Do students want to succeed? Peer group choice, social influence, and undergraduate performance

Gigi Foster*, University of South Australia

December 30, 2006

*Working Paper** - Please Do Not Quote

Abstract

Students routinely choose others to partner with in studying or discussing course material. The influence of study groups on their members may differ depending upon the student-level preferences expressed in this choice. This study searches for evidence of such heterogeneity by exploiting a unique classroom experiment in study group assignment within a University of South Australia introductory economics course. Drawing a conceptual framework from the treatment effects literature, I decompose the peer effect and measure the strength and nature

*I thank Paul Frijters, Andrew Leigh, Chikako Yamauchi, two anonymous referees, and seminar participants in the economics group of the Research School of Social Sciences of the Australian National University, and in the Department of Economics and Finance at the Queensland University of Technology, for helpful comments. I also acknowledge UniSA's School of Commerce and Centre for Regulation and Market Analysis for support over the course of this project. All viewpoints expressed within the paper and errors that remain are solely my own.

of selection versus other determinants of peer group composition and peer group influence. I then use the expected returns yielded by this process to judge whether students choose to sort into peer groups in a way that could be consistent with differential returns to peer ability.

1 Introduction

In a variety of settings, individuals are free to make choices that influence their economic and social outcomes at a later stage. In the context of public policy, the existence of free choice has long been seen by researchers as an impediment to evaluation (see Heckman & Robb (1986), Meyer (1995), and Angrist, Imbens & Rubin (1996), among many others). Even in an experimental or quasi-experimental setting, the classic treatment effects presentation explains imperfect compliance with a randomized treatment state as a explicit threat to the unbiasedness of treatment effect estimation. Choice—also known, more broadly, as selection—is perceived to be a problem.

Whether or not imperfect compliance is managed adequately within a sample, decisions that might be informed by treatment effect estimation are typically made without access to empirically validated selection equations. This problem forms the core of the ongoing investigation into treatment effect estimation in the public economics literature. Indeed, founding contributors (Heckman & Robb 1986) to this literature admitted as much 20 years ago (p. 106, emphasis added):

> Until these implicit behavioral assumptions [underlying the selection decision] are made explicit and the limitations of mathematical statistics are clearly recognized, there will be no convergence in views on the validity, and limits, or competing approaches to the selection bias problem. *The "solu-*

tion" to the selection bias problem lies outside of formal statistics.

A solution outside of formal statistics has not, to the author's knowledge, been attempted in economics. While a few recent authors have developed novel and exciting ways to measure the extent of selection (e.g., Echenique & Fryer (2005)) or have taken pains to explicitly defend their assumptions regarding the extent of selection (e.g., Altonji, Elder & Taber (2005), Krauth (2005)), little work exists that examines the underlying behavioral determinants of selection. What has been done in economics (e.g., Sacerdote & Marmaros (2006), Foster (2005)) has not linked selection processes to outcomes.

The selection problem is widely believed to plague the unbiased estimation of peer effects, via endogenous peer group selection. As in the treatment effects literature, selection has been viewed in this context as a problem to be overcome through randomization, instrumentation, or other clever statistical tricks. Yet the peer effect relevant to a random-assignment context may differ from that applicable to a context with self-selection. Only a few papers (e.g., Altonji et al. (2005), Arcidiacono, Foster, Goodpaster & Kinsler (2005)) have realized some of the extent to which the rich panel data sets that are now regularly obtained to estimate peer effects can be exploited to explore the nature of the selection problem. Few papers have attempted to ascertain the nature of peer effect heterogeneity beyond crude interactions of observable characteristics with the main peer effect, and in spite of indications of remarkable heterogeneity from such exercises, no unifying explanation has been offered.

In this paper, I use a panel of information on students' actual social choices and educational performance to explore the importance and nature of selection when estimating the returns to peer ability. I first examine whether the returns to randomly-derived versus self-selected peer ability are identical. I then allow for different returns to peers for different students, and ask whether free choice results in a learning environment where students whose returns to peers are the greatest are observed in the highest-ability peer groups. The primary technique employed is the empirical specification of a model of the undergraduate selection mechanism within classrooms exposed to a unique experimental design; this is coupled with a policy-relevant decomposition of the selection that is observed. I argue that the free-choice complication can, if properly understood, be used as a tool to guide educational policymakers in tailoring their policies to suit the heterogeneous needs and proclivities of students.

As pointed out by Hoxby & Weingarth (2006), if peer effects are in fact linear in means (as often assumed in the peer effects literature), then no policy recommendation can be made regarding student mix that is based solely on efficiency. Only distributional concerns can guide policy in this area if it is indeed the case that every student benefits equally from peers with a fixed average ability.[1] For this reason, exploring whether

[1]While Hoxby & Weingarth (2006) imply that observed patterns of ability sorting are inexplicable if peer effects are linear in means, this is not the case. Even if no concern motivated an individual's choice of peers except the impact that he anticipated they would have on his outcomes, peer effects that were linear in means could generate ability-based sorting because of the inherent scarcity of high-achieving peers. If social partnering works as a typical market does, and information regarding peer influence is relevant to students' selection decisions, then under the linear-in-means model, the highest achievers enjoy the greatest demand for their influence—and will therefore be able to exert the greatest power in selecting peers for themselves. Rationally, they will choose the next-highest achievers in order to maximize the social influence of which they themselves may partake. Low-ability individuals have little power, due to lack of demand for them from others, and hence end up in groups with each other despite their preferences not to be with one another due to the comparatively less beneficial peer effects that are generated in low-ability groups. This could account what appears to be same-seeking selection by ability in any given population, and it is accommodated conceptually in this paper through the use of

peer effects really are homogeneous is potentially policy-relevant. Doing so with the treatment effect framework to motivate the model of selection and the search for peer effect heterogeneity provides a degree of conceptual structure that is rarely brought to the peer effects question.

2 Conceptual foundation

In the classic linear treatment effect framework with discrete treatment status, an individual's outcome of interest is assumed to be a function of his attributes, his treatment status, and an error term:[2]

$$Y_i = X_i\beta + d_i\alpha_i + U_i \tag{1}$$

where d_i is 1 for individuals exposed to treatment, and 0 for those not exposed, and for generality, the treatment effect α is allowed to be heterogeneous across individuals. Heckman & Robb (1986) bifurcate the treatment effect α_i into two components, one of which ($\bar{\alpha}$) is common across all individuals, and one of which (ϵ_i)varies by individual:

$$Y_i = X_i\beta + d_i\bar{\alpha} + U_i + d_i\epsilon_i \tag{2}$$

In estimation, the sum of U_i and $d_i\epsilon_i$ forms the error. A selection problem plaguing the estimation of $\bar{\alpha}$ may then result either from an expected value of U_i that depends on the value of d_i, or from an expected value of ϵ_i that depends on d_i.[3] The levels

a fixed effect capturing both preferences and social power.

[2]For a full exposition of the classic framework and statistical subtleties within it, readers may refer to any of the numerous papers by Heckman, Robb, Angrist, Imbens, and Rubin, as listed in the reference list of this paper, or to their antecedants.

