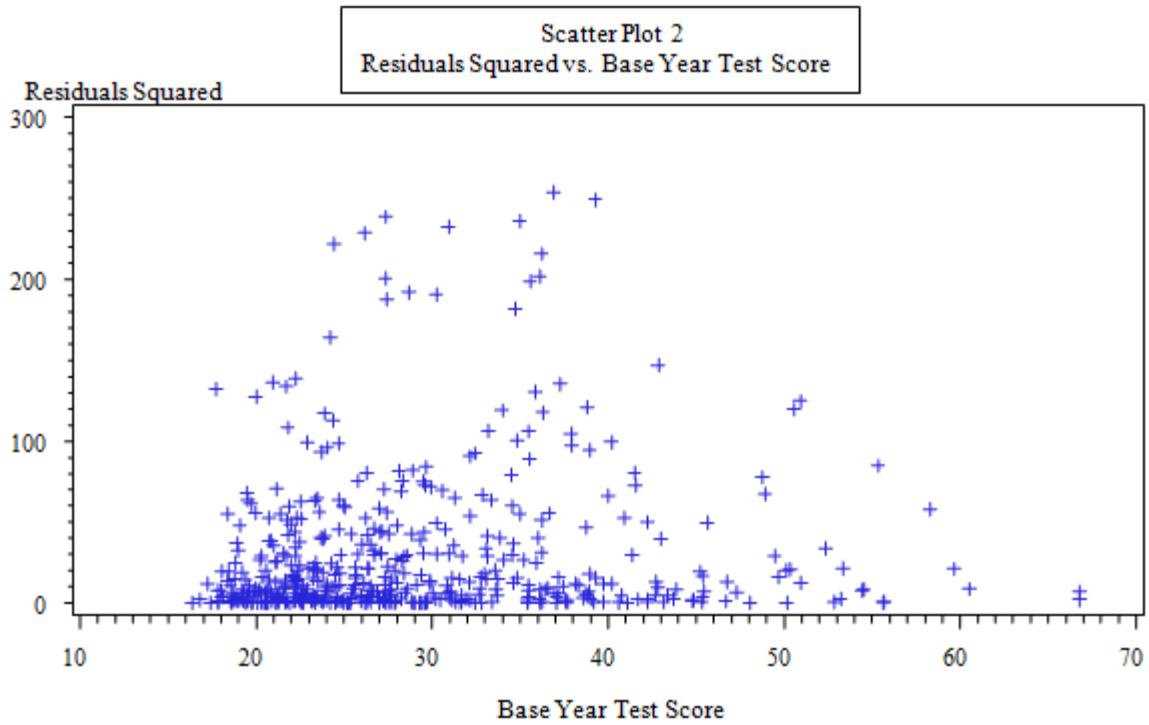
4.8 Heteroskedasticity

Definition 3: Regression disturbances whose variances are not constant across observations are said to be *heteroskedastic*.

As is often the case with cross-sectional data, there is a clear violation of the homoskedasticity condition in this data set. That is, the ‘equal variance’ condition on the errors in the Gauss-Markov theorem does not hold. As can be seen from Scatter Plot 1 and Scatter Plot 2, the residuals of the OLS estimation of equation (1) appear to be related to the base year test score, and are clearly not spread equally around the zero line. Further investigation also provides reason to believe that the residuals are also correlated with teacher experience.



To formally test for heteroskedasticity, we use two tests: the White test and the Breusch-Pagan test. The White test regresses the errors against all of the independent variables and tests the hypothesis that all of the coefficients are equal to zero. The Breusch-Pagan test regresses the errors against any subset of the independent variables that may be suspicious. This model fails the White test and the Breusch-Pagan test using teacher experience, base year test score and base year test score squared.



Heteroskedasticity is a problem for two reasons. First, it entails a violation in one of the assumptions of the Gauss-Markov theorem, which means that we cannot use this theorem to make valid inferences. That is, our estimators are no longer BLUE.

