The Formation and Human Capital Consequences of High School Social Networks

Sean Lewis-Faupel*

Last Updated: February 1, 2016

For the most current version of this paper, please click here.

Abstract

This paper improves our understanding of how education policy interacts with peer effects to impact academic outcomes. In particular, I show that altering a student’s prospective friends – classmates who a student encounters and may befriend – affects high school and college success through friendships that subsequently develop. Policies such as ability tracking, the establishment of charter schools, and student busing are important determinants of these prospective friends. In this paper, I undertake two related analyses. First, I show that the attributes of a student’s prospective friends are associated with academic performance during and after high school using variation across grade levels within schools. Second, I measure heterogeneity across schools in the rate at which different types of students meet one another using a dynamic model of network formation. The model allows me to differentiate between preferences for friends, school composition, and meeting rates, which is key for evaluating a range of policies. Relative to other empirical models of social network formation, my model has two distinct advantages. First, I identify the meeting rates separately from preferences for friend attributes using data on friend rankings. Second, students in my model form expectations over future meetings and take into account peer effects. Using the estimates, I evaluate the relationship between meeting rates and a measure of classroom tracking. Finally, I use the estimated model to simulate changes to meeting rates and their impact on high school and college outcomes.

*The author is at the University of Wisconsin–Madison. This research received support from the grant T32 HD007014, awarded to the Center for Demography and Ecology at the University of Wisconsin–Madison by the Eunice Kennedy Shriver National Institute of Child Health & Human Development.
1 Introduction

A friendship begins with a meeting. In the context of high school, the composition of meetings that a student experiences depends on the makeup of the student body as well as the structure of course enrollment within the school. Consequently, variation in these conditions could have a significant impact on the structure of high school social networks. To the extent that friends are important in determining high school academic outcomes, the conditions under which friendship networks form are important to students’ high school success.

This paper makes several empirical contributions related to understanding this relationship. First, I document an association between the composition of the pool of potential friends (that is, the characteristics of students who are likely to encounter and befriend one another) and measures of academic success and human capital. While the importance of peers has been demonstrated many times over, I highlight evidence that suggests actual friends are a key channel through which peers at large matter. By using variation within high schools across grade levels in the attributes of peers, I control for grade-invariant selection into schools. The results indicate that students who are exposed to potential friends who are more likely to succeed academically are themselves more likely to perform well. The results are also consistent with realized friendships as one channel through which prospective friends matter.

These results motivate a more thorough model of the mechanisms that lead from who students meet to who they befriend to the interactions that ultimately occur between those friends. The model, unlike the regressions described above, allows me to consider several inputs to the network formation process: preferences for friend attributes, school composition, and the rate at which different types of students meet one another. Discerning between these factors is integral to evaluating the social effects of policies, such as ability tracking (which could change meeting rates) or school choice programs (which could change school composition), while accounting for students’ preferences for friends. These preferences are likely to, one, stay relatively constant across policies, and two, potentially play a large role in network formation regardless of the other inputs. Failure to account for these preferences could easily allow one to overstate the potential effects of policies.
Given these extensive findings, one of the main contributions of this paper is to separately identify these preferences from the rate at which students encounter prospective friends. To this end, I use the patterns observed in students’ rankings of their friends to infer preferences for traits (Beggs, Cardell, and Hausman (1981) provide the framework for the estimation).

Controlling for these friend preferences, I calculate meeting rates among prospective friends from a dynamic network formation model that identifies these rates from the empirical probability that each type pair is friends. The model is based on work done by Christakis et al. (2010) and Goldsmith-Pinkham and Imbens (2013). This dynamic, finite horizon approach has advantages over simultaneous move games, infinite horizon models, search models, and standard logistic models. Aside from the focus on meeting rates, I extend the existing literature on empirical network formation by allowing the agents to partially account for the future effects of a friendship when deciding whether to form a connection. In particular, the students have expectations over both the remaining students they may encounter as well as the incentives generated by peer effects once the network has formed. Without modeling this potential foresight, I would fail to account for the fact that students may become friends because of the expectation that the friends will incentivize later behavior (either good or bad). By estimating this model separately for each school, I obtain a distribution of meeting rates. To link these estimates to policy, I measure the relationship between the meeting rates and a measure how representative a typical course is on the school’s population.

---

1 Allowing friendship formation to depend on the current state of a dynamic network avoids issues with multiple equilibria that would occur in a simultaneous move game.

2 Allowing an infinite number of meetings (and thus extending history to be infinite), in the network formation literature, generates steady state Markov transition probabilities under certain conditions. However, modeling an infinite number of meetings destroys the notion that some friendships may not form because the pair never encountered one another, thus eliminating the policy lever this paper is interested in. Furthermore, even in moderately-sized social settings, it seems unlikely that all participants would be exposed to one another to the extent that would allow for a friendship to form.

3 In a typical labor search model, the order in which workers encounter firms does not matter because workers can only have one job at a time and firms typically do not have decreasing returns to labor. Relaxing either of these assumptions complicates calculations since the point in history at which the parties meet then matters. The model used here can be thought of as a finite time search model where workers can work multiple jobs simultaneously (but are still time constrained) and firms have either decreasing returns to labor, a fixed number of vacancies, or simply cannot search forever.

4 Compared to a standard logistic model of friendship formation (where an strategic interactions are ignored), the dynamic model is better equipped to evaluate counterfactuals since it accounts for the manner which changes to others’ meetings rates might filter through. The dynamic model should also be able to better fit the data since it allows for an additional dimension of randomness (i.e., stochastic meetings).
Finally, in order to link changes in network formation to outcomes, I estimate a binary choice model where friends influence each other’s decisions. In the standard model I employ, friend effects are identified because of the non-linearity of choice probabilities as well as the network structure of data (Brock and Durlauf, 2001; Lee, Li, and Lin, 2014). Using the combined network formation and peer effects model, I estimate the implications of several changes in the environment of friendship formation, each corresponding to a potential policy: changes in meeting rates, changes in school size, and changes in school composition.

There is a wide literature describing the role peers can play in academic decisions in high school. Evidence from randomized experiments, quasi-random data, and a variety of identification techniques consistently suggests that peer effects exist and can be substantial, in particular for the academic outcomes of children and young adults. For reviews, see Epple and Romano (2011) and Sacerdote (2011). Given the importance of preferences, the value of linking policy to peer effects is potentially quite high. One the few papers to directly tackle policy and peer effects, Carrell et al. (2013), suggests that individuals’ preferences for friends are strong enough that, at least in the context of the Air Force, it may be difficult to design a policy that takes advantage of peer effects. I complement this previous work by examining the case of a representative sample of U.S. high school students. My model also allows me to test a much wider range of policies than those the authors addressed with their single experiment.

This is also one of the few papers to combine network formation and peer effects in a single model. Among the others who have taken this approach, of greatest relevance are Badev (2013) and Goldsmith-Pinkham and Imbens (2013), with the former modeling peer effects simultaneously with network formation and the latter modeling the two sequentially. While the focus of Goldsmith-Pinkham and Imbens differs from mine, the model has much in common with the one used here. But, as discussed above, I extend their approach to allow for greater foresight among agents during the network formation.

In both Badev (2013) and Goldsmith-Pinkham and Imbens (2013), as well as in most other network formation papers, the estimation of network formation cannot differentiate between parameters linked to students’ preferences and parameters determining meeting rates. Currarini et al. (2010) estimate a simple model that identifies meetings separately.
from preferences. However, their approach relies mostly on functional form assumptions and restricts preferences for friends to a single parameter. Richards-Shubik (2012) decomposes preferences for friends and the availability of sexual partners in high school (in the form of peer effects and equilibrium behavior, respectively). While the author’s approach is conceptually similar to mine, the model is in fact describing a very different process: to Richards-Shubik, students’ preferences are an equilibrium object (whereas I fix preferences) and meetings are not a function of any institutional policy. Perhaps most in the spirit of my approach, Mayer and Puller (2008) model prospective friend meetings and friend preferences distinctly and use Facebook data to estimate parameters for college students. The largest drawback of their approach relative to mine is a reliance on an assumption that the meetings are determined by a set of variables distinct from those over which preferences operate, and that no unobservables link the two. Furthermore, Mayer and Puller restrict agents to be myopic and do not link the network formation to peer effects or additional outcomes.

The paper proceeds with a discussion of the data used in each part of the analysis. In Section (3), I present results that suggest the attributes of a student’s prospective friends impact academic performance through actual friends. In Section (4), I describe a more detailed model of network formation and peer effects. The estimation of this model is related in Section (5). Section (6) presents results from the model estimation and several counterfactuals. Section (7) concludes.

2 Data

The data used for analysis primarily comes from the National Longitudinal Study of Adolescent to Adult Health (Add Health). Add Health is a school-based study that includes nearly the full population of eighty U.S. high schools. Collected in 1994 and 1995, the data includes 90,118 students students in grades 7-12. A subsample of 20,745 were given a longer interview at home (including a verbal intelligence test, known as the Peabody Picture Vocabulary Test, or PVT for short), and most of this subsample were followed in subsequent waves of the survey. Parents of students in the subsample were also interviewed in the first wave. During both the home and school interviews, students were asked to list
their friends. In later waves, high school transcripts were collected with grades and classes taken. In addition, income and educational attainment were collected in the 2007-2008 wave.

Table (1) lists summary statistics for students at the school level. Moments from the distributions of proportions and means of each variable listed are given. The typical school has over 500 students and is majority white, with black, Hispanic, and Asian students making up the remaining 40 percent of the student body. The average family in a typical school earns around $40,000 annually, with a parent having a college degree in only slightly over a third of the cases. In the typical school, 84 percent of students graduate, but only 30 percent continue on to college. In all of these measures, there is great diversity across schools in the sample.