[3]In the latter case, any such selection problem would result from imperfect modeling of the selection rule in pursuit of hurdling the inherent problem of non-identification of $\bar{\alpha}$ if no selection rule is specified.

of these dependencies vary according to an underlying behavioral selection equation which economists have not attempted to specify to date, to the author's knowledge, in anything but a manner motivated by statistical logic.[4]

More recent literature in the field of treatment effect estimation has shown how the effects of continuous treatments on economic outcomes can be modelled through an extension of the assumption that selection occurs only with respect to observable characteristics.[5] While the extension to continuous treatments is valuable, this recent literature (like its antecedents) is concerned primarily with statistical solutions to the potential problem of systematic differences in treatment status that are correlated with outcomes, rather than with ascertaining the extent to which these differences are evident and the association that they actually have with outcomes.

In the present paper, I attempt to examine these questions directly, in the context of peer effects. The standard linear-in-means reduced-form model of peer effects can be written as

$$outcome_i = X_i\beta' + peers_i\gamma_i + U_i \tag{3}$$

where γ_i represents the effect of the (continuous) average of peers' attributes to which student i is exposed ($peers_i$) on that student's outcome. The endogeneity concern that garners researchers' attention is the potential for correlation between $peers_i$ and U_i, due to students' selection of peer groups. However, the observed peer group is not only potentially a direct function of individuals' preferences (hence the concern, to the extent

No such problem exists in the case of estimating the treatment effect on the treated, but this latter effect must be generalized to the population at large using an assumption about future selection rules if one wishes to use it to predict policy outcomes.

[4]A classic example of this is the argument presented in Altonji et al. (2005).

[5]For example, Hirano & Imbens (2004) extend the propensity score conditioning technique to the case of continuous treatments.

that such preferences are correlated with outcomes) but also likely to be only related to—not the same as—the peer group that would perfectly suit those preferences. This may be due to both external and internal factors. Prototypical examples of each of these sources of influence, respectively, would be the nature of the available pool of students from whom to choose peers, and the individual's own level of assertiveness or passivity in forming groups. If students respond heterogeneously to the peers they choose and the peers they don't choose, as has been suggested by Foster (2006), then artificially eliminating the correlation between $peers_i$ and U_i (for instance, through randomization) will not necessarily permit the recovery of the peer effect relevant to policy settings which feature peer group selection.

To investigate the nature of the conventional endogeneity problem in more detail, we can write Equation 3 to follow the treatment effects framework:

$$outcome_i = X_i\beta' + peers_i\bar{\gamma} + U_i + peers_i\epsilon_i \tag{4}$$

where $\epsilon_i = \gamma_i - \bar{\gamma}$, the component of the treatment effect that is not common to all individuals. Again here, the error term in estimation is $U_i + peers_i\epsilon_i$. It is clear in this formulation that not only the level, but the effect on outcomes, of peer attributes may be heterogeneous across individuals, and the two may be correlated. In particular, the portion of the peer effect applicable to individual i, ϵ_i, may be correlated with $peers_i$. This would occur if those students most influenced by their peers chose peer groups with the most beneficial attributes—the analog of the case where the most needy individuals in a population seek job training.[6] There are therefore two possible routes to endogeneity plaguing peer effect estimation: first, students with good unobservables may

[6] An important distinction should be drawn here between a student for whom the peer effect is greater than any other student, and a student for whom the "best" peers will help more than any other peers. From an individual perspective, each student may very well want to choose high-ability peers,

be observed disproportionately with observably good peers; and second, if returns to peers are heterogeneous across individuals, students whose proclivities to be influenced by peers are the strongest across all students may be observed disproportionately with observably good peers.

In order to progress both empirically and conceptually, we require a second equation describing how exactly peer groups form (the analog of a Heckman-style selection equation). Suppose an empirical context in which students, indexed by i, sort into small peer groups within a classroom. Every student i has a certain broad group j (all students in the classroom) from which to choose his peer group k. We can then write down an underlying peer group selection equation such as

$$peers_{ijk} = \beta_0 + broadpeers_{ij}\delta + peers_i^*\omega_i + \mu_{ijk} \tag{5}$$

where β_0 is each student's "expected" peer group composition, in the absence of increases or decreases in response to either internal or external factors, which is mechanically determined by the distribution of attributes over the entire sample; $broadpeers_{ij}\delta$ represents the component of peers' attributes to which the individual is exposed that results mechanically from the mix of available peers in broad group j (a subset of the entire sample); $peers_i^*$ is the individual's latent first-best peer group composition preference; and ω_i is the individual-specific modifier of this preference capturing the extent to which the individual is socially assertive or passive relative to other group members in putting his will regarding the composition of his own peer group into practice.[7] This

since (compared to other choices) these peers will help them more. The test in this paper is whether those students who are, compared to all other students, *most* affected by their peers are also observed to choose the best peers.

[7]A departure from unity in this parameter, in either direction, would indicate passivity or low social power. Because the formation of social partnerships requires the consent of all members, some students

equation assumes as environment where students choose a nuclear group (e.g., a group of friends or study partners) from amongst a larger group to which they are arguably exogenously assigned(e.g., a class or section) of a population of interest that is randomly sampled (e.g., a school or university course).[8]

Importantly, the empirical identification of both person-specific effects—$peers_i^*$ and ω_i—is infeasible unless the researcher can access a panel data set with some exogenous variation in either social power or underlying preferences within individuals. The present paper exploits a data set richer than most, but no obvious candidate for either of these sources of variation presents itself. Therefore, the two effects are estimated in aggregate using a fixed effects framework, and remain indistinguishable in this paper. The implications are that we will not know, for example, whether a high fixed effect for a given student means that he has an underlying preference to be with high-ability peers, or that he is relatively powerful in satisfying those preferences, or both. A high fixed effect could also mean that the student in fact prefers to be with low-ability peers, but is passive in making that preference known.

The key academic contribution of the paper is to suggest the estimation of a selection equation underlying peer effects, in order to better inform our understanding

whose preferences are to be with top-notch peers may be shunted to lower-quality peer groups only because those top-notch peers do not want them. There is no exact analog to this feature of the problem in the treatment effects framework, although it is somewhat similar to the potential denial of services to applicants due to program budget constraints.

[8]The exercise of estimating a selection equation underlying a determinant of outcomes is also coherent with the concern regarding "ability bias" on estimated returns to own education, due to unobservably more able individuals disproportionately choosing to acquire schooling. See Angrist & Krueger (1991), Card (2001) and similar literature for a review of the concept of omitted ability bias in estimating the wage returns to own education.

of how both selection and peer effects operate in the undergraduate educational setting. This understanding then informs a consideration of both randomized and self-selected educational settings, and is used to test whether observed student sorting may be socially optimal from the perspective of peer effects on learning. The key policy goal of the paper is to derive information relevant to educational policymakers regarding the assignment of students of heterogeneous ability into learning groups.