Next, the way we estimate our standard errors and our significance depends crucially on the fact that our disturbances have constant variance. That is, the variance/covariance matrix for our coefficients, Σ_{β^*} , is evaluated (Greene 2003) as

$$\Sigma_{\beta^*} = (X'X)^{-1}X'(\sigma^2I)X(X'X)^{-1} \quad (8)$$

where X is an $(n \times k)$ matrix with each column represents a control variable and each row represents an observation. Here, I is an $(n \times n)$ identity matrix, so $(\sigma^2 I)$ is a diagonal matrix with σ^2 in every diagonal entry. Also, σ^2 represents the variance of the disturbances which we have assumed to be constant. As such, using simple algebra, we can simplify the VCV matrix down to,

$$\Sigma_{\beta^*} = \sigma^2 (X'X)^{-1}(X'X)(X'X)^{-1} = \sigma^2 (X'X)^{-1} \quad (9)$$

This gives us a $(k \times k)$ variance/covariance matrix where the k^{th} diagonal entry represent the square of the standard error of variable k . Since σ^2 is not known, it is estimated by the disturbances, but is still assumed to be constant across the disturbances. Clearly, if we have a situation where the σ^2 are not equal across observations, we cannot make the simplification, as we did in equation (9), and our standard errors that we estimate will not be correct.

To deal with this problem, I will take two approaches: One will adjust our estimation of standard errors and the other will entail a completely different way to estimate the model.

As we can see, heteroskedasticity causes the estimated standard errors to be incorrect because the standard error calculation supposes that the variances of the error terms are all the same. To address this, I will estimate robust standard errors (White 1980). This is, in a sense, a more conservative way to estimate the standard errors, thus causing us to make

inferences more cautiously than we otherwise would. To estimate the robust standard errors, we estimate σ^2 for each observation using the actual residuals. That is, in our computation of equation (8), we replace Σ_{β^*} with an estimation of $\Sigma_{\beta^*}^{HC}$, a heteroskedasticity-consistent estimator of the variance/covariance matrix using the ε_i^2 's, the squared error term, namely:

$$\Sigma_{\beta^*}^{HC} = (X'X)^{-1}X'\text{diag}(\varepsilon_i^2)X(X'X)^{-1} \quad (10)$$

Intuitively, this would make the sampling variance increase in the total error for each variable and thus the standard errors would increase in total error. White proved that this formulation of the standard error converges to the actual standard error asymptotically, so that with a large sample, we are not too much is sacrificed in terms of efficiency.

An important advantage to this approach is that it does not require us to know the exact functional form of the heteroskedasticity. In addition, since it causes the standard errors to increase in total squared disturbance, it provides us with conservative estimates of standard errors that would tend to steer us towards caution in inference.

The disadvantage to this approach is that our estimators are inefficient relative to other methods such as weighted least

squares (Greene 2003). Our estimators could potentially be chosen to have a lower variance. However, if the estimators remain significant with the robust standard errors, the inferences should not be misleading.