3 Prospective Friends and Human Capital Outcomes

If friendships form from the pool of students a person encounters at school and if friend effects are meaningful, we should expect to find an association between a student’s own outcomes and the characteristics of classmates with whom the student is likely to befriend. In this section, I measure this relationship in the data. Because it is likely that better students (or their parents) select into schools with better peers, I include school fixed effects and use variation across grade levels within schools to identify the association. This variation across grades may be attributable to small-sample noise caused by factors such as genetics, frictions in the housing market, or incomplete information. However, I cannot easily rule out unobservables that are correlated with the error generating at least some of this variation; thus, I interpret these results as descriptive evidence. Despite this limitation, the absence of a detectable relationship might call into question the idea that the environment of friendship formation matters.

I focus on three outcomes of success in and after high school: high school graduation, high school GPA (out of 4.0), and college attendance. I regress each of these outcomes on the mean characteristics of a group of students who we expect might be the pool from which friends are drawn, controlling for a student’s own characteristics. These mean characteristics are PVT score, number of years of parental education, and number of
parents in the household. As a student’s prospective friends, I use those students in the
same school grade in the first specification. The regressions take the form

\[ y_i = \beta' \bar{X}_{g(i),s(i)} + \gamma' X_i + \phi_{s(i)} + \theta_{g(i)} + \varepsilon_{i,s(i)} \] (1)

where \( y_i \) is an outcome for student \( i \), \( \bar{X}_{g(i),s(i)} \) is a vector of mean characteristics for those
in student \( i \)’s grade and school; \( X_i \) is a vector of those same attributes for the student \( i \),
plus several other individual controls; and \( \phi_{s(i)} \) and \( \theta_{g(i)} \) are school and grade intercepts
for student \( i \), respectively.

The results are presented in Table (2.1). A statistically significant negative association
between having prospective friends with single parents and the outcomes is observed,
though the magnitude is modest. Moving from none to all of grademates with single
parents corresponds to a 0.414 grade point drop in high school GPA. A similar negative
effect on prospective friends’ average PVT score is counterintuitive. However, this may be
the result of competition: Given two students with the same PVT score at the same high
school competing with differently skilled classmates for, for example, a limited number of
A grades, we might expect the student with weaker competition to perform better.

In the next specification, I use grademates of the same sex and gender as the group of
prospective friends. If students generally prefer friends with attributes similar to their
own, students of the same race and gender may matter more than the attributes of the
entire grade level. I replace school fixed effects with school by type (gender and race)
fixed effects so that I am now comparing two students in the same school and gender by
race bin but in different grades (I was previously controlling for gender and race but their
effects were not allowed to vary across schools in Table (2.1).) These fixed effects have
the advantage of controlling for students selecting into schools based on the characteristics
of students of their own sex and race (at the school level – selection at the grade level
remains a potential issue).

The specification now takes the form

\[ y_i = \beta' \bar{X}_{t(i),g(i),s(i)} + \gamma' X_i + \phi_{s(i),t(i)} + \theta_{g(i)} + \varepsilon_{i,s(i),t(i)} \] (2)

where the notation matches that above with the added index of student \( i \)’s type (gender
and race): \( t(i) \). The results are presented in Table (2.2). Prospective friends’ PVT scores are no longer a significant factor associated with the outcome. Since grademates of the same gender and race are less representative of those with whom a student competes, these findings are in line with the competition story. Prospective friends with single parents and with high school graduate mothers correspond to significant changes in the high school outcomes, though magnitudes remain small. A one hundred percentage point increase in the latter is associated with almost four tenths of a grade point higher high school GPA.

While I have posited that prospective friends matter primarily due to actual friends, an alternative hypothesis is that peers at large have a direct effect on students unrelated to actual friends. To test for this, I run the regressions from Table (2.2) controlling for friend characteristics. These results are presented in Table (3). Most (but not all) of the coefficients that were previously significant drop in magnitude. Significant results load onto the friend effects, and the peer effects are almost entire statistically insignificant. These results suggest that the friendships may be the primary channel through which these peer matter. Again, these results are intended to support but not necessarily confirm the theorized relationship.

Overall, these patterns are consistent with a story of the pool of prospective friends affecting actual friends who in turn influence each other. Unfortunately, it is difficult to draw policy conclusions from such specific and small (the differences across grades in a school are not large on average) variation in school composition. I move next to a model of network formation and peer effects that accounts for meeting rates and school composition simultaneously and identifies their relative importance using variation within high schools across types of students in numbers of friends, rankings of friends, and actions of friends.
4 A Model Network Formation and Peer Effects

4.1 Environment and Preferences

4.1.1 Notation

The network formation model employed here is largely based on the model used by Christakis et al. (2010), with the addition of forward-looking students and a distinction between meeting rates and preferences for friend traits. Students exist in a predetermined, isolated school of size $I$. Each student is defined by an $A$-dimensional vector $X_i$ of constant attributes. This vector contains observables such as race, ability, and parental education. The set of attribute vectors across students in the school is aggregated by the $I \times A$ matrix $X$.

Friendships are notated in the symmetric, $I \times I$ matrix $D$ as a 1 for entries $d_{ij}$ and $d_{ji}$ when $i$ and $j$ are friends. These entries are zero otherwise. (This matrix is often referred to as the adjacency matrix.) Friendships are assumed to be undirected, and so $D$ is symmetric. A single row of $D$, notated as $D_i$, defines the set of friends of student $i$. Student $i$’s number of friends is given by $q_i = \sum_{j=1}^{I} d_{ij}$. Students will want to keep track of the mean of their friends’ attributes: $\bar{X}_i = (D_iX)' / q_i$. As will become clear, the network evolves across time, and so networks in different periods are noted with an $n$ subscript as $D_n$.

4.1.2 Timing

Prior to high school, pairs of students meet in a randomly determined sequence. The length of this sequence for a given school is fixed at $N$, and no meeting is repeated. Therefore, a given student will have some number of meetings (less than $N$) spaced (stochastically) throughout this process. The vector $M_{i,n}$ is defined to be this set of meetings for student $i$ up to period $n$. The set of all meetings for all students is given by the $N \times 2$ matrix $M$, where $m_{n,1}$ and $m_{n,2}$ identify the two individuals involved in the $n$th meeting. These meetings and their order are determined by a random process that corresponds to drawing one meeting at a time from the pool of all possible meetings, without replacement and with weights on each meeting of $\omega(X_i, X_j)$, where $X_i$ and $X_j$ are the characteristics of the
two students involved in a potential meeting. It is assumed that $i$ meeting $j$ is equivalent to $j$ meeting $i$, and so it is natural to have $\omega(X_i, X_j) = \omega(X_j, X_i) \forall X_i, X_j$.

When forming friendships, it is assumed that students are unaware of the structure of the social network, aside from knowing the identity of their own friends. During a meeting, students observe each other’s $X_i$, as well as an individual, mean-zero, i.i.d shock related to the pairing, $\epsilon_{j,i}$. Each decides independently whether forming a friendship is beneficial, and if both agree it is, a friendship is formed.\footnote{Other friendship formation conditions may be interesting to examine. For instance, one could assume transferable utility so that a friendship would form when the sum of the individual utilities is positive. However, because agents are forward looking, calculating the expected utility transfers is not computationally trivial. In addition, assuming individual, independent utilities rather than a joint utility production function allows for attribute preferences to be more easily identified and estimated.} There is no cost to accepting a friendship if rejected by the potential friend, and students in the model will never have an incentive to be untruthful about whether a friendship is individually beneficial. In this model, friendships are not allowed to be destroyed, even when there may be an incentive to do so.

Once social networks have formed, high school begins, and students make a binary choice regarding some academic outcome, $e \in \{0, 1\}$. While the model allows this choice to be quite general, I will proceed with the idea that the choice is high or low effort in school. Immediately prior to this decision, the students learn two things. The first is the structure of the network – they can now perfectly observe all social connections in the school. The second is a stochastic shock to their preference for the choice of effort. Students choose their level of effort simultaneously but cannot observe the actions of others (they do not know the preference shocks of others). Following the effort decision, students receive utility.

\section*{4.1.3 Preferences}

For a student with friends, utility comes from the mean traits and actions of the student’s friends ($\bar{X}_i, \bar{e}_i$), from the number of friends ($q_i$), from idiosyncratic friend payoffs ($\epsilon_{f,i}$), and
from their choice of effort \((e_i)\). The utility function takes the form

\[
U_{f, i}(X, e, D, \varepsilon_f, \varepsilon_e) = \alpha_0 \cdot q_i + \alpha_1 \cdot |X_i - \bar{X}_i| + D_i \cdot \varepsilon_{f,i} \\
+ \delta \cdot [\beta_0' X_i + \beta' X X_i + \varepsilon_{e,i}'] \cdot e_i + \beta_e' \cdot |e_i - \bar{e}_i|.
\] (3)

The parameter \(\alpha_0\) is a marginal cost or benefit for each additional friend. The degree to which students care of about differences in their friends’ traits is determined by the vector \(\alpha_1\), which weights the differences in \(X_i\) between the student and her friends. If we expect students to prefer friends with similar traits to their own, we should expect the entries of \(\alpha_1\) to be negative. The term \(\delta\) is only important for estimation and can be disregarded during the theoretical discussion of the model. The utility that the students receive from their choices of effort depend on their own traits, their friends’ mean traits, and an i.i.d shock. Finally, students’ utility depends on the difference between their choice of effort and the mean choice of effort among their friends \((\bar{e})\). The weight on this term, \(\beta_e\), will be negative if students wish to conform to their friends’ actions.