3 Data

The University of South Australia (UniSA) is a public university with approximately 29,000 students and 2,000 staff, whose main campuses are located in Adelaide, South Australia. The Division of Business represents approximately one-third of the university. Within this division, all students (each of whom enrolls in a particular, typically 3-year, program of study) undertake a core of business-related courses in their first year before continuing on to more specialized courses. One course of eight taught to full-time students in their first year is Economic Principles. This course is taken by most students in their first semester, although it can also be taken in the second year by students in some programs. It is essentially a first-year microeconomics course whose principal text is an Australian adaptation of the introductory American textbook by McConnell.

Delivery of this course to local students consists of two hours of lectures and a weekly 1.5-hour tutorial. Given that more than 800 students typically enroll in this course in the first semester of every academic year, small group work is emphasized in tutorials, with typically 2 to 4 students per group and the total number of students per tutorial around 25. During the semester in which the data for this study were collected, there were 30 tutorials, taught by 11 different tutors, and 661 students observed to attend at least one of the five observed tutorials during the semester.[9]

There are several components of formative and summative assessment in the course. Individually, students must complete, for credit towards their final course grade, an economics-specific online questionnaire entitled “Reflections on Learning Inventory” (RoLI), which collects an array of background and preference data and is designed to help students become aware of and improve their learning style and capacity; they also must individually complete a second online questionnaire, which is a version of the standard Test of Economic Literacy.[10] The RoLI and TEL are both completed in the first two weeks and again in the final two weeks of the 13-week semester. The final piece of individual assessment is an end-of-semester closed-book examination.

Groupwork-based assessment completed individually in class, which is the main subject of the present paper, consists of true/false tutorial quizzes of ten questions apiece administered throughout the semester, for which preparation and discussion of potential responses in small groups is encouraged. During the semester in which data were collected, questions were distributed and discussed at each week's tutorial; only in five pre-arranged weeks, however, were quiz answer sheets collected and the answers marked and put towards students' final course grades.[11] Discussion and debate regarding the correct answers occurred in small groups during the first 30 minutes of each tutorial. Tutors were not involved in the discussions, but only collected individual answer sheets at the end of the 30 minutes. At the close of discussion, each individual recorded his own answers to each question, and was asked to record on the back of his answer sheet

[9]Of these, 577 had nonnmissing values for the key variables used in the empirical analysis.

[10]The TEL has been used elsewhere to measure economics understanding; see, for example, Beck & Krumm (1991), McKenna (1994), and Whitehead & Halil (1989).

[11]It is for these weeks that data is available on student attendance and peer group composition, as described below.

the names and student numbers of the other members of his small discussion group.[12]

Within this context, an experimental design regarding small group assignment was implemented as follows. Each week of the semester, students in half of the 30 tutorials were randomly re-assigned to small groups; students enrolled in the other 15 tutorials were permitted to self-select into small groups as they wished each week.[13]

In this context, the peer group of interest is the small study group. The broad group from which the peer group may form each week is the set of students observed to attend that tutorial in that week (which differs from the entire tutorial group at any point in time both because of student attrition from the course during the semester, and because not every enrolled student attended tutorial each week). The observation of actual student attendance and grouping patterns allows a much more up-close look at students and peers than has typically been possible in the peer effects literature.

3.1 Variables

The following variables observed in the data are used in the present study: students' scores on the first and last Test of Economic Literacy (TEL), broken down by question; students' first Reflections on Learning Inventory (RoLI) responses (which include information regarding how well they expect to do in the course, whether English is their first language, and several other background details); their tutorial quiz answers

[12] Additional groupwork was required out of class by students in completing assignments; this information is not used in the present paper.

[13] The randomization was implemented in practice through the creation of a new randomized student grouping list each week, for every randomized tutorial. These lists were taken by assistants to each class, along with quiz sheets, and students were instructed to form the designated groups before beginning work on their quizzes for that week.

for five of the 13 weeks of the semester (the basis of the dependent variable used in the present paper), along with the method of assignment to small groups used in their tutorial (either self-selected or random); the identity of other attending students in their tutorial for each of the five weeks; and the identity of students in the small groups in which they reported working on tutorial quizzes each week.[14] I also observe extensive background information about each student drawn from UniSA's student record files, including birthdate, sex, citizenship status, method of admission to university, TER score (if admitted via a standard high school pathway),[15] courseload, part-time student status, home state, and University program. Finally, I have information on the tutor, size of tutorial, day of the week, and time of day that the tutorial was held.

I construct a measure of each student's initial competency in economics by coding each of the 40 questions on the initial Test of Economic Literacy as having one correct response, worth 1 point, and three incorrect responses, worth 0 points each. Most of the

[14]Small group composition is derived for both types of tutorials (selected and randomized) from the student identification numbers and names of their small group members that students were instructed to write on the backs of their quiz sheets.

[15]The TER, or Tertiary Entrance Rank, is a student's rank out of 100 on a national scale applicable only to year-12 students ("school leavers"—as opposed to mature-age entry students or other non-traditional University applicants). The TER is computed on the basis of both a student's marks (grades) in key high school subjects, and the marks (grades) that fellow students in his graduating cohort achieved. Each student's overall score from this weighted calculation (called the Tertiary Education Score) is then ranked against all others in the student's cohort, to produce a number for each university applicant. Universities then compare applicants' TERs to their entrance criteria (stated partly in terms of required TERs) for each program of study. The TER is therefore best thought of as the Australian analog of the American Scholastic Aptitude Test, although the analogy is not perfect. For more information on the higher education market and funding institutions in Australia, see http://www.dest.gov.au/.

other explanatory variables used in the analysis, with the exception of age, are included as dummies or indexes.

The first few rows of Panel A of Table 1 give some descriptive statistics on the sample used, broken down by whether students were assigned to a tutorial with self-selecting (287 students) or randomized (290 students) small groups. The "ability" variable underlying the own and group ability measures whose distributions are shown on this table is the portion of the average quiz score of each student that is predictable from observable pre-quiz characteristics. (See subsection 4.3 of the paper for more discussion of this measure, and alternative possible measures of baseline ability.) On the Test of Economic Literacy taken at the start of the course, students in the sample scored on average about 24 out of a possible 40 points. On average, students in the sample (who had to have attended at least one tutorial) attended about 4 tutorials out of five that were observed. Students were cautiously optimistic about their chances of success in the course, with an average score of about 2.5 out of 4 on expectations (denoting a value between "OK - a pass - about average" and "Very well - better than average"). Over 80 percent of students were Australian, and slightly over half were female.