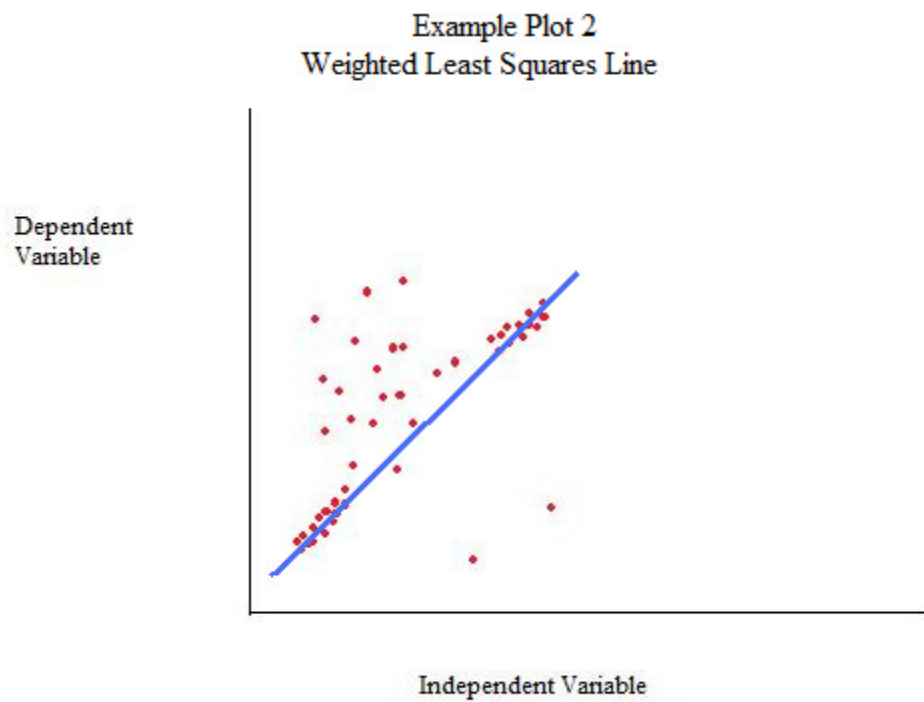
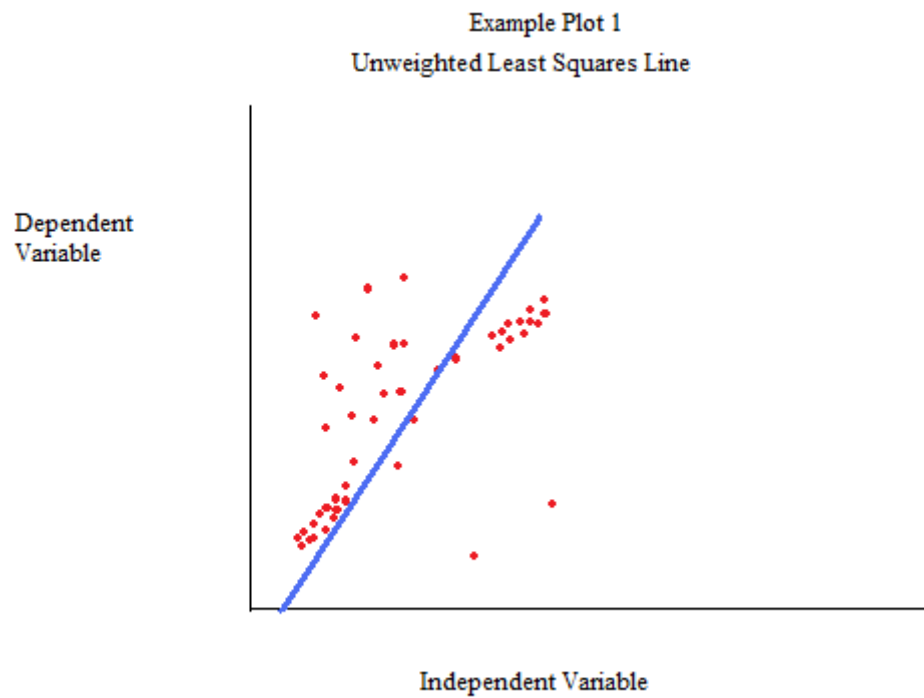
Next, to try and find more efficient estimators, I specify a weighted least squares (WLS) model. To do this, I note that the residuals appear to be related base year score and teacher experience, and I model the squared residuals as a function of base year score, base year score squared and teacher experience:

$$\varepsilon_i^2 = \beta_0 + \beta_1 Y_8 + \beta_2 Y_8^2 + \beta_3 E, \quad (11)$$

where Y_8 is 8th grade test score and E is teacher experience.