For a student with no friends, utility is only determined by the choice of effort. In this case, the utility function takes the form

\[
U_s(X_i, e_i, \varepsilon_{e,i}) = \delta \cdot (\beta' X X_i + \varepsilon_{e,i}) \cdot e_i.
\]

### 4.2 Optimal Behavior

Consider how students choose effort. Because this choice occurs simultaneously, students cannot observe \(\bar{e}\). Instead, they choose an \(e\) to maximize this utility given rational expectations over \(\bar{e}\). In this binary choice setting, rational expectations must satisfy

\[
\psi = Pr(\varepsilon_e > -\beta_0 - \beta' X - \beta' X \bar{X} + \beta_e \cdot [1 - 2D\psi]),
\] (4)

where \(\psi\) is the vector of rational expectations for each student’s \(e\) and \(D\) is the row-normalized form of \(D\). This structure simply says that students’ expectation of friends’ effort is consistent with the probability of that their friends choose high effort, given those expectations. Brock and Durlauf (2001) and Lee, Li, and Lin (2014) show that an
equilibrium exists given this setting. Furthermore, if $|\beta_e| < 2$, the rational expectations equilibrium is unique. The importance of this uniqueness is discussed in the identification section.

Now consider the optimal choices during the friendship formation phase. Unlike past empirical models, students in this model consider future utility when making these decisions. Recall that during friendship formation, students do not know the structure of the network. During the final meeting of the network formation process, student $m_{N,i}$ chooses whether to accept or reject a friendship with student $m_{N,j}$ based on the expectation of utility that will follow conditional on the friends and information about the network the student has accumulated over the course of previous meetings. This set of friends and information is summarized by the state object $S$. Thus, the binary choice of whether or not to accept friendship is the result of the following maximization

$$\max_{\{\text{accept}, \text{reject}\}} \left\{ E[U_{f,i}(X, e, D, \varepsilon_f, \varepsilon_e) \mid S'_{i,N}(j)], E[U_{f,i}(X, e, D, \varepsilon_f, \varepsilon_e) \mid S'_{i,N}(-j)] \right\} \quad (5)$$

where expectations are over the space of possible networks and

\begin{align*}
S'_{i,n}(-j) &= \{X, D_{i,n-1}, \varepsilon_{f,i,n}, \{M_{i,n-1}, j\}\}, \\
S'_{i,n}(j) &= \{X, \{D_{i,n-1}, j\}, \varepsilon_{f,i,n}, \{M_{i,n-1}, j\}\}.
\end{align*}

In earlier meeting periods, the choice is similar but the student also accounts for the possibility of future meetings. The expected utility that results from the possible paths through meetings and forming friends is described by the value function $V_{i,n}$ in period $n$ for student $i$. Therefore, the choice of whether to accept or reject a friendship in period $n$ results from the following maximization

$$\max_{\{\text{accept}, \text{reject}\}} \left\{ V_{i,n+1}(S'_{i,n}(j)), V_{i,n+1}(S'_{i,n}(-j)) \right\}. \quad (6)$$

This value function weights the values of the possible states the student might move to in

---

While the student does not know the network structure during the formation period, the student can discern some details about the network based on who she has met and whether friendships have formed. While this learning process may be of interest, its complexity will require simplifying assumptions in the estimable version of the model.
the following period by the probability of moving to each state. Three paths to the next period are possible. First, a student could meet no one. A second outcome is a meeting that does not lead to a friendship. In that case, the student’s state is only updated with a record of the meeting. The third possibility, of course, is that the student meets someone and becomes friends with that person. In that case, the state is updated with the new friend. This leads the value function to take the following form

\[
V_{i,n}(S_i) = \Pr(\text{no meeting}|S_i,n)V_{n+1}(S_i) + \sum_{j \neq i} \Pr(i \text{ meets } j|S_i) \\
\times \left[ \Pr(i,j \text{ become friends} | S_i,n) E[V_{i,n+1}(S_i'(j)) | i, j \text{ become friends}] \\
+ (1 - \Pr(i,j \text{ become friends} | S_i)) V_{i,n+1}(S_i'(-j)) \right].
\]

4.3 An Estimable Model

There are three challenges to solving the above model in a reasonable amount of time. First, the state space is large. Aside from the other components, keeping track of the identity of each friend is infeasible for all but the smallest schools in the sample. The additional objects in the state space are of an even greater dimension. A second challenge is calculating the probability of friendship formation in Equation (7). While the student can accurately forecast her own probability of acceptance, the probability that a meeting ends in friendship also depends on the probability of other students accepting. This probability, in turn, is a function of the status of the network and, not knowing this information, the student would integrate over all possible networks that could have formed up to that point in time, weighting each network based on what she knows from past meetings. Given the high dimensionality of even a small network, this integral cannot be feasibly calculated. Finally, the expectation of the final utility in Equation (5) requires integration over the same space and is equally impractical to compute.

To proceed, I make an observation and three assumptions. First, I observe that the list of friendship shocks, \( \varepsilon_{f,i} \), can be dropped from the state space. While they affect final utility, they will not inform future decisions because the shocks are assumed to be independent and

\footnote{Allowing the student to observe the full network in this stage only shifts the problem to the state space – the full network is much too large for use as a state when computing the value functions.}
additive. The first assumption I make is that students cannot recall the identities of those they previously met but do remember the average attributes and number of their friends: $\bar{X}_i$ and $q_i$. This allows me to collapse the state space to these two dimensions.

I next make an assumption regarding how students forecast the probability that others will accept friendship. In particular, I assume that students believe the friends of their potential friends to be identical in attributes to their potential friends (that is, when meeting $j$, $\bar{X}_j = X_j$) and of a deterministic number. This belief of the number of friends of potential friends is allowed to vary over the course of the network formation according to function $g(n)$. These beliefs dispense with the need to integrate over possible networks when the student forecasts future friendship formation since, in the eyes of a student, the network structure from which potential friends are drawn is fixed.

Finally, I assume similar heuristic behavior when a student forecasts how peers will impact their choice of effort. The student takes into account the fact that a friend will influence what effort they choose, but I assume the student ignores that fact that they may convince their friend to choose a different effort than they would otherwise. Specifically, a student forecasts $\bar{e}$ as if they have one friend with $X_j = \bar{X}_i$ who does not respond to peer effects.\(^8\) This results in the following heuristic expectations

$$\tilde{E}[\bar{e}] = 1 - f_e(-\beta' X \bar{X}_i)$$

$$\tilde{E}[e] = 1 - f_e(-(1 - 2\tilde{E}[\bar{e}]) - \beta' X X_i)$$

where $f_e(x) = \Pr(\varepsilon_e < x)$.

Given these assumptions, the value functions now have a tractable form. While these assumptions do bound the rationality of the student, they are key to allowing one to estimate a model with forward looking agents, a challenge which has not previously been overcome in the empirical network formation literature. Appendix 1 contains the full enumeration of the value functions used in estimation.

\(^8\)An alternative way of stating this assumption is that the student assumes her friends’ friends are evenly split between the high and low $e$. That is, $\bar{e}_j = 0.5$. 
5 Estimating the Model

5.1 Identification

5.1.1 Meeting Rates and Friend Preferences

A student who has many friends with similar attributes could be explained either by a preference for similar attributes or by an environment where that student meets others with similar attributes. In order to separate these two channels, I use data on both the numbers of friendships and the ranking students assign to these friendships. Assuming that who students have as friends does not affect how they rank their friends, friend orderings indicate how much students care about each trait forming friends. If the determinants of the extensive margin of friendship differ from those that determine the intensive margin, this method may fail. This could occur if, for instance, students learn about their preferences for traits of friends based on their experience with friends. Alternatively, preferences could be a function of who the student meets. While estimation of preferences would still be consistent, conclusions about counterfactual meeting rates could be incorrect.

Once the preferences are measured, the likelihoods generated from the network formation model described above can identify the meetings rates from the observed mixtures of friends. By estimating the formation model separately for each high school, I utilize only within high school variation. Identification requires that two individuals with the same friend preferences in the same high school, but who differ in their \( X_i \), have a different mix of friends (on average) only due to differential meeting rates. This assumption may fail if, for instance, there is an unobserved dimension that influences friendship formation but varies in expectation across the two observed types.

Note that \( \alpha_0 \), the marginal payoff to having an additional friend, is not identified from the ranking of friends. This parameter determines the number of meetings that result in friendships across all types (it is assumed constant). As a result, the number of meetings is not separately identified from \( \alpha_0 \). In the estimation results presented below, the number of meetings is normalized to be five times the population of observed students.
5.1.2 Peer Effects

Manski (1993) first outlined a traditional identification problem in linear peers effects models. The simultaneity and linearity of a linear-in-means peer effects model results in a “reflection problem” that prevents identification. However, more recent work (Brock and Durlauf (2001), for instance) has demonstrated that the non-linearities of discrete choice models render this issue moot. Furthermore, even in the linear case, Bramoullé et al. (2009) have shown that peer effects in a social network are identified with very weak restrictions on the network structure. Finally, Lee, Li, and Lin (2014) demonstrate that, under a condition discussed below, peer effects on a network with a binary choice are identified.

A complication that remains, however, is multiple equilibria. Brock and Durlauf (2001) and Lee, Li, and Lin (2014) show that if the magnitude of peer effects is limited, there is a unique rational expectations equilibrium. Since we cannot observe which rational expectations equilibrium was in play, we cannot evaluate the likelihood at each parameter value without an equilibrium selection mechanism. As a result, I restrict the friend effect such that a unique equilibrium exists: $\beta_e \in (-2, 2)^9$.