4 Empirical design

4.1 Selection equation estimation

The approach followed in this paper begins with a first-stage estimation of the peer group selection equation proposed above, of the form

$$peers_{ijk} = \beta_0 + broadpeers_{ij}\delta + \eta_i + \mu_{ijk} \tag{6}$$

where i identifies individuals, j identifies broad groups (each broad group is the group of students attending a given tutorial in a given week, excluding self), and k identifies the small study group (which, like the broad groups indexed by j, is in practice distinct for each individual because it excludes the self). Time subscripts are omitted from this equation for expositional clarity, but they are implicit: variance across time is what allows the estimation of η_i. The dependent variable $peers_{ijk}$ and the independent variable $broadpeers_{ij}$ can be thought of as measures of average peer ability, at two distinct levels (the small group and the broad group, respectively), neither of which contains the ability of student i.[16] An intercept distinct from the fixed effect ($peers_i^*\omega_i$) is included in Equation 6 to facilitate the post-estimation decomposition of the peer effect in preparation for the second stage. Within this framework,

$$\eta_i = peers_i^*\omega_i \tag{7}$$

or in other words the underlying preferences and social power of individual i are aggregated into one individual-specific fixed effect η_i.[17]

Once Equation 6 has been estimated, I can decompose an individual's observed peer group composition each week into three components: (1) a mechanical component derived from the distribution of the entire population of individuals, and of the supersetting broad group j ($\hat{\beta}_0 + broadpeers_{ij}\hat{\delta}$); (2) a function of the individual's fixed preferences and/or social power within the broad group ($\hat{\eta}_i$);[18] and (3) random luck

[16]For expositional clarity, I postpone discussion of exactly how the key variables $peers_{ijk}$ and $broadpeers_{ij}$ are measured until subsection 4.3.

[17]This effect would also include any early luck that persists across time due to social forces. Early luck can only be persistent in self-selecting tutorials, which may mechanically augment the heterogeneity in both peer group attributes and individual outcomes seen in self-selecting tutorials as compared with randomized tutorials.

[18]Because most students do not switch tutorials during the semester, the individual's preferences

on the day for that individual ($\hat{\mu}_{ijk}$).[19] μ_{ijk} can be interpreted as the component of peer group composition that is due purely to luck on the day (and is uncorrelated with student preferences). Isolating the $\hat{\eta}_i$s allows us to recover, in a second stage, an upper bound on the impact that the portion of peer group composition attributable to students' observed peer group selection has on outcomes.[20] Note that $\hat{\mu}_{ijk}$, as well as the mechanical portion of peer group composition, could still impact outcomes despite the fact that they reflect portions of the environment that were not chosen. In fact, researchers capitalizing on randomized peer groups in order to estimate the peer effect are necessarily limited to estimating the effect on outcomes of that portion of peer group

notated by η in the equation above are best thought of as preferences and/or power conditional on choosing, and attending, a given tutorial in a given week. By specifying this model, I am disregarding the possibility that students enroll in tutorials in a way that might systematically influence the contribution to their small peer group composition that is made by the composition of their observed tutorial group each week. First-year students are typically unfamiliar with tutors and other students, and would therefore probably have little basis other than scheduling conflicts on which to prefer one tutorial over another. Furthermore, because a substantial portion of students do not attend tutorial each week, the extent to which students are effective in choosing the pool from which they then must form small groups on a weekly basis, via their initial choice of tutorial, is argued to be limited at best. Empirical evidence to support this view includes a lack of positive correlation between the ability of the broad peer group and $\hat{\eta}_i$ across student weeks, for either selected or randomized tutorials.

[19] Note that if η_i happened to capture everything about the individual that was simultaneously relevant both to his outcomes and to his choice of peers—in other words, if all unobserved individual effects on outcomes occurred through individuals' selection of peer groups—then there would be no concern about bias in conventional peer effects estimation.

[20] This effect will be overestimated to the extent that student-level preferences for high-achieving peers, and/or their social power levels, are positively correlated with students' own unobserved proclivities to achieve. This is the source of concern regarding ability bias in conventional peer effects estimation.

composition that is represented in Equation 6 as $\beta_0 + broadpeers_{ij}\delta + \mu_{ijk}$.

4.2 Outcome equation estimation

Leaving aside the possibility of individual-specific peer effects for the moment, and with the three portions of peer composition in hand from the first-stage selection equation, I first estimate a second-stage equation assuming the same peer effect for every individual from each of these three portions. The decomposition of peer group ability is made explicit through the use of three subscripts (m = "mechanical," s = "selected" and l = "luck") on the three components of peer ability, and their respective effects:

$$outcome_{ijk} = X_i\beta_1' + \nu_j + pe\bar{e}rs_{ijk,m}\bar{\gamma}_m + pe\bar{e}rs_{ijk,s}\bar{\gamma}_s + pe\bar{e}rs_{ijk,l}\bar{\gamma}_l + U_{ijk} \tag{8}$$

(Again, time subscripts are omitted for expositional clarity.) Note that despite the inclusion of each portion of peer ability—including the portion chosen by students—there should be no bias on the estimates of $\bar{\gamma}_m$ or $\bar{\gamma}_l$ if the selection and outcome equations are correctly specified. However, if the broad group has an effect on outcomes that does not flow through the small study group, then that effect will be captured by $\bar{\gamma}_m$. Given these caveats, our interpretations of the coefficients should be as follows: $\bar{\gamma}_m$ is the effect on individual performance of attributes of the broad group (whether flowing through the small study group, or independently); $\bar{\gamma}_s$ is the effect of the chosen portion of peer group composition on performance, whose estimate, given the estimation framework, may be upwardly-biased due to correlation between $pe\hat{e}rs_{ijk,s}$ (shown in the first stage, Equation 6, as η_i) and U_{ijk}; and $\bar{\gamma}_l$ is the effect of that portion of peer group composition which is due to pure luck on the day. Finally, ν_j is a vector of tutorial-specific effects, which in the present study is composed of indicators for the tutor, whether the tutorial was held on a Monday, and whether the tutorial was scheduled for the morning.

With heterogeneous (individual-specific) treatment effects, the treatment-effects style outcome equation of interest in this context, slightly expanded from the one discussed previously, is

$$outcome_{ijk} = X_i\beta_1 + \nu_j + pe\hat{e}rs_{ijk,m}\bar{\gamma}_m + pe\hat{e}rs_{ijk,s}\bar{\gamma}_s + pe\hat{e}rs_{ijk,l}\bar{\gamma}_l + peers_{ijk}\epsilon_i + U_{ijk} \quad (9)$$

If peer effects are heterogeneous with respect to individuals' selection proclivities, then the ϵ_is should be related to the η_i from the selection equation (Equation 6), the latter of which appears in estimated form in the outcome equation as $pe\hat{e}rs_{ijk,s}$. The two variables would be positively correlated across students in the state of the world where individuals are observed to sort into peer groups in a way aligned with the benefit that they receive from them. They would be uncorrelated or even negatively correlated if something prevents this potentially socially efficient solution from appearing—whether that obstacle is related to preferences, social power, or both.[21]

As implied in Equation 9, in the event of treatment effect heterogeneity that is related to the choice of peer groups, all three components of peer composition may differentially affect outcomes. If sorting induces a socially efficient matching of students to peer groups, then ϵ_i should vary across individuals in a way correlated with different values of η_i (shown in the second stage, once estimated, as $pe\hat{e}rs_{ijk,s}$). This implies the inclusion of an interaction term in a re-specified model such as the following:

$$outcome_{ijk} = X_i\beta_1'' + \nu_j + pe\hat{e}rs_{ijk,m}\bar{\gamma}_m + pe\hat{e}rs_{ijk,s}\bar{\gamma}_s + pe\hat{e}rs_{ijk,l}\bar{\gamma}_l + peers_{ijk}pe\hat{e}rs_{ijk,s} + U_{ijk} \quad (10)$$

The above are the main estimating equations used in the paper. After choos-

[21]This logic undeniably rests on the correct specification of the two equations. Further deductive analysis aimed at discovering the most likely nature of these equations has the potential to advance this discussion substantially.

ing a variable to use as a proxy for underlying ability, I run the peer group selection equation and recover the three portions of peer ability ($\hat{\beta}_0 + broadpeers_{ij}\hat{\delta}$, $\hat{\eta}_i$, and $\hat{\mu}_{ijk}$) which are then input into the outcome equation respectively as $pe\hat{e}rs_{ijk,m}$, $pe\hat{e}rs_{ijk,s}$ and $pe\hat{e}rs_{ijk,l}$. I then experiment with the inclusion or exclusion of the different components of $peers_{ijk}$ to explore the patterns in student reactions to peers. Selection and outcome equations are estimated at the student-week level separately for students in tutorials with randomized and self-selected small groups, and results are compared.