Then, the original model is estimated again, but each observation is weighted by the reciprocal of the predicted value of the squared residuals, following Greene (2003). When the measurements of the errors are uncorrelated but have different variances (i.e. are heteroskedastic), Greene points out that a weighting mechanism ought to be used. If a weighted sum of the squared residuals is minimized, β^* is BLUE if the reciprocal of the measurement's variance (i.e. the predicted residual) is used as the weight. If the squared residuals are modeled correctly, the resulting estimators will be efficient and inferences from the standard errors will be valid. The intuition behind the approach is that some observations have better predictive power than others, and we ought to pay more attention to the ones that

show a more systematic relationship to the independent variables. See the figures below.



While these plots are highly exaggerated, they are illustrative of the motivation behind the weighted least squares formulation. In the first example plot, you can see that for the middle values of the independent variable, the errors are quite large. When estimating the line, we take every observation to have equal value and minimize the sum of the squared distances from the prediction line. As a result, our line is systematically wrong for all values of the independent variable because the observations with high variance are influencing the estimation.

In the second example plot, we put a higher priority on minimizing the squared distances from the prediction line at the values of the independent variable where the errors are likely to be lower. That way, we can see the systematic relationship between the independent and dependent variables more clearly and with less error for some values and with greater error for other values of the independent variable.

The disadvantage to this approach is that it is difficult to know the exact functional form of the squared residuals, though the plots do give an intuitive idea. In addition, this method devalues observations with large residuals. It is possible that those observations actually contain a lot of information. When making inferences, we must keep in mind the behavior of the residuals, as it is certain that our model has

less predictive power over some subset of the independent variable. However, upon using this estimation using weights, the model passes both the White and the Breusch-Pagan tests.

4.9 Results of Approach I - Robust Standard Errors

In Table 4, results from the initial OLS estimation are presented with the original, incorrect, standard errors as well as the heteroskedasticity consistent standard errors. While some of the robust standard errors are larger than those estimated assuming homoskedasticity, none of them are drastically different. In fact, if we accept a 5% level of significance, none of our variables change from significant to insignificant or vice versa.

The variables relating to track changes show some of what we expected. Switching from a low track to either an average track or a higher track is associated with higher test score gains. Of course, the mean base year test scores show that these students are significantly brighter than average among the students in our sample, implying that this coefficient does not necessarily represent a valid estimate of the consequences of moving a typical student to a higher track. Interestingly, despite the fact that those students switching from a low track to an untracked class have higher initial scores, the parameter associated with a switch to an untracked class is negative,

though not significant. If we take a 95% confidence interval of the effect of mixed classes,

Table 4 Parameter Estimates for Equation (7) with Robust Standard Errors $R^2 = 0.226$, 556 Observations, 66 with Missing Values							
Variable	Parameter Estimate	Standard Error	t Value	Pr > t	Heteroscedasticity Consistent		
					Std Error	t Value	Pr > t
Intercept	-2.822	2.811	-1.00	0.315	2.521	-1.12	0.263
Base Year Mathematics Test Score	0.444	0.167	2.66	0.008	0.156	2.85	0.004
Base Year Mathematics Test Score Squared	-0.008	0.002	-3.38	0.001	0.002	-3.77	<0.001
Base Year Reading Test Score	0.143	0.046	3.08	0.002	0.045	3.16	0.002
“Average Levels of Ability” in 10 th grade	2.805	0.618	4.54	<0.001	0.612	4.58	<0.001
“Varying Levels of Ability” in 10 th grade	-0.056	0.805	-0.07	0.945	0.831	-0.07	0.947
“Higher Levels of Ability” in 10 th grade	6.293	1.018	6.18	<0.001	1.129	5.57	<0.001
Percent of Students in Remedial Math	-0.093	0.027	-3.45	<0.001	0.027	-3.42	<0.001
Socioeconomic Status Composite	1.026	0.357	2.87	0.004	0.363	2.82	0.005
Sex (1 = female)	-1.142	0.516	-2.21	0.027	0.509	-2.24	0.025
Teacher Experience	-0.308	0.100	-3.07	0.002	0.102	-3.01	0.003
Class Size	0.036	0.038	0.94	0.349	0.037	0.98	0.330
Asian	-1.976	1.182	-1.67	0.095	1.277	-1.55	0.123
African American	-0.486	0.774	-0.63	0.530	0.717	-0.68	0.498
Hispanic	1.225	0.785	1.56	0.119	0.735	1.67	0.096
Native American	2.259	2.022	1.12	0.265	1.761	1.28	0.200