5.1.3 Additional Parameterizations

Because the model allows students to look forward to the incentives friends create in the choice of effort, an additional reason a student may be friends with others of a certain type is due to these peer effects. Because the peer effects are identified from separate estimation, their scale relative to the shock in the friendship formation process ($\delta$) must be estimated with meeting rates. In order to identify $\delta$, there must be two pairs of types who have the same meeting rate and who respond to $\delta$. (For instance, a pair of two of the same types do not respond to $\delta$ because their similarity leads to no peer effect.) This restricts slightly the flexibility of the meeting rate estimation, and if no two pairs of types

---

$^9$Multiple equilibria require the imposition of some equilibrium selection mechanism to perform counterfactuals. Since both identification and implementation of such a mechanism is challenging, I rule out these cases. In addition, estimation is complicated outside of this set since the fixed point which determines rational expectations may not exist (see the Estimation section for more details). Note that this assumption downwardly biases the counterfactual results since larger peer effects would only serve to magnify changes in the network structure.
have same meeting rates, δ is not identified. In practice, the meeting rates are restricted both to reduce the number of parameters estimated and to aid in the identification of δ. The weights on pairs of types meeting correspond to which dimensions they vary across. Given the set of parameters \( \{ \lambda_a \}_{a=1}^A \), the weights are

\[
\omega(X_i, X_j) = \exp \left( \sum_{a=1}^A \lambda_a \mathbf{1}\{x_{a,i} \neq x_{a,j}\} \right).
\]

In the data used for estimation, students name only five friends of each gender. Because of this data restriction and to limit the size of the state space in the value function calculations, the number of friends a student may have in the model is capped at five.\(^{10}\) The shocks in the network formation process \( \varepsilon_{f,i} \) are assumed to have a type I extreme value distribution.\(^{11}\) The shock in the discrete choice of \( e \) is assumed to be logistically distributed.

### 5.2 Estimation Implementation and Likelihood Functions

#### 5.2.1 Overview

Estimation proceeds in three stages, each leading respectively to the following sets of parameters: factors that determine the choice of effort, including friend effects; friend preferences (\( \alpha_1 \)); and meeting rates, along with \( \alpha_0 \) and \( \delta \). The first stage is estimated from data combined across all schools in the estimation sample. The second and third states derive estimates from each school entirely separately.

#### 5.2.2 Peer Effects

The estimation procedure for peer effects mirrors exactly that of Lee, Li, and Lin (2014). The method differs only slightly from that of Brock and Durlauf (2001), who consider the

---

\(^{10}\) In some cases, students may have more than five friends in the data since it is assumed that a nomination by either one of two students indicates a friendship for both. When this occurs, all friendships ranked (by either side) at or below the friendship that would exceed the limit are dropped.

\(^{11}\) In two cases, this distribution is approximated. First, the probabilities in the value functions of future friendships forming uses a fast approximation of the entire distribution which has an average error of less than a one thousandth of a percent and is accurate within a tenth of a percent over the entire distribution. The second case is the conditional expectation of \( \varepsilon_f \) in the value functions. Here, the integral is precalculated over a grid with spacing 0.01.
case where individuals respond to the mean of the full group, not just friends. In both cases, since students choose an optimal $e$ based on rational expectations of others’ choices of $e$, a necessary task, in addition to those of standard discrete choice estimation, is finding a consistent set of expectations given model parameters. For the model here, assuming a logistic error term, the expectation must obey

$$\psi = \frac{1}{1 + \exp(-\beta_0 - \beta'_X \bar{X} - \beta'_e \cdot [1 - 2\tilde{D}\psi])},$$

where $\psi$ is the vector of rational expectations for each student’s $e$ and $\tilde{D}$ is the row-normalized form of $D$.

Given a set of rational expectations, the likelihood is straightforward to calculate:

$$L(\beta) = \prod_{i=1}^{L} \left( \frac{1}{1 + \exp(-\beta_0 - \beta'_X \bar{X} - \beta'_e \cdot [1 - 2\tilde{D}\psi])} \right)^{e_i} \times \left( \frac{1}{1 + \exp(\beta_0 + \beta'_X \bar{X} + \beta'_e \cdot [1 - 2\tilde{D}\psi])} \right)^{1-e_i}.$$

For each guess of parameter values, rational expectations are calculated and the likelihood computed. This procedure is repeated until the likelihood is maximized. This likelihood is calculated across all schools in the estimation sample, holding fixed $\beta_X$, $\bar{X}$, and $\beta_e$. The parameter $\beta_0$, however, is allowed to vary across schools.

### 5.2.3 Friend Attribute Preferences

Friend preferences are estimated using a rank ordered logistic model (sometimes referred to as an exploding logit), first introduced by Beggs, Cardell, and Hausman (1981). The model assumes a series of alternatives with associated attributes and with (potentially partial) rankings for each individual. Utility is a linear combination of the attributes plus an error term distributed as a type I extreme value. The process that generates estimates from this model can be thought of as a series of multinomial logistic regressions for each individual, where each “regression” considers the probability of choosing each ranked alternative as best above those ranked below it.

The multinomial logistic nature of the estimation means that the fact that the rankings
are conditional on friendships does not impact the consistency of the estimates. Under logit’s I.I.A. assumption (an identical assumption is needed for the rank ordered logit), how a student ranks non-friends will not impact the rankings of friends. Thus, the students omitted from ranking do not affect estimates. Additionally, the fact that the shocks, $\varepsilon_f$, associated with friendships of those who are different from the student must be higher to offset the disutility from the differences is explicitly accounted for in the likelihood function. In a standard multinomial logit, observations where individuals choose a higher priced option, for example, must also have higher shocks on average, all else equal. In either case, the estimates ultimately reflect how often someone chooses a friend/option with a differing trait or high price over one with no difference or a low price.

Here, students rank their existing friends. Thus, the alternatives are each individual in the school. Attributes are absolute differences between a student’s scalar characteristic, $x_i^a$, and that of her friend, $x_j^a$. The likelihood function, modified slightly from Beggs, Cardell, and Hausman to account for the fact that students rank different sets of alternatives, is given by

$$L(\alpha_1) = \prod_{i=1}^{J_i} \prod_{j=1}^{J_i-1} \frac{\exp(\alpha_1 \cdot |X_i' - X_{r_j}^i|)}{\sum_{h=j}^{J_i} \exp(\alpha_1 \cdot |X_i' - X_{r_h}^i|)}$$

where $r_j^i$ is the index for $i$’s $j$th best friend out of the $J_i$ friends listed. This likelihood is maximized separately for each school.

### 5.2.4 Network Formation

Conditional on the set of meetings that took place, the probability of observing a given network $D$ is the product of independent probabilities that the network evolved as observed from one meeting to the next. Because a pair can only meet once, there is a unique path to any network given a set of meetings. The probability of an observed network can therefore be decomposed as follows:

$$\Pr(D|M, X, \omega, \alpha) = \prod_{n=1}^{N} \Pr(D_n|D_{n-1}, M, X, \omega, \alpha)$$

where $d_{ij}^n = 1$ if both $d_{ij} = 1$ and $\{i, j\} \in \{m_{n',1}, m_{n',2}\}_{n'=1}^n$. 

19
Now consider the choice of whether to add a friend. Recall that students compare the draw of \( \varepsilon_f \) to the difference in value functions. For a student who currently has friends, the probability she accepts is

\[
p^f_n(X_i, \bar{X}_i, q_i, X_j) \equiv \Pr \left( \varepsilon_f > V^f_{n+1}(X_i, \bar{X}_i, q_i) - V^f_{n+1}(X_i, \frac{q_i\bar{X}_i + X_j}{q_i + 1}, q_i + 1) \right)
\]

The probability of the network evolving from \( \mathbf{D}_{n-1} \) to \( \mathbf{D}_n \) is the probability that, for both \( m_{n,1} \) and \( m_{n,2} \), the shock was smaller than \( \phi \) if the friendship is observed (\( d_{ij} = 1 \)), or the complement of that probability if the friendship is not observed (\( d_{ij} = 0 \)). This can be written formally as

\[
\Pr(\mathbf{D}_n | \mathbf{D}_{n-1}, M, X, \omega, \alpha) = [p^f_n(X_i, \bar{X}_i, q_i, X_j) \cdot p^f_n(X_j, \bar{X}_j, q_j, X_i)]^{d_{ij}}
\times [1 - p^f_n(X_i, \bar{X}_i, q_i, X_j) \cdot p^f_n(X_j, \bar{X}_j, q_j, X_i)]^{1-d_{ij}}
\]

where \( i = m_{n,1} \) and \( j = m_{n,2} \). A similar probability for cases when one or both students have no friends can be derived using the appropriate value functions.

Next, the probability of a given set of meetings, conditional on \( N \), is

\[
\Pr(M | X, \omega, \alpha) = \Pr(M | \omega) = \prod_{n=1}^{N} \frac{\omega(t(m_{n,1}), t(m_{n,1}))}{W - \sum_{n'=1}^{n-1} \omega(t(m_{n',1}), t(m_{n',2}))}
\]

where

\[
W = \sum_{s=1}^{T} \sum_{s'=1}^{T} \left( \sum_{k=1}^{I} 1\{t_k = s\} \right) \left( \sum_{k=1}^{I} 1\{t_k = s'\} \right) - 1\{s = s'\} \right) \frac{\omega(s, s')}{2}.
\]

Finally, the likelihood of the observed network is a function of the probabilities above:

\[
\mathcal{L}(\omega, \alpha | D, X) = \Pr(D | X, \omega, \alpha) = \sum_{M \in \mathcal{M}} \Pr(D | M, X, \omega, \alpha) \Pr(M | X, \omega, \alpha)
\]

If the number of meetings \( N \) is fixed within a population of \( I \) students, there are \( \frac{(I+I(I-1)/2)!}{(T+I(I-1)/2-N)!} \) elements to the above sum. In a very conservative situation of only 25 students with only 5
percent of possible meetings taking place, there are approximately $10^{40}$ values to calculate and sum. If $N$ is unknown, the number is even higher. As a result, direct calculations are not feasible and simulation methods are used.