4.3 Choice of peer attributes

To estimate these models, a measure of "relevant peer ability" is required. The choice of this measure is integral to the results of the entire exercise, and there is virtually no theoretical guidance available within the discipline on this matter. The ideal "ability" measure would capture, in more general terms, all peer attributes that are at the same time performance-relevant for the individual and variant systematically by individual. From this perspective, a variety of observable social attributes (e.g., race, gender, age, etc.) that have little to do with what economists would typically call ability may be more appropriate than more direct performance measures. This is because students may perform differentially when they are in observably homogeneous or heterogenous groups, and may tend to be observed in homogeneous or heterogeneous groups systematically, for reasons originating in psychology rather than economics. This is not to say that such processes cannot be modelled by economists; many researchers have been attempting to join economics and social psychology through structural models of sorting and outcomes (see, for example, Akerlof & Kranton (2000) and its spawn over the last several years, such as Fletcher (2005)). Notably, this strand of literature has not been empirically joined to the peer effects literature, something that would be a valuable endeavor to undertake in future research.

In this paper, I measure peers' potential value academically rather than socially. I take the predicted values from the following regression run cross-sectionally, using one observation for each student, to form the student-level underlying ability measures that are then aggregated (by taking the group-wide mean, excluding self) to create the peer composition measures $peers_{ijk}$ and $broadpeers_{ij}$:

$$avgquizscore_i = \psi_1 + Y_i \psi_2 + e_i \tag{11}$$

where $avgquizscore_i$ is the average tutorial quiz score of student i (the sum of all points earned on quizzes divided by the number of quizzes handed in), and Y_i is a vector of all measures on the student observed in advance of any quizzes taking place that might predict tutorial quiz performance.[22] This equation is able to predict roughly 7 percent of the variation in quiz score performance across students. The remaining variation in quiz score performance may be due both to chance and to student-specific and/or contemporaneous factors that the current model structure does not accommodate. As such, the current paper estimates only the impact of a linear function of peers' performance-

[22] In practice, this vector includes the following variables: score on the initial Test of Economic Literacy; dummies for whether the student reported on the initial Reflections on Learning Inventory liking math, being the first in his family to attend university, to have taken economics elsewhere, to have attended a public school, to have English as a first language, and to enjoy working alone; TER score (and a dummy for its absence, if the student entered university through nontraditional means); the student's reported expectations of success in the course, on a scale of 1 to 4; the number of courses the student is taking in the current semester; and dummies for South Australian residence, Australia as the student's home country, female, University entry type (TAFE/year-12/STAT/tertiary transfer), part-time student status, and each of six different major programs of study.

relevant observable characteristics on own performance.

The student-level predicted values of Equation 11 are used to construct peer ability at both levels (the small group and the broad group). The term "ability" is used henceforth to denote the predicted value of this equation for each student. Different aggregations of this ability measure are used to construct the various peer measures. "Small group ability" ($peers_{ijk}$) is the main peer measure analyzed in the paper, and is the average of the ability measures of one's small study group (excluding self); "broad group ability" ($broadpeers_{ij}$) is the average of the ability measures of the broad group—all students who attend one's tutorial in a given week (excluding self). Descriptive statistics on the estimating samples (one sample each for the tutorials with randomized versus those with self-selected small groups) are presented at the student level, and the student-week level, in Panels A and B respectively of Table 1.

5 Estimation and results

First, the three distinct portions of small group ability are isolated through running the selection equation (as specified in Equation 6). The results of these regressions are shown in Table 2. Several points are worth mentioning about this table.

Perhaps the most striking aspect of the table is the consistent and potent association of the ability of the broad peer group (the pool from which students may choose their peers) with the ability of the small study group. While mechanical, this dependence of peer group ability on broad group ability highlights the importance of the broad learning environment in creating the sort of variance in subgroup peer measures with which peer effects are often identified. In generating any interpretation of a peer effects result, it is worth bearing in mind that observed variance at a sub-level is naturally a function of variance at a higher level, which could in fact be the causal path to outcomes.[23] In the case at hand, it may be that the ability of the broad group—not just the small study group—affects student performance. Even in the absence of any true effect from small group ability, there is such a strong correlation between the abilities of the small and broad groups that were we to include only small group ability (and not broad group ability) in the outcome equation, we might incorrectly infer importance of the small group in determining outcomes.

Table 2 also shows an interesting and expected pattern in variance explained when student-level fixed effects are added. For students enrolled in tutorials with self-selecting groups, adding such effects gains us a massive 37 percentage points of additional variance explained. However, in the tutorials with randomized small groups, we only gain 3 percentage points. This is expected, as students who are not supposed to be choosing their peers should not be observed to influence their peer group composition systematically. Note that those three percentage points are present, however: the randomization, like most in natural experiment undertakings, is not perfect.

Table 3 shows the results of quiz performance equations run separately for students in tutorials with selected versus randomized small groups. The first and fourth columns show results from a specification which does not bifurcate the small group ability into three portions, but instead uses small group ability as an aggregate along with broad group ability, and a set of student-level covariates, as explanatory variables. From these regressions, one would conclude that while broad group ability (the ability of the set of students attending the tutorial in the given week) may be a positive influence on student performance, its effect is not statistically significant for students in either the

[23]This is the crux of the difficulty with the instrumental variables approach taken by Evans, Oates & Schwab (1992) and ensuing authors, despite its intuitive appeal.

self-selected or randomized assignment protocol. Further, students in randomized small groups do not seem to benefit from higher ability in those small groups. Students in self-selected small groups may benefit from small group ability, but since the aggregate measure of small group ability includes some portion that was chosen by students and may be correlated with the error in the outcome equation, we cannot be sure.