the *upper limit* of the interval has the student improving by only 1.5 questions as a result of the change. The lower limit has the student doing worse by about 1.7 questions. In view of the higher initial test scores, this number may even be biased upward. These numbers are hardly enough to inspire widespread reform.

Somewhat surprisingly, class size shows a positive relationship with test score gains. While this is consistent with findings from Argys et. al. (1996), it seems to contradict the more common practice of making lower-track classes small. This question almost surely warrants further research to determine what exactly is causing the positive class size effect. Also consistent with Argys et. al, being Asian shows a negative and not quite significant association with test score improvement in this sample. Finally, it is extremely interesting to note that teacher experience shows a strong negative association with test score gains. It seems that the students in a low track benefit more from inexperienced teachers than from experienced ones. Perhaps the experienced teachers who have proven themselves as high quality instructors have been rewarded with higher tracks and the only remaining experienced teachers in the low tracks are the largely unsuccessful ones. If that is the case, then it is not surprising that the inexperienced teachers would have a greater positive effect on

test scores than the experienced ones, on average. This suggests that perhaps closer attention ought to be paid to the process that matches teachers to classrooms.

As might be expected, socioeconomic status shows a strong positive test score effect. Also, the percent of students in the school needing remedial math shows a strong negative effect on test score gains. It seems that there are intra-school peer effects in mathematics. The positive spillovers may come from higher ability students' tendency to be wealthy, from parental volunteer time or simply from providing motivation. This finding supports Figlio's conclusion that high ability and average ability students leaving the school would be detrimental to low ability students.

Of course, some of our inferences about significance may be understated, as the robust standard errors approach produces consistent, but inefficient estimators. Even though the robust standard errors do not appear much different from the original ones, we may be able to improve our estimation with the weighted least squares.

4.10 Results of Approach II - Weighted Least Squares

To estimate the weighted least squares model, the residuals squared must first be regressed against base year mathematics test score, base year mathematics test score squared, and teacher experience to obtain weights. The results from that

estimation are presented in Table 5. As expected from the scatter plots, both of the test score terms are significant and the squared term is negative, creating the concave down appearance. Also as expected, the residuals squared are positively related to teacher experience.

Table 5 Estimation of Residuals Squared, Equation (11) $R^2 = 0.0475$				
Variable	Parameter Estimate	Standard Error	t-Value	Pr > t
Intercept	-61.776	20.506		
8 th grade Mathematics Test Score	5.001	1.236	4.05	<.0001
8 th grade Mathematics Test Score Squared	-0.065	0.017	-3.72	0.0002
Teacher Experience	1.891	0.793	2.38	0.0176

From here, I use the reciprocal of the predicted value of the residual squared as a weight for the original regression, following Greene (2003). If the weights are correct, the resulting estimators should be both consistent and efficient. The results of the weighted regression are presented in Table 6. While it is difficult to know for sure the functional form of the disturbances, we will have some measure of success or failure by tests for heteroskedasticity after the weighted estimation. This weighted least squares model passes both the White and Breusch-Pagan tests with p-values of 0.16 and 0.99 respectively. Prior to weighting, this specification failed the tests with p-values of 0.01 and 0.0001. So while we may not

have the exact functional form of the residuals, we now do not reject the hypothesis that our errors are homoskedastic.