I base the estimation routine for the network formation parameters on that used in Christakis et al. (2010) and Goldsmith-Pinkham and Imbens (2013). This approach uses Markov Chain Monte Carlo (MCMC) methods, which are often employed in estimating network models as well as many other high-dimensional problems across multiple disciplines. This routine allows for sampling from the distribution of possible meetings and possible parameters in a way that moves through the distributions more efficiently than sampling at random from the entire space of potential values. This approach lends itself to Bayesian estimation of the parameters, which proves to be more straightforward and faster than attempting to maximize an approximation of the likelihood function or match simulated moments. Given a prior distribution for the model parameters, I use the MCMC routine to sample from the posterior.

To describe the process briefly, I first discretize the state dimension of $\bar{X}_i$ and use nearest neighbor interpolation when calculating values. I begin with an initial guess for the parameter values, and the routine draws a new set of parameter values from an arbitrary candidate distribution. The chain moves to the new set of parameter values with a probability equal to the ratio of the likelihood of the new parameter draw given the data to that of the original values. Next, a similar procedure occurs with $M$. Past analysis suggests an optimal rate at which new values are accepted, and so the variance of the candidate distributions must be adjusted to match those transition probabilities. This process of stepping through the relevant distributions is repeated until some criteria are met.

Estimation proceeds in two stages. In the first, a single MCMC chain updates a normally distributed prior for the parameters with mean zero and covariance equal to the identity matrix. After adjusting the transition rates appropriately, this chain proceeds for 2000 iterations. After removing the first 500 draws, the resulting posteriors are fit to a multivariate normal distribution. In the second stage, this fitted distribution serves as the prior. Five MCMC chains are run simultaneously starting from five unique random draws from the prior distributions with variance multiplied by 100 (to ensure that they are dispersed). After adjusting transition probabilities and discarding the first 500 steps, these five chains
continue until, for each parameter, the ratio of the total variance of draws across all chains to the mean of the variances of each chain is below 1.1, as suggested in previous work. The complete details of this process are given in Appendix 2.

6 Results and Counterfactuals

6.1 Sample Selection and Variable Definitions

The results below come from a subset of 16 medium-sized schools in the Add Health dataset. I exclude male students in this version of the estimation. The schools are large enough (> 100 girls) that the social networks are not too sparse for identification, and small enough < 250 girls) to permit timely estimation. Types with three or fewer students in a school are dropped, as are any students with missing data for variables used in the model. The vector of attributes used for estimation includes the following four binary variables: 11th or 12th grader; any parent completed college; white (race); ability. Ability is measured by regressing self-reported grades on self-reports of effort and TV watching, race, parental education, grade level, and PVT score for all students for whom these items are recorded. I then subtract the fitted values for all terms except ability (which is only measured for a subsample) from self-reported grades. This residual is then converted into a binary indicator by splitting the full sample at the median. Effort is measured by combining the self-reported effort and TV watching using the coefficients from the regression described above as weights. This measure is also converted into a binary outcome based on the student’s effort rank relative to the full sample median.

6.2 Estimates

The model described above is estimated separately for each school. Table (4) lists summary statistics for the estimated parameters across schools. Because the distributions of some parameters are quite skewed, means as well as medians are reported. The estimates suggest that student prefer friends from a similar grade, grade, and ability level (a negative parameter implies disutility from differing in that characteristic). Parental education does not appear to directly influence utility in the typical school. The estimates for race
vary most widely across schools, though a smaller observed number of interracial friendships (relative to other friend differences) could be the source of the variance. Figure (1) plots the distribution of each of the friend preference parameters across schools.

The meeting rate parameters can be interpreted as the percentage difference in the probability of meeting a student who differs in the respective trait but is similar in all others. The estimates suggest that in a typical school, a student is less likely to meet someone of same grade or parental education but is more likely to meet someone of a different race or ability level. This is consistent with classrooms being well mixed in terms of race and ability but less so in terms of grade-level (which is not surprising, of course) and parental education. The full distribution of these parameters across schools is shown in Figure (2). Finally, the estimate for \( \beta_e \), which is estimates simultaneously for all schools, is -1.31. That this parameter is below zero indicates disutility from not conforming to one’s friends choice of effort, which is consistent with intuition and past findings.

I next measure a potential link between the estimated meeting rates and observed classroom composition. As a measure of class segregation, I use the average deviation of the typical student’s classmates from the makeup observed in the school. Formally, the composition measure is

\[
\theta_s \equiv \frac{1}{I_s} \sum_{i=1}^{I_s} \sum_{t=1}^{T_s} \pi_{s,t} \left( 1 - \frac{\pi_{i,t}}{\pi_{s,t}} \right)^2
\]

where \( \pi_{s,t} \) and \( \pi_{i,t} \) are, respectively, the shares of school s’s population and of student i’s classmates who are type t. A \( \theta_s \) close to 0 indicates that the typical student at school s has classmates that look like the general population of the school, while a value closer to 1 indicates that a student’s classmates tend to differ from those in the school at large. Although students are likely to choose classes based on what classes their friends take, a student’s friends are likely to make up only a small portion of the statistic. Figure (3) shows the distribution of this measure across the schools in the estimation sample.

To measure the association between this statistic and the estimated meeting rates, I regress each of \( \{\lambda_{s,t}\}_{t=1}^{T_t} \) on \( \theta_s \). While the measured coefficients may be causal, if there are other unobserved school attributes which affect meetings rates and are correlated with \( \theta_s \), this interpretation is not correct. Table (5) presents these results. The small numbers of schools makes inference difficult. However, in one case I observe a significant correlation. Less
representative classrooms (higher tracking) is significantly associated with lower estimated meeting rate parameters for race. These results suggest that, as intended, the estimated meeting rates give some indication of the role of the institutional environment in friendship formation.

### 6.3 Additional Outcomes and Model Fit

To consider academic outcomes other than effort in school, I estimate a production function on the subsample for whom these additional outcomes are available. For a given outcome $y_i$, this function takes the form

$$y_i = \gamma_0 + \gamma_e e_i + \gamma' X_i + \gamma_\bar{e} \bar{e}_i + \gamma_{\bar{X}} \bar{X}_i + \eta_i$$

The parameters for this function are estimated by OLS, and the outcomes include high school graduation, high school GPA, and college attendance.

Table (6) shows several statistics across schools from both empirical data and data simulated from the estimated model. The model fits the academic outcomes very well. In Figure (4), the true and simulated distributions of effort across schools are plotted. Across the distribution, the model performs well in matching the observed data. However, the model predicts a lower number of friendships than are observed in the data. The model also fails to capture the observed clustering of friendships. While this model has no mechanism to generate such clustering, a similar model where friends of friends have higher meeting probabilities could do better in this dimension.

### 6.4 Counterfactuals

In the first series of counterfactuals, I examine the impact of changing how often students encounter one another. These alternative scenarios are suggestive of what might result from policies such as ability tracking that affect which types of students interact at school. I first equalize meeting rates so that any given student is equally likely to meet any other student in the school. This scenario is mostly useful for comparison relative to cases where similar students are either more or less likely to meet one another relative to different
students. Next, I repeat the exercise twice more using, respectively, the lowest estimated meeting rate parameters and the highest estimated meeting rate parameters.

Table (7) shows the resulting change in $e$ (academic effort) as well as the outcomes estimated separately (high school graduation, college attendance, and college graduation). In all three counterfactual cases, the outcomes change little in response to the different meetings rates. The most extreme difference is found in college attendance, where equal meetings rates lead to 1.5 percentage point increase in attendance. Elsewhere, differences tend to be on the order of a few tenths of a percentage point. To test whether more extreme meetings rates could lead to bigger differences in outcomes, I simulate the model with the most extreme meeting meeting rate parameters below zero exaggerated by a factor of 5 (very low) and by a factor 10 (very, very low). This is equivalent to making meetings between students of differing attributes less and less likely. Table (8) presents results from this exercise. Even with the extreme parameter values, very small effects are observed. College attendance responds the most, increasing 1.4 percentage points. All other differences are below 1 percentage point.

Additionally, I test the impact of shifting the composition of schools. In particular, I simulate students sorting into schools based on parental education – a situation that may mirror the opening of a charter school that draws students whose parents have the means to ensure their enrollment. To do this, I pair schools randomly and move as many of the high parental education students as possible to one of the schools randomly. I do so holding the size and the composition of the schools along other dimensions fixed. Changing the distribution of $X$ in a school will mechanically have non-social effects through each type’s base propensity to choose high or low effort (through $\beta_X$). To hold this constant, I change the distribution of $X_i$ only during the network formation process; for the choice of $e$, I use the school’s observed distribution of $X_i$.

Table (9) gives the outcomes under the alternative level of school-level heterogeneity. While the resulting achievement levels shift more than under the alternative meeting rates, the changes remain quite small. Furthermore, the results are mixed. High school graduation increases by 1.4 percentage points while college attendance falls by 0.6 percentage points.
7 Conclusion

Past evidence suggests that social interactions play a strong role in determining success in high school. Little attention has been given to the role that the high school environment may play in the formation of the ties through which those interaction operate. This paper demonstrates a connection in the data between the high school social environment and high school success that is consistent with the environment operating through friendship formation: Both classmates and friends with attributes that are linked to positive outcomes are associated with better individual outcomes. In addition, classmates’ attributes are strongly associated with the composition of students’ actual friends.