The ensuing columns of Table 3 show estimation results when small group ability is broken into its component parts, based on the first-stage selection equation, and included separately as distinct outcome predictors. The standard errors for Columns 2, 3, 5, and 6 of the table have been bootstrapped to take account of the sampling bias present in the predicted values taken from the first stage of estimation. For students in tutorials with self-selected small groups (and to a lesser extent for those in tutorials with randomized groups), we would anticipate that the estimated effect of chosen small group ability would be higher than that of the portion of small group ability due to random luck on the day, since the former may be upwardly biased due to correlation between selection proclivities and unobservable outcome determinants. This is in fact the case, for students in both assignment protocols. The effect of the unpredictable portion of small group ability (labelled "Unpredictable small group ability" in the table)—arguably the most accurately estimated of the three small group effects—is estimated to be lower than that of the chosen portion of small group ability (labelled "FE (small group ability)" in the table). The estimated effect of the chosen portion is positive and significant for students in self-selected tutorials, and as this may be a function of upward bias, it should be viewed as an upper bound. The effect of the unpredictable portion of peer ability, however—a portion that is, by design, *not* subject to upward bias—is insignificant for students in both types of tutorials, and even negatively signed for those in randomized tutorials.

The portion of small group ability that is a linear function of the available pool of peers (labelled "Mechanical small group ability" in the table) is estimated to have a positive and significant effect on outcomes for students in self-selected groups. As noted above, the influence directly from the broad group, rather than through the small group—shown in column 1 of Table 3 to be estimated at 3.7216 for students in self-selected groups—is included in the "Mechanical small group ability" estimated influence of 4.3576. By deduction, the influence of the portion of small group ability that is mechanically determined by the composition of the broad group is roughly 4.3576 - 3.7216, or .6360. This is in the same ballpark as the estimated influence of the unpredictable portion of small group ability (.7791), and both of these estimates fall short of the presumably upwardly-biased estimate of the influence of the chosen portion of peer ability (1.2895).

For students in randomized groups, the story is somewhat different. The portion of small group ability that is due to random chance is actually estimated to have a *negative* (although insignificant) influence on student performance (-.3839). Consistent with this, the estimated total impact of the ability of the broad group, whether by itself or through small group ability, is less than the estimated direct influence of broad group ability (2.0665 versus 2.0963, respectively), implying again a *negative* influence of the portion of small group ability that is a function of broad group ability. Despite the lack of statistical significance, these results tend to indicate that unlike students in self-selected groups, students in randomized groups are not positively responding to the portion of the ability of their small group that is outside their direct control. We do see a positive, although insignificant, estimated effect of the chosen portion of small group ability for those in randomized groups, which only serves to remind that the randomization was not perfect. As with those students in self-selected small groups, this effect may be an artifact of correlation between selection proclivities and/or social power (which are muted in this sample, compared with the sample of students in self-selected groups) and unobserved performance determinants.

Leaving aside the lack of statistical significance, the economic importance implied by the point estimates of the most accurately estimated peer group influence—that of the unpredictable portion of peer ability—is small. For students in self-selected small groups, increasing the unpredictable portion of average ability of a student's small group peers by one standard deviation would be estimated to yield a change in quiz performance equivalent to roughly 6 percent of a standard deviation. This predicted effect is slightly more than twice the size of the effect on quiz scores estimated to result from a one-standard deviation increase in own Test of Economic Literacy score, again notwithstanding the fact that neither effect is statistically significant.

It may be that equations that assume a homogeneous peer effect for all students cannot adequately capture the influence of peers. If free choice leads to a socially efficient solution, then we should observe higher returns to peer ability for individuals who are observed with higher-ability peers. Table 4 shows the results of testing, using two alternative functional forms, whether undergraduate students are observed to choose small study groups in a way that is aligned with their individual-specific returns to peer ability.

The first column of Table 4 shows results for students in self-selected groups from regressions including all three portions of small group ability, the set of explanatory variables used in Table 3, and an interaction term equal to the following:

$$peers_{ijk} * (pe\hat{e}rs_{ijk,s} > 0) \tag{12}$$

The third column shows corresponding results for students in randomized groups. The second and fourth columns of Table 4 show results using a different form of the interaction, namely

$$pe\hat{e}rs_{ijk_s} * pe\hat{e}rs_{ijk,s} \tag{13}$$

in order to see whether perhaps only the returns to that part of peer ability that is chosen, rather than returns to peer ability from any source, are correlated with students' selection decisions. All standard errors in Table 4 are bootstrapped across the entire two-stage estimation procedure.[24]

While the first formulation does not reveal any meaningful heterogeneity—and does not fit the data as well as the linear model—the second heterogeneity formulation reveals the expected pattern. There appear to be increasing returns to peer ability, but only with respect to that portion of peer ability which is chosen by the student. This is consistent with the notion that those students who stand to benefit most from peer ability are those who are observed systematically to be exposed to the highest levels of it. It is also consistent with increasing returns to peer ability across all students. However, the sole formulation of the heterogeneity term that is significant parameterizes these returns as occurring only for the systematically selected portion of peer ability; and even this formulation is only positive and significant for those students in self-selected small groups. These aspects of the results support the contention that it is choice rather

[24]The following variants of the interaction term were also tried: $(pe\hat{e}rs_{ijk,s} > 0) * pe\hat{e}rs_{ijk,m}$;$pe\hat{e}rs_{ijk,s} * pe\hat{e}rs_{ijk,m}$; and $pe\hat{e}rs_{ijk,s} * ability$. None was significant for students under either assignment protocol.

than randomization that brings about increasing returns. Student selection of peers may therefore be aligned to some extent with the relative returns to peer ability—or more specifically, to *chosen* peer ability—across students.

Some evidence is provided below that the chosen portion of peer ability is not correlated with observable predictors of success. However, due to the potential upward bias arising from imperfect control for own ability in these equations and the possible correlation of this unobserved portion of own ability with chosen peer ability, it is impossible to conclude with certainty that peer ability selection is aligned with returns. It also may be that these results reflect social power differentials rather than underlying ability differentials or differences in susceptibility to peer influence. All that we can say based on these results is that we cannot reject the notion that students' selection of peers is aligned with their heterogeneous returns to peers.

As a general point, it is of interest that more of the variation in student performance is explained by student-level attributes such as expectations of course success, nationality, residence, mode of entry, and so on for students in self-selected groups than for those in randomized groups. In addition, students in self-selected groups do slightly worse on average than those in randomized groups. These facts might lead us to surmise two things. First, when students are allowed to select their own peer groups, their choices may through some mechanism unlock returns to their own attributes. This mechanism may be psychological or social more than economic, perhaps having to do with feeling comfortable or in one's element. As a result, students' more typical behavioral patterns surface, so (speaking roughly) the well-prepared students excel and the more poorly-prepared students do not. When faced with the unpredictable, via the randomization assignment protocol, students are forced outside of their comfort zones in the learning process, and their own ex ante attributes are not as important as what happens in that new environment. Randomization may therefore be a leveler—not good for students who are well-prepared to succeed (and don't need exogenous shocks to aid learning), but perhaps a method of pushing more disadvantaged students out of their comfort zone and into an unfamiliar environment where more learning may take place than otherwise would.[25] Another route to "leveling" through randomization could be that those with low social power (who may in fact be observably disadvantaged students) do not feel the effects of it, so that while individuals' (potentially efficient) peer ability preferences are not realized, neither is any social exclusion that might be correlated with relative returns to peer ability. The result of randomization may not be socially optimal, in the sense of aligning relative returns with assignment to groups, but there is no weighting of group assignments by social power.