Table 6 Weighted Least Squares Estimation of Equation (7) $R^2 = 0.226$				
Variable	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	-2.759	2.474	-1.12	0.2654
8 th Grade Mathematics Test Score	0.447	0.153	2.91	0.0037
8 th Grade Mathematics Test Score Squared	-0.008	0.002	-3.76	0.0002
8 th Grade Reading Test Score	0.120	0.047	2.57	0.0105
“Average Levels of Ability” 10 th grade	2.428	0.578	4.20	<.0001
“Varying Levels of Ability” 10 th grade	-0.108	0.752	-0.14	0.8858
“Higher Levels of Ability” 10 th grade	6.298	1.019	6.18	<.0001
Percent of Students in Remedial Math	-0.080	0.024	-3.30	0.0010
Socio-Economic Status	1.063	0.338	3.14	0.0018
Sex	-0.988	0.477	-2.07	0.0389
Teacher Experience	-0.279	0.101	-2.77	0.0058
Class Size	0.059	0.035	1.71	0.0872
Asian	-2.419	1.126	-2.15	0.0323
African American	-0.593	0.677	-0.88	0.3816
Hispanic	0.581	0.695	0.84	0.4030
Native American	3.085	2.110	1.46	0.1444

From the table we can see that most of the variables have the same signs and interpretations, though some of the magnitudes have changed a small amount. The only variable that changed from significant to insignificant at the 5% level is Hispanic. Notably, the track change variables did not change appreciably- in fact, it shifted downward. A change from a low

track to an untracked class is still negative and insignificant, with a relatively tight confidence interval centered slightly left of zero, while changes from low to average and high are positive and significant.

5 Where Else Could We Go From Here?

5.1 A Look into Matching - Background and Introduction

As mentioned above, the methods I use for modeling these questions are all limited in different ways. With the limited powers of the available models and data, what is the best course of action?

The theory of matching processes could be used to model an educational environment where students are given a choice among classes within a school. One could use the empirical model of education production as well as the empirical discrete choice model as inputs into the design of indifference solving mechanisms for the matching process. From here, it would be valuable to estimate the potential costs and efficiency gains of such a scheme.

The most common application of the matching process is what is commonly known as the stable marriage problem (Gale and Shapley, 1962). That process starts with two sets: n men comprise set M and n women comprise the set W . For a marriage to occur, an element of M and an element of W must agree to

marry. To have stable marriages, scenarios where a change could make everyone better or equal off must not be possible.

Matching processes have been applied to problems ranging from college football games to matching medical school graduates to residencies. Roth et al (2008) use an extension of the matching process to match students to schools in New York City. Various tie breaking mechanisms must be analyzed to resolve indifferences in preferences.

The matching process could apply to the choice of a classroom within a school, as opposed to schools within a district, with some changes to the structure of the model. The design of such a model should be guided, in part, by empirical results regarding education production and revealed preferences of students.

Why might we want to consider such a scheme? The empirical literature on education production is vast. Krueger & Whitmore (2000) find that some categories of students are more responsive to class size reductions than others. Hanushek (2003) suggests that peer effects play a role in schools and that different categories of students have different responses to the ability of their peers. Hanushek (1986) finds many inputs that have far different effects on different categories of students. Keeping track of how each individual will respond to inputs and trying to provide them accordingly could bring up information and

administrative cost problems akin to those of central planning. Indeed the findings in this paper showed some surprising results for a small subset of the student population.

Allowing choice within a school may allow students and parents to maximize their own utility within a school, alleviating some of the costs of ineffective or unwanted inputs as well as allowing students to directly express their preferences for learning. To my knowledge, there are no empirical or theoretical inquiries into this kind of a scheme.

5.2 - Proposed Model Idea:

To model an intra-school class choice scheme using a matching process. The following research questions must be addressed:

- (1) What are the Model Parameters:** What can we learn from the existing applications of the matching process? What assumptions about the initial parameters must we make? Given those, what rules for the process will provide efficiency and/or strategy-proofness?
- (2) How would we Design Such a Mechanism?:** How can we optimally design a set of initial options for each school so that the cost of changing options after students express their preferences can be minimized? What rules can be put in place to allow schools to be flexible in the face of varying preferences from year to year? Given a set of options, how

will groups of students make choices? How will we account for student indifferences or default assignments?