A structural model of friendship formation and social interactions allows for several additional levels of analysis. Aside from offering an alternative source of identification, the model separates meeting rates from friend preferences. The novel use of friend rankings allows the two channels to be separately identified. The model also accounts for school size and composition. Importantly, the general equilibrium effects of changes to the school environment are incorporated as students’ friendship formation depends both on past meetings and the forecast of future meetings. Allowing students to look forward during the network formation process is a novel contribution to the empirical network formation literature.

Finally, the estimated model is used to predict the changes in high school outcomes under alternative meetings rates and compositions of schools. The estimated impacts of these alternatives are consistently small, even when they are pushed to extremes. On one hand, these results are discouraging: Education policy cannot be easily used to influence outcomes through social networks. On the other hand, it is valuable to know that for the policies considered in this paper, social effects are unlikely to substantially offset other potential benefits from these policies, such as more directed curricula in ability tracked classrooms or greater competition among charter schools. To the extent that some levels of homophily could have negative consequences, this research also suggests that policies aimed at shifting preferences will be much more successful than those that only change who one encounters.
Appendix 1: Value Functions

First, let the function $\rho$ updates a student’s $\bar{X}$ when she makes a new friend:

$$\rho(\bar{X}, X, q) = \bar{X} \cdot q + X \over q + 1.$$ 

Given the assumptions placed on the students’ expectations, the state space $S$ can be collapsed to $\{X_i, \bar{X}_i, q_i\}$. The value functions that take a point in this state space as an argument are written with a tilde: $\tilde{V}$.

Students forecast the probability that they meet any other student in a given period as if no meetings had yet occurred. The probability of $i$ and $j$ meeting depends on their individual characteristics, $X_i$ and $X_j$:

$$p_m(X_i, X_j) = \frac{\omega(X_i, X_j)}{\sum_{k=1}^l \sum_{\ell=1, \ell \neq k} \omega(X_k, X_\ell)/2}.$$ 

where the denominator sums over the weights of all possible meetings.

The heuristic used by a student $i$ with $q_i > 0$ friends to predict the probability that a meeting with a student with attributes $X_j$ leads to friendship in period $n$, is

$$p_f^f(X_i, \bar{X}_i, q_i, X_j, n) = Pr(\varepsilon_f > \tilde{V}_{n+1}^f(X_i, \bar{X}_i, q_i) - \tilde{V}_{n+1}^f(X_i, \rho(\bar{X}_i, X_j, q_i), q_i + 1))$$

$$\cdot Pr(\varepsilon_f > \tilde{V}_{n+1}^f(X_j, X_j, g(n)) - \tilde{V}_{n+1}^f(X_j, \rho(X_j, X_j, g(n)), g(n) + 1),$$

where $g$ has been calibrated to $g(n) = \text{nint}(3n/N)$ and $\text{nint}(\cdot)$ rounds to the nearest integer.

The same probability for a student with no friends is

$$p_f^s(X_i, X_j, n) = Pr(\varepsilon_f > \tilde{V}_{n+1}^s(X_i) - \tilde{V}_{n+1}^f(X_i, X_j, 1))$$

$$\cdot Pr(\varepsilon_f > \tilde{V}_{n+1}^f(X_j, X_j, g(n)) - \tilde{V}_{n+1}^f(X_j, \rho(X_j, X_j, g(n)), g(n) + 1).$$

The expected value of $\varepsilon_f$ for student $i$ with $q_i > 0$ friends conditional on becoming friends with a student with attributes $X_j$ in period $n$ is

$$\varepsilon_f(X_i, \bar{X}_i, q_i, X_j, n) = E[\varepsilon_f | \varepsilon_f > \tilde{V}_{n+1}^f(X_i, \bar{X}_i, q_i) - \tilde{V}_{n+1}^f(X_i, \rho(\bar{X}_i, X_j, q_i), q_i + 1)].$$

For a student with no friends, the same conditional expectation is

$$\varepsilon_s(X_i, X_j, n) = E[\varepsilon_f | \varepsilon_f > \tilde{V}_{n+1}^s(X_i) - \tilde{V}_{n+1}^f(X_i, X_j, 1)].$$

Thus, the value function for students with at least one friend is

$$\tilde{V}_{n}^f(X_i, \bar{X}_i, q_i) = \left(1 - \sum_{j \neq i} p_f^f(X_i, X_j, n)\right) \tilde{V}_{n+1}^f(X_i, \bar{X}_i, q_i) + \sum_{j \neq i} p_m(X_i, X_j)$$

$$\times \left[ p_f^f(X_i, \bar{X}_i, q_i, X_j, n) \left( \tilde{V}_{n+1}^f(X_i, \rho(\bar{X}_i, X_j, q_i), q_i + 1) + \varepsilon_f(X_i, \bar{X}_i, q_i, X_j, n) \right) + (1 - p_f^f(X_i, \bar{X}_i, q_i, X_j, n)) \tilde{V}_{n+1}^f(X_i, \bar{X}_i, q_i) \right].$$
And the value function for students with no friends is

\[ \tilde{V}_n^s(X_i) = \left( 1 - \sum_{j \neq i}^I p_f^s(X_i, X_j, n) \right) \tilde{V}_{n+1}^s(X_i) + \sum_{j \neq i}^I p_m(X_i, X_j) \]

\[ \times \left[ p_f^s(X_i, X_j, n) (\tilde{V}_{n+1}^f(X_i, X_j, 1) + \epsilon_s(X_i, X_j, n)) + (1 - p_f^s(X_i, X_j, n)) \tilde{V}_{n+1}^s(X_i) \right]. \]

The values for period \( N + 1 \), when utility is realized, result from the network interactions described in the text. For a student with friends, this is

\[ \tilde{V}_{N+1}^f(X_i, \bar{X}_i, q_i) = \alpha_0 \cdot q_i + \alpha_1 \cdot |X_i - \bar{X}_i| \]

\[ + \delta \cdot \left[ (\beta_0 + \beta_X' X_i + \beta_{\bar{X}}' \bar{X}_i) \cdot \tilde{E}[e] + \beta_e \cdot |\tilde{E}[e] - \bar{E}[\hat{e}]| \right]. \]

For a student with no friends, this is

\[ V_{N+1}^s(X_i) = \max_{e_i} \{ \delta \cdot (\beta_0 + \beta_X' X_i) \cdot e_i \}. \]
Appendix 2: Markov Chain Monte Carlo Estimation

This estimation routine is largely based on Christakis et al. (2010).
Define the set of estimation parameters to be \( \Theta = \{ \} \)

1. A single initial chain is calculated.
   (a) An initial prior for the parameters is set to have density \( \Upsilon_0(\Theta) = N(0, \sigma_{\Upsilon,0}) \) where \( \sigma_{\Upsilon,0} \) is the identity matrix.
   (b) A candidate distribution variance for the parameters is defined to be \( \sigma_{c,0} = \sigma_{\Upsilon,0}/6 \).
   (c) A candidate transition parameter for the set of meetings is set to \( p_c = 0.01 \)
   (d) Initial values for all parameters, \( \Theta_0 \), are set to 0.
   (e) The value functions are solved, starting with the last period and moving to the first.
   (f) An initial set of meetings are generated from the observed friendships and a randomly selected set of meetings drawn with equal probability from the remaining list of meetings.
   (g) A test chain is calculated.
      i. An MCMC step in theta is taken.
         • Begin with the current set of parameters, \( \Theta_r \).
         • A candidate set of parameters \( \tilde{\Theta} \) are drawn from \( N(\Theta_r, \sigma_{c,0}) \).
         • The value functions are re-solved with \( \tilde{\Theta} \).
         • The following is calculated: \( \zeta_{\Theta} = \frac{Pr(D|M,X,\tilde{\Theta})\Upsilon_0(\tilde{\Theta})}{Pr(D|M,X,\Theta_r)\Upsilon_0(\Theta_r)} \).
         • If \( \zeta_{\Theta} > 1 \), \( \Theta_{r+1} = \tilde{\Theta} \). Otherwise, with probability \( \zeta_{\Theta} \), \( \Theta_{r+1} = \Theta_r \).
      ii. An MCMC step in meetings is taken.
         • Begin with the current set of meetings, \( M_r \).
         • Define \( r = \lceil p_c N \rceil \). Move \( r \) random meetings simultaneously to a new random position in the sequence of meetings. Replace \( r \) random meetings with an equally size set of meetings chosen randomly from the set of meeting not occurring. Again move \( r \) random meetings simultaneously to a new random position in the sequence of meetings. Define the new set of meetings to be \( \tilde{M} \).
         • The following is calculated: \( \zeta_{M} = \frac{Pr(D|M,X,\tilde{M})Pr(\tilde{M}|D,X,\Theta)}{Pr(D|M_r,X,\Theta)Pr(M_r|D,X,\Theta)} \).
         • If \( \zeta_{M} > 1 \), \( M_{r+1} = \tilde{M} \). Otherwise, with probability \( \zeta_{M} \), \( M_{r+1} = M_r \).
      iii. Steps (i) and (ii) are repeated 200 times.
(h) The probabilities that $\tilde{\Theta}$ and $\tilde{M}$ became the next step in the chain (the transition probabilities) are calculated.

(i) If both transition probabilities fall between 0.3 and 0.5, the program continues. Otherwise, $\sigma_{c,0}$ and $p_c$ are adjusted in proportion to how extreme the transition probabilities are, prior steps are discarded, and (g) is repeated with starting with the initial values set in (a)-(f).

(j) An additional 2300 steps are added to the chain as in step (g) for a total of 2500 steps. The first 500 are discarded.