Some mild evidence of support for the contention that randomization is an outcome-leveler in this context is provided by Table 5, which shows the average tutorial quiz scores for low- versus high-background ability students, by group assignment protocol (randomized versus self-selected). Students with below-average scores on the Test of Economic Literacy, and those with below-average predicted average quiz performance, do better in randomized tutorials than in self-selected tutorials. Students with above-average values on those measures do worse in the randomized protocol than they do when allowed to select their own groups. However, perhaps surprisingly, students with higher-than-average expectations of success in the course do *not* perform better under the self-selection protocol. Students who expect to do better in the course than the average student expects to do perform worse, on average, in self-selected tutorials than

[25]However, this learning does *not* seem to come from peer ability, as can be seen from the results in Table 3.

in randomized tutorials—while those with lower-than-average expectations of success perform very slightly better in self-selected groups. However, it should be noted that none of the above-mentioned differences is statistically significant.

Table 6 presents a few summary statistics at the small group level for selected (N=329) versus randomized (N=353) groups. The average size of a selected group is 2.75 students, compared to 2.51 for a randomized group. Other observable characteristics are similar. The degree of selection on a given observable characteristic that is actually occurring when self-selected groups form can be gauged by a comparison of the average of the standard deviations of the characteristic within all self-selected small study groups, with the average of the standard deviations of the same characteristic within randomized small groups.[26] This comparison reveals some degree of sorting on expectations of success in the course (average standard deviations of .63 and .58 within randomized versus self-selected groups, respectively), Australian citizenship (.19 versus .16) and gender (.39 versus .23); but there is little sorting on the more direct ability measures of Test of Economic Literacy scores (5.86 versus 5.84) and predicted average quiz scores (.24 versus .23).

Table 7 presents some correlations at the student level of the selection proclivities estimated in the selection equations ($\hat{\eta}_i$) and various observable attributes. Of interest, there is no statistically significant relationship (at the 5% level) between selection proclivities and any observable characteristic for students in randomized groups. For students in self-selected groups, both gender and expectations are related to sorting proclivities, although more conventional measures of ability are not. It appears, from this analysis and the comparison of average standard deviations reported above, that

[26]This analysis method for gauging the degree of selection is suggested in Arcidiacono et al. (2005).

students' natural selection processes are a blend of social and academic components which are not well-proxied by the academic component alone.

6 Limitations

A number of caveats should be borne in mind when considering the results of this work as it stands. First, I have taken no account of differential attendance levels across students or wholesale attrition from the course. Second, I have not used the additional measures of course performance (i.e., scores on out-of-class group assignments, the individual final exam, and the final Test of Economic Literacy) that are available in the data set. The implication of this is that any performance effects found from the exercises above may be ephemeral. In future work in this vein, I plan to test the longevity of social effects observed in the classroom by capitalizing on these additional performance measures.

I have also used a very limited measure of peer attributes, and despite robustness checks using two other measures, more could be done. In future work it would be desirable to develop a method to test a variety of different peer attributes that may be important to students, with the goal of divining the weighted combination of them that best expresses what students care about when selecting their peers in a particular context. Of note, when running this analysis using initial Test of Economic Literacy score as the measure of peer ability, results very similar to what is reported in this paper are found. However, when using average quiz performance (instead of the predictable portion of that performance, based on observable background characteristics), students in randomized groups are estimated to react more strongly to all forms of peer ability than students in self-selected groups. This may indicate the presence, and heightened importance in the randomized learning context, of spontaneous or interactive learning effects within randomly-formed small groups that are difficult to predict using conventional econometric techniques. It may also further underline the potential effect on learning of erasing the correlation of social power with selection.

Lastly, as alluded to elsewhere in the paper, the estimation structure used here does not take account of the fact that group formation is a social phenomenon as much, or more, than an individual phenomenon. As such, it would be desirable to test whether selection as a function of individual variability (that is, estimated on individual-level data or below, as is done in this paper) or as a function of group variability (that is, estimated at a level above that of the individual) best fits observed empirical patterns. This is an enterprise beyond the scope of current research in the area, but may develop through time as economists borrow concepts and analysis methods from other social sciences.

7 Conclusion

This paper makes three main contributions. The first is to add some conceptual clarity to economists' often labored empirical efforts to identify peer effects. This contribution includes a caveat with regard to the search for randomization as an identification strategy. To the extent that (1) peer composition results from random variation in environments, and/or (2) individuals react to the random portion of peer ability identically to the way they react to peer ability which is a function of their own choices, randomization is a reasonable strategy to pursue. However, to the extent that peer group composition is a function of individual preferences, then if the effect of the "chosen" portion of peer group composition is stronger or weaker than the effect of the "random" portion, using randomization to identify peer effects may result in over-or under-estimation of the effect of peers in a non-experimental context.

Second, while economists (including those working in the treatment effects and ability bias literatures) have long had to assume that individuals' choice of context would be guided by levels of anticipated returns to that context, this assumption has rarely been introduced into a social context. While I cannot test the individual rationality of students directly, this paper considers the notion that free social choice in the context of undergraduate education may lead to a socially efficient allocation of high-ability peers. I find evidence consistent with this, but cannot establish for certain that my estimators are free of the conventional upward bias due to sorting by unobserved ability.

Finally, educational policymakers may find the results of this analysis useful in designing programs to help both disadvantaged and well-prepared students. Preliminary evidence suggests, in broad strokes, that those who are well-prepared best capitalize on their preparation when they are allowed to find their own way, while poorly-prepared students may benefit from facing educational contexts that they would not necessarily have chosen themselves, whether because they would not have preferred them or because they would have been excluded from them. Informal policies that subtly break up low-achievement cliques or study partnerships and replace them with re-assignment to other milieus may be effective in helping marginal students.[27] More generally, to revisit a key educational policy choice, we might look to these results to determine whether random assignment to study groups or self-selection into study groups would be predicted to be optimal for student learning. The preliminary evidence shown here indicates that

[27] The results in the paper that support these recommendations are in line with the results of a much larger randomization experiment conducted a very different context, looking at a very different and wider array of outcomes—the Moving to Opportunity experiments (see, for example, Katz, Kling & Liebman (2001)). They also accord with anecdotal evidence from primary school classroom teachers.

randomization may lead students who are poorly prepared to do better than they would under self-selection, while simultaneously leading students who are well-prepared to do worse. This, in turn, implies that the policy maker may face a distributional trade-off: neither policy is optimal for all students.

Preliminary evidence presented in this paper suggests that randomization may produce different effects on student learning than self-selection, and that students *may* use their power of choice to improve the social efficiency of observed sorting into peer groups, if typical models of selection and outcomes are accurate. Much remains to be done in economics to sort out the complex nature of, influences on, and effects of, free choice. In future peer effects research, we will hopefully find new ways to use the information available in panel data sets with changing contexts to explore how people choose their groups in practice, and what impact these choices have on their outcomes.

References

Akerlof, G. & Kranton, R. (2000), 'Economics and identity', *Quarterly Journal of Economics* **115**(3), 715–753.

Altonji, J. G., Elder, T. E. & Taber, C. R. (2005), 'Selection on observed and unobserved variables: assessing the effectiveness of Catholic schools', *Journal of Political Economy* **113**, 151–184.