(3) What are the Consequences in terms of efficiency?

5.3 - Addressing the Model Questions:

(1) Model Parameters

View class selection as a matching game, matching students to classes. The Roth (2008) model assumes a fixed capacity of each school. Instead of a fixed capacity for each class, we will assume a minimum number of students per class and a fixed school size. Students will express their preferences, and the matching will be achieved in a way similar to the Roth model. However, we will relax the assumption of a fixed set of classes. That is, if all students in a school want the same type of class, some or all of the classes can change to provide several copies of a specified class. This will likely complicate the matching process. The metamorphosis could drastically change the implications in the Roth model with regards to strategy and efficiency. I would like to eventually relax the assumption of fixed school size to allow for school expansion in a more dynamic model.

(2) Design

Using intuition gained from my current empirical research along with data such as the *National Education Longitudinal Study of 1988*, I will empirically model education production

functions for a number of subsets of students. This data follows students from eighth grade into the workforce, surveying students, teachers, parents and administrators. Criteria other than expected test scores may factor into students choices, as suggested by Hallinan (1996), so a discrete choice model including extra-curricular activities, peer group composition, teacher characteristics and other factors as inputs should be specified to determine what students value when choosing a class. Guided by these empirical results regarding efficiency and revealed preference, initial course offerings can be designed with the purpose of minimizing the need for changes in course offerings after student preferences are known. These results will also allow us to make a thoughtful decision about the default class placements of indifferent students.

(3) Consequences

Using predictions based on the empirical models, we can estimate the potential efficiency gains of this plan. After implementation, we can test whether the prediction held true.

5.4 - Broader Impacts of the Proposal

Through this research, I plan to contribute to the field in three very important ways: (1) extend the theory of matching processes to make the model more accessible to other applications where capacity constraints are not realistic and changes in one of the matching sets are feasible; (2) create

realistic policy proposals that could potentially increase efficiency in the public school system; (3) bring into the academic and policy conversation the idea of students taking ownership in their educations. The fostering of this kind of attitude has potential to help all kinds of students. To my knowledge, no such model of the distribution of students across classrooms within a school has yet been developed.

6 Conclusions

Ability tracking is a controversial practice at all levels of education as evidenced by the plethora of literature on the subject. However, the nature of tracking in different settings is remarkably different. The theory of tracking assumes that instruction will be targeted at a specific group, raising the efficiency of classroom placements. Of course, as with all theories, there are places for unintended consequences. Much literature in the economics of education is devoted to the theory of school choice. Perhaps school choice may be extended to include class choice as described in the matching model above, as class placement at the high school level can be viewed as a utility maximizing decision for students, and there may be a separating equilibrium based on varying classroom and teacher characteristics. Further research on this subject may prove extremely useful to teachers, parents, and administrators alike. Perhaps the only difference in many tracks is the label, and the

label is what is actually detrimental to both the teacher's view of the class and the student's view of himself or herself. Evidence from this paper shows that students switching from a track labeled "low" to a track labeled "average" show greater improvement in test scores than those who remain in the low track. Of course, selection bias may be driving this result, and with a larger sample size could very well have been significant. However, the focus of this paper is to determine whether students who switch from a lower achieving class to an untracked class perform systematically better than their counterparts who remain in the lower performing track. Evidence from this paper provides evidence to refute that hypothesis. The upper end of the 95% confidence interval would have a student improving by about 1.4 questions, or 0.14 standard deviations. This number is not significant enough to warrant widespread reform.

The claim of critics that students are "locked in" to particular tracks and unable to move seems to be completely unfounded, as more than half of our sample of eighth graders initially in a low track exercised mobility and switched mathematics courses. As expected by some, both academic and non-academic factors are related to the decision to switch classes. With a larger data set and more variables, we may be able to deduce more about the nature of track mobility and the

possibility of making sure that our school systems can compensate for initial errors in placement and differences in student preferences with regards to the composition of a classroom. At any rate, it appears that students who perform above average for their track can and do exercise vertical (and tracked to untracked) mobility in mathematics courses.