2. A posterior distribution for $\Theta$ is calculated from the 2000 points in the initial chain by fitting a joint normal distribution via maximum likelihood. Call this $\Upsilon_p(\Theta)$.

3. Five distinct chains are calculated. For each:

   (a) The prior for the parameters is set to be $\Upsilon_p(\Theta)$.

   (b) A candidate distribution variance for the parameters is defined to be $\sigma_{c,p} = \sigma_{\Upsilon,p}$.

   (c) A candidate transition parameter for the set of meetings is set to $p_c = 0.01$.

   (d) An initial set of parameters are drawn from $\Upsilon_p(\Theta)$ but with the variance-covariance matrix multiplied by 100 (to ensure that the starting values are widely distributed).

   (e) Steps (1)(e)-(i) are repeated with the model parameters and transition parameters defined above.

   (f) The chain is continued for at least 2500 steps, discarding the first 500, and until the condition below is met.

4. All five chains are continued until for each scalar $\vartheta \in \Theta$, $\frac{\text{var}(\{\vartheta\}_{\text{chain}=1})}{\sum_{\nu=1}^5 \text{var}(\{\vartheta\}_{\text{chain}=\nu})/5}$. 

5. The pooled steps from all five chains are the posterior distribution and the source of the estimates reported in the text.
References


<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>5th Percentile</th>
<th>95th Percentile</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>0.59</td>
<td>0.31</td>
<td>0.00</td>
<td>0.03</td>
<td>0.933</td>
<td>1.00</td>
</tr>
<tr>
<td>Black</td>
<td>0.18</td>
<td>0.24</td>
<td>0.00</td>
<td>0.00</td>
<td>0.741</td>
<td>0.94</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.15</td>
<td>0.18</td>
<td>0.00</td>
<td>0.02</td>
<td>0.544</td>
<td>0.90</td>
</tr>
<tr>
<td>Asian</td>
<td>0.04</td>
<td>0.08</td>
<td>0.00</td>
<td>0.00</td>
<td>0.221</td>
<td>0.54</td>
</tr>
<tr>
<td>Mother Completed College</td>
<td>0.37</td>
<td>0.15</td>
<td>0.00</td>
<td>0.20</td>
<td>0.615</td>
<td>0.85</td>
</tr>
<tr>
<td>Father Completed College</td>
<td>0.43</td>
<td>0.17</td>
<td>0.16</td>
<td>0.21</td>
<td>0.750</td>
<td>1.00</td>
</tr>
<tr>
<td>Family Income ($10,000s)</td>
<td>0.04</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.073</td>
<td>0.18</td>
</tr>
<tr>
<td>PVT Score</td>
<td>0.49</td>
<td>0.13</td>
<td>0.01</td>
<td>0.27</td>
<td>0.664</td>
<td>0.81</td>
</tr>
<tr>
<td>HS Graduate</td>
<td>0.84</td>
<td>0.11</td>
<td>0.35</td>
<td>0.65</td>
<td>0.978</td>
<td>1.00</td>
</tr>
<tr>
<td>Attended College</td>
<td>0.30</td>
<td>0.17</td>
<td>0.00</td>
<td>0.07</td>
<td>0.603</td>
<td>0.86</td>
</tr>
<tr>
<td>Number of Students in School</td>
<td>515</td>
<td>422</td>
<td>6</td>
<td>46</td>
<td>1292</td>
<td>2414</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS Grade Average PVT</td>
<td>-0.0281</td>
<td>-0.345**</td>
<td>-0.158**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0622)</td>
<td>(0.158)</td>
<td>(0.0775)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS Grade Fraction w/</td>
<td>0.163*</td>
<td>0.112</td>
<td>-0.0588</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS Graduate Mother</td>
<td>(0.0975)</td>
<td>(0.248)</td>
<td>(0.0869)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS Grade Fraction w/</td>
<td>0.00908</td>
<td>0.195</td>
<td>0.0906</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College Graduate Father</td>
<td>(0.0506)</td>
<td>(0.136)</td>
<td>(0.0683)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS Grade Fraction w/</td>
<td>-0.110**</td>
<td>-0.414***</td>
<td>-0.0743</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Parent</td>
<td>(0.0523)</td>
<td>(0.158)</td>
<td>(0.0685)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>15174</td>
<td>11821</td>
<td>15175</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < 0.10; ** p < 0.05; *** p < 0.01

This table reports estimates from four linear regressions. HS Graduation is a dummy equal to 1 if the student graduated from high school. HS GPA is measured on a 4.0 scale. College Attendance is a dummy equal to 1 if the student attended any college. Income is measured in 2008 dollars. Fixed effects included: school, grade. Student-level regressors: race, sex, parental education, parental income, parents in household, PVT score (percentile/100), missing data indicators. Grade-level regressors not shown: average years of mothers’ education, average years of fathers’ education, fraction with high school graduate fathers, fraction with college graduate mothers, average family income, number of grademates. Standard errors clustered at school level.
Table 2.2: Association Between Own Type Grademate Characteristics and Human Capital Outcomes (HS by Type Fixed Effects)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HS Graduate</td>
<td>HS GPA</td>
<td>Attended College</td>
</tr>
<tr>
<td>HS Grade Own Type</td>
<td>0.0393</td>
<td>0.00394</td>
<td>0.00762</td>
</tr>
<tr>
<td>Average PVT</td>
<td>(0.0418)</td>
<td>(0.101)</td>
<td>(0.0570)</td>
</tr>
<tr>
<td></td>
<td>0.207***</td>
<td>0.290*</td>
<td>0.0669</td>
</tr>
<tr>
<td>Fraction w/ HS Graduate Mother</td>
<td>(0.0587)</td>
<td>(0.158)</td>
<td>(0.0634)</td>
</tr>
<tr>
<td></td>
<td>-0.0250</td>
<td>0.125</td>
<td>0.0818*</td>
</tr>
<tr>
<td>Fraction w/ College Graduate Father</td>
<td>(0.0288)</td>
<td>(0.0819)</td>
<td>(0.0422)</td>
</tr>
<tr>
<td></td>
<td>-0.0808***</td>
<td>-0.183**</td>
<td>-0.0869**</td>
</tr>
<tr>
<td>Fraction w/ Single Parent</td>
<td>(0.0310)</td>
<td>(0.0922)</td>
<td>(0.0435)</td>
</tr>
<tr>
<td>Observations</td>
<td>14798</td>
<td>11533</td>
<td>14799</td>
</tr>
</tbody>
</table>

* p < 0.10; ** p < 0.05; *** p < 0.01

This table reports estimates from four linear regressions. HS Graduation is a dummy equal to 1 if the student graduated from high school. HS GPA is measured on a 4.0 scale. College Attendance is a dummy equal to 1 if the student attended any college. Income is measured in 2008 dollars. Types are defined by the combination of race and sex. Fixed effects included: school by type, grade. Student-level regressors: parental education, parental income, parents in household, PVT score (percentile/100), missing data indicators. Grade by type regressors not shown: average years of mothers’ education, average years of fathers’ education, fraction with high school graduate fathers, fraction with college graduate mothers, average family income, number of own type grademates. Standard errors clustered at school-type level.
Table 3: Association Between Own Type Grademate Characteristics and Human Capital Outcomes, Controlling for Friends

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HS Graduate</td>
<td>HS GPA</td>
<td>Attended College</td>
</tr>
<tr>
<td>HS Grade Own Type</td>
<td>0.0353</td>
<td>-0.153</td>
<td>0.0741</td>
</tr>
<tr>
<td>Average PVT</td>
<td>(0.0621)</td>
<td>(0.191)</td>
<td>(0.109)</td>
</tr>
<tr>
<td>HS Grade Own Type</td>
<td>0.121</td>
<td>0.416</td>
<td>0.0721</td>
</tr>
<tr>
<td>Fraction w/ HS Graduate Mother</td>
<td>(0.0913)</td>
<td>(0.261)</td>
<td>(0.124)</td>
</tr>
<tr>
<td>HS Grade Own Type</td>
<td>-0.0482</td>
<td>0.181</td>
<td>0.0541</td>
</tr>
<tr>
<td>Fraction w/ College Graduate Father</td>
<td>(0.0457)</td>
<td>(0.130)</td>
<td>(0.0877)</td>
</tr>
<tr>
<td>HS Grade Own Type</td>
<td>-0.0113</td>
<td>-0.125</td>
<td>-0.111</td>
</tr>
<tr>
<td>Fraction w/ Single Parent</td>
<td>(0.0451)</td>
<td>(0.147)</td>
<td>(0.0883)</td>
</tr>
<tr>
<td>Friends Average PVT</td>
<td>0.114***</td>
<td>0.835***</td>
<td>0.339***</td>
</tr>
<tr>
<td></td>
<td>(0.0188)</td>
<td>(0.0559)</td>
<td>(0.0296)</td>
</tr>
<tr>
<td>Friends Fraction w/</td>
<td>0.0106</td>
<td>-0.284*</td>
<td>-0.0816</td>
</tr>
<tr>
<td>HS Graduate Mother</td>
<td>(0.0986)</td>
<td>(0.167)</td>
<td>(0.0582)</td>
</tr>
<tr>
<td>Friends Fraction w/</td>
<td>0.0271**</td>
<td>0.110***</td>
<td>0.0975***</td>
</tr>
<tr>
<td>College Graduate Father</td>
<td>(0.0110)</td>
<td>(0.0285)</td>
<td>(0.0181)</td>
</tr>
<tr>
<td>Friends Fraction w/</td>
<td>0.0374</td>
<td>-0.128</td>
<td>0.0321</td>
</tr>
<tr>
<td>Single Parent</td>
<td>(0.0922)</td>
<td>(0.133)</td>
<td>(0.0503)</td>
</tr>
<tr>
<td>Observations</td>
<td>5673</td>
<td>4624</td>
<td>5673</td>
</tr>
</tbody>
</table>