Considering all of the literature on the subject and the nature of tracking at the high school level, it seems that the call for complete detracking is unfounded. Taking away the choice of a more difficult class from a young adult contradicts much of what the education system is trying to do in encouraging students to be ambitious and successful. Indeed, the evidence does not even support the claim that students initially in lower performing classes improve with a switch to a class of varying levels of performance. However, administrators should think about the way classes are labeled and the way in which resources are divided among classes, as the evidence shows that different students respond differently to different educational inputs.

7. References

- Akerhielm, Karen (1995), Does Class Size Matter? Economics of Education Review 14(3) pp. 229-241.
- Argys, Laura M., Daniel I. Rees and Dominic J. Brewer (1996), Detracking America's Schools: Equity at Zero Cost? Journal of Policy Analysis and Management 15(4), pp. 623-645.
- Boozer, Michael and Cecilia Rouse (2001), Intraclass Variation in Class Size: Patterns and Implications. Journal of Urban Economics 50, pp. 163-189.
- Cramer, J.S. (1991). *The Logit Model: An Introduction for Economists*. New York: Routledge, Chapman and Hall.
- Figlio, David N. and Marianne E. Page (2000), School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Equality? Working Paper No. 8055, National Bureau of Economic Research.
- Finn, Jeremy D., and Charles M. Achilles (1990), Answers and Questions about Class Size: A Statewide Experiment, American Educational Research Journal 27(3), pp. 557-577.
- Greene, William H. (2000). *Econometric Analysis*. Upper Saddle-River: Prentice-Hall.
- Hallinan, Maureen T. (1994), Tracking: From Theory to Practice, Sociology of Education 67(2), pp. 79-84.
- Hallinan, Maureen T. (1996), Track Mobility in Second School, Social Forces 74(3), pp. 983-1002.
- Hanushek, Eric A., John F. Kain, and Steven G. Rivkin (1998), Does Special Education Raise Academic Achievement for Students with Disabilities? Working Paper No. 6690, National Bureau of Economic Research.
- Hanushek, Eric A. (1999), Some Findings From an Independent Investigation of the Tennessee STAR Experiment and From Other Investigations of Class Size Effects. Education Evaluation and Policy Analysis 21(2), pp. 143-163.

Hoover-Dempsey, Kathleen V., Otto C. Bassler, and Jane S. Brissie (1987), Parent Involvement: Contributions of Teacher Efficacy, School Socioeconomic Status, and Other School Characteristics. *American Educational Research Journal* 24(3), pp. 417-435.

Krueger, Alan B. (1997), Experimental Estimates of Education Production Functions. Working Paper No. 6051, National Bureau of Economic Research.

Krueger, Alan B. and Diane M. Whitmore (2000), The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR," Working Paper No. 7656, National Bureau of Economic Research.

Oakes, Jeannie (1992), Can tracking research inform practice? Technical, normative, and political considerations. *Educational Researcher* 21(4), pp. 12-21.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain (2005), Teachers, Schools, and Academic Achievement. *Econometrica* 73(2), pp 417-458.

Shapson, Stan M, Edgar N. Wright, Gary Eason, and John Fitzgerald (1980), An Experimental Study of the Effects of Class Size. *American Educational Research Journal* 17(2), pp. 141-152.

Slavin, Robert E (1990), Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis. *Review of Educational Research* 60(3), pp. 471-499.

Summers, Anita A. and Barbara L. Wolfe (1975), Equality of Educational Opportunity Quantified: A Production Function Approach. Philadelphia Federal Reserve Papers, 1975.

White, Halbert (1980), A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48(4), pp 817-838.

Zimmer, Ron (2003), A New Twist in the Educational Tracking Debate. *Economics of Education Review* 22, pp. 307-315.