* p < 0.10; ** p < 0.05; *** p < 0.01
This table reports estimates from four linear regressions. The specifications are identical to those in Table (3) but with the addition of each grade-type measure for the student’s friends.
### Table 4: Estimates Across Schools, Population Weighted

<table>
<thead>
<tr>
<th></th>
<th>Median</th>
<th>Mean</th>
<th>S.D.</th>
<th>Fraction Different from 0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Friend Preferences ($\alpha_1$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td>-0.562</td>
<td>-0.428</td>
<td>0.485</td>
<td>0.647</td>
</tr>
<tr>
<td>Parental Education</td>
<td>0.0613</td>
<td>0.00506</td>
<td>0.21</td>
<td>0.686</td>
</tr>
<tr>
<td>White</td>
<td>-0.337</td>
<td>-0.334</td>
<td>0.633</td>
<td>0.392</td>
</tr>
<tr>
<td>Ability</td>
<td>-0.128</td>
<td>-0.0465</td>
<td>0.225</td>
<td>0.614</td>
</tr>
<tr>
<td><strong>Meeting Rate Parameters ($\lambda/\omega$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td>-0.196</td>
<td>0.308</td>
<td>1.23</td>
<td>0.375</td>
</tr>
<tr>
<td>Parental Education</td>
<td>-0.162</td>
<td>0.755</td>
<td>1.73</td>
<td>0.5</td>
</tr>
<tr>
<td>White</td>
<td>0.991</td>
<td>6.67</td>
<td>9.19</td>
<td>0.25</td>
</tr>
<tr>
<td>Ability</td>
<td>0.301</td>
<td>1.04</td>
<td>2.28</td>
<td>0.188</td>
</tr>
<tr>
<td><strong>Weight on $e$ Utility ($\delta$)</strong></td>
<td>1.41</td>
<td>1.65</td>
<td>0.673</td>
<td>1</td>
</tr>
<tr>
<td><strong>Friend $e$ Utility ($\beta_e$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-1.31</td>
<td>(2.13)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table lists statistics from the distribution of estimates across schools, weighted by the number of students in each school (most parameters are estimated independently for each school). The parameter $\beta_e$ is estimated simultaneously and thus a single estimate and standard error are reported above.
Table 5: Linear Regression, Meeting Rate Parameters on Measure of Tracking

<table>
<thead>
<tr>
<th></th>
<th>$\lambda_{\text{Grade}}$</th>
<th>$\lambda_{\text{Parental Education}}$</th>
<th>$\lambda_{\text{White}}$</th>
<th>$\lambda_{\text{Ability}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.01</td>
<td>-0.952</td>
<td>29.7</td>
<td>3.19</td>
</tr>
<tr>
<td></td>
<td>(1.21)</td>
<td>(1.75)</td>
<td>(9.77)</td>
<td>(1.99)</td>
</tr>
<tr>
<td>Tracking</td>
<td>6.25</td>
<td>8.82</td>
<td>-110</td>
<td>-11.5</td>
</tr>
<tr>
<td></td>
<td>(5.87)</td>
<td>(8.52)</td>
<td>(43.2)</td>
<td>(9.68)</td>
</tr>
</tbody>
</table>

This table shows results from four regressions of each of the estimated meeting rate parameters on the measure of tracking described in the text.
Table 6: Observed and Simulated Data Across Schools, Population Weighted

<table>
<thead>
<tr>
<th></th>
<th>Observed Data</th>
<th>Simulated Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>Effort (e)</td>
<td>0.508</td>
<td>0.0452</td>
</tr>
<tr>
<td>High School Graduation</td>
<td>0.881</td>
<td>0.0986</td>
</tr>
<tr>
<td>High School GPA</td>
<td>2.83</td>
<td>0.291</td>
</tr>
<tr>
<td>College Attendance</td>
<td>0.387</td>
<td>0.178</td>
</tr>
<tr>
<td>Number of Friends</td>
<td>2.64</td>
<td>0.685</td>
</tr>
<tr>
<td>Clustering Coefficient</td>
<td>0.232</td>
<td>0.0773</td>
</tr>
</tbody>
</table>

This table lists moments from the observed data and from data simulated using the estimated parameter values. In both cases, statistics are weighted by the number of students in each school.
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effort (e)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>0.513</td>
<td>0.0454</td>
</tr>
<tr>
<td>Equal Meeting Rates</td>
<td>0.515</td>
<td>0.0458</td>
</tr>
<tr>
<td>Lowest Meeting Rate Estimates</td>
<td>0.515</td>
<td>0.0456</td>
</tr>
<tr>
<td>Highest Meeting Rate Estimates</td>
<td>0.515</td>
<td>0.0435</td>
</tr>
<tr>
<td><strong>High School Graduation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>0.871</td>
<td>0.0951</td>
</tr>
<tr>
<td>Equal Meeting Rates</td>
<td>0.876</td>
<td>0.106</td>
</tr>
<tr>
<td>Lowest Meeting Rate Estimates</td>
<td>0.871</td>
<td>0.115</td>
</tr>
<tr>
<td>Highest Meeting Rate Estimates</td>
<td>0.872</td>
<td>0.127</td>
</tr>
<tr>
<td><strong>High School GPA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>2.74</td>
<td>0.31</td>
</tr>
<tr>
<td>Equal Meeting Rates</td>
<td>2.78</td>
<td>0.291</td>
</tr>
<tr>
<td>Lowest Meeting Rate Estimates</td>
<td>2.78</td>
<td>0.29</td>
</tr>
<tr>
<td>Highest Meeting Rate Estimates</td>
<td>2.75</td>
<td>0.326</td>
</tr>
<tr>
<td><strong>College Attendance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>0.384</td>
<td>0.2</td>
</tr>
<tr>
<td>Equal Meeting Rates</td>
<td>0.401</td>
<td>0.205</td>
</tr>
<tr>
<td>Lowest Meeting Rate Estimates</td>
<td>0.397</td>
<td>0.208</td>
</tr>
<tr>
<td>Highest Meeting Rate Estimates</td>
<td>0.39</td>
<td>0.214</td>
</tr>
</tbody>
</table>

This table lists moments for several outcomes across schools for the estimated parameters (Baseline) and several counterfactual meetings rates described in the text. In all cases, the statistics are weighted by the number of students in each school.
### Tables 8: Outcomes Across Schools for Extreme Meetings Rates, Population Weighted

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effort (e)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>0.513</td>
<td>0.0454</td>
</tr>
<tr>
<td>Very Low Meeting Rate Parameters</td>
<td>0.515</td>
<td>0.0441</td>
</tr>
<tr>
<td>Very, Very Low Meeting Rate Parameters</td>
<td>0.515</td>
<td>0.0444</td>
</tr>
<tr>
<td><strong>High School Graduation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>0.871</td>
<td>0.0951</td>
</tr>
<tr>
<td>Very Low Meeting Rate Parameters</td>
<td>0.872</td>
<td>0.11</td>
</tr>
<tr>
<td>Very, Very Low Meeting Rate Parameters</td>
<td>0.871</td>
<td>0.113</td>
</tr>
<tr>
<td><strong>High School GPA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>2.74</td>
<td>0.31</td>
</tr>
<tr>
<td>Very Low Meeting Rate Parameters</td>
<td>2.80</td>
<td>0.27</td>
</tr>
<tr>
<td>Very, Very Low Meeting Rate Parameters</td>
<td>2.80</td>
<td>0.277</td>
</tr>
<tr>
<td><strong>College Attendance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>0.384</td>
<td>0.2</td>
</tr>
<tr>
<td>Very Low Meeting Rate Parameters</td>
<td>0.398</td>
<td>0.205</td>
</tr>
<tr>
<td>Very, Very Low Meeting Rate Parameters</td>
<td>0.398</td>
<td>0.204</td>
</tr>
</tbody>
</table>

This tables lists moments for several outcomes across schools for the estimated parameters (Baseline) and several counterfactual meetings rates described in the text. In all cases, the statistics are weighted by the number of students in each school.
Table 9: Outcomes Across Schools for Counterfactual Sorting Scenario, Population Weighted

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effort ($e$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>0.513</td>
<td>0.0454</td>
</tr>
<tr>
<td>Sorting</td>
<td>0.508</td>
<td>0.0432</td>
</tr>
<tr>
<td>High School Graduation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>0.871</td>
<td>0.0951</td>
</tr>
<tr>
<td>Sorting</td>
<td>0.885</td>
<td>0.0645</td>
</tr>
<tr>
<td>High School GPA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>2.74</td>
<td>0.31</td>
</tr>
<tr>
<td>Sorting</td>
<td>2.72</td>
<td>0.306</td>
</tr>
<tr>
<td>College Attendance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>0.384</td>
<td>0.2</td>
</tr>
<tr>
<td>Sorting</td>
<td>0.378</td>
<td>0.21</td>
</tr>
</tbody>
</table>

This table lists moments for several outcomes across schools for the estimated parameters (Baseline) and a counterfactual where students sort into schools as described in the text. In all cases, the statistics are weighted by the number of students in each school.
Figure 1

Distribution of Friend Preference Parameter Values Across Schools, Population Weighted

Figure 2

Distribution of Meeting Rate Parameter Values (Normalized) Across Schools, Population Weighted
Figure 3

Measure of Tracking Across Schools, Population Weighted

Figure 4

Observed and Simulated $e$ Across Schools, Population Weighted

Observed
Simulated