Angrist, J. D., Imbens, G. W. & Rubin, D. B. (1996), 'Identification of causal effects using instrumental variables', *Journal of the American Statistical Association* **91**(434), 444–454.

Angrist, J. & Krueger, A. (1991), 'Does compulsory school attendance affect schooling and earnings?', *Quarterly Journal of Economics* **106**, 979–1014.

Arcidiacono, P., Foster, G., Goodpaster, N. & Kinsler, J. (2005), Estimating spillovers in the classroom with panel data. Working Paper.

Beck, K. & Krumm, V. (1991), 'Economic literacy in German speaking countries and the United States: First steps to a comparative study', *Economia* **1**(1), 17–23.

Card, D. (2001), The causal effect of education on earnings, *in* D. Card & O. Ashenfelter, eds, 'handbook of Labor Economics', North-Holland Publishers.

Echenique, F. & Fryer, R. (2005), On the measurement of segregation. NBER Working Paper 11258.

Evans, W. N., Oates, W. E. & Schwab, R. M. (1992), 'Measuring peer group effects: A study of teenage behavior', *Journal of Political Economy* **100**(5), 966–991.

Fletcher, J. (2005), Evaluating economic and social interventions that increase college enrollment: Academic standards, social status, and the quest for identity. Working Paper.

Foster, G. (2005), 'Making friends: A nonexperimental analysis of social pair formation', *Human Relations* **58**(11), 1443–1465.

Foster, G. (2006), 'It's not your peers, and it's not your friends: Some progress toward understanding the peer effect mechanism', *Journal of Public Economics* **90**, 1455–1475.

Heckman, J. J. & Robb, R. (1986), Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes, *in* H. Wainer, ed., 'Drawing Inferences from Self-Selected Samples', Springer, pp. 63–107.

Hirano, K. & Imbens, G. W. (2004), The propensity score with continuous treatments, *in* A. Gelman & Xiao-Li, eds, 'Missing Data and Bayesian Methods in Practice', Wiley Press, New York.

Hoxby, C. & Weingarth, G. (2006), Taking race out of the equation: School reassignment and the structure of peers effects. Working Paper.

Katz, L. F., Kling, J. R. & Liebman, J. B. (2001), 'Moving to opportunity in Boston: Early results of a randomized mobility experiment', *Quarterly Journal of Economics* **116**(2), 607–654.

Krauth, B. V. (2005), 'Peer effects and selection effects on smoking among Canadian youth', *Canadian Journal of Economics* **38**(3), 735–757.

McKenna, K. (1994), High school economics in Australia, *in* W. Walstad, ed., 'An International Perspective on Economic Education', Kluwer Academic Publishers, pp. 219–232.

Meyer, B. D. (1995), 'Natural and quasi-experiments in economics', *Journal of Business and Economic Statistics* **13**(2), 151–161.

Sacerdote, B. & Marmaros, D. (2006), 'The importance of race versus location in determining social interactions', *Quarterly Journal of Economics* **Forthcoming**.

Whitehead, D. & Halil, T. (1989), *The test of economic literacy: Standardisation in the U. K.*, Institute of Education, University of London, London. Research Papers in Economics Education.

Table 1: Descriptive statistics: Means and standard deviations

Panel A: *Student Level*[a]	Selected groups	Randomized groups
Australian	**0.86** (-)	**0.81** (-)
Female	**0.52** (-)	**0.51** (-)
Test of Economic Literacy score	**24.49** (6.97)	**24.03** (6.79)
Number of observed tutorials attended	**3.94** (1.26)	**3.99** (1.20)
Expectation of doing well in course	**2.51** (.72)	**2.46** (.82)
Average quiz score	**6.74** (1.11)	**6.82** (.97)
Own ability	**6.73** (.30)	**6.80** (.29)
Average small group ability	**6.75** (.20)	**6.81** (.19)
Average observed broad group ability	**6.74** (.10)	**6.81** (.10)
Average mechanical small group ability	**6.75** (.11)	**6.81** (.10)
Average unpredictable small group ability	**0.00** (.00)	**0.00** (.00)
Average fixed portion (FE) of small group ability	**-0.00** (.17)	**-0.01** (.16)
N	287	290
Panel B: *Student-Week Level*	Selected groups	Randomized groups
Quiz score	**6.82** (1.80)	**6.84** (1.76)
Own ability	**6.75** (.30)	**6.82** (.30)
Small group ability	**6.75** (.23)	**6.82** (.25)
Broad group ability	**6.74** (.10)	**6.81** (.10)
Mechanical small group ability	**6.75** (.11)	**6.82** (.10)
Unpredictable small group ability	**0.00** (.13)	**0.00** (.19)
Fixed portion (FE) of small group ability	**0.00** (.17)	**0.00** (.14)
N	904	886

[a]Panel A of this table provides means (in bold) and standard deviations (in parentheses) of key analysis variables at the level of the student, for all students appearing in the analysis samples; Panel B shows similar statistics at the student-week level (multiple observations per student), for all students appearing in the analysis samples. See Sections 3.1 and 4.3 for details on variable definitions.

Table 2: Peer group selection equations

Dep. Var: Small Group Ability[a]	Selected groups		Randomized groups	
Constant	**.5960***	**-1.0165**	**-.0527**	**.1728**
	(.2586)	(1.4080)	(.2571)	(1.6399)
Broad group ability	**.9121****	**1.1513****	**1.0086****	**.9755****
	(.0384)	(.2087)	(.0377)	(.2408)
Student-level FE?	no	yes	no	yes
N	904	904	886	886
Adjusted R-squared	.1527	.5247	.1729	.2041

[a]Each column shows the estimates from a separate regression. Standard errors are clustered at the tutorial-week level. See text for details on variable definitions.

Table 3: Quiz performance equations

Dep. Var: Tutorial Quiz Performance[a]	Selected groups			Randomized groups		
Small group ability	**1.0696****	–	–	**-.1056**	–	–
	(.2271)			(.2137)		
Broad group ability	**3.7216**	–	–	**2.0963**	–	–
	(2.1317)			(2.6133)		
FE (small group ability)	–	**1.2895****	–	–	**.7587**	–
		(.2990)			(.3655)	
Unpredictable small group ability	–	**.7791**	**.7791**	–	**-.3839**	**-.3839**
		(.4931)	(.4680)		(.3451)	(.3669)
Mechanical small group ability	–	**4.3576****	**4.0326****	–	**2.0665**	**2.1321**
		(1.0169)	(1.8825)		(1.2995)	(2.7909)
Test of Economic Literacy score	**.0053**	**.0049**	**.0073**	**.0027**	**.0009**	**.0026**
	(.0090)	(.0090)	(.0105)	(.0103)	(.0107)	(.0100)
Expectations=2	**.2736**	**.2920**	**.1841**	**.0819**	**.0492**	**.0812**
	(.1854)	(.2516)	(.2382)	(.2176)	(.2047)	(.2584)
Expectations=3	**.3178**	**.3467**	**.1771**	**.2535**	**.2156**	**.2528**
	(.2180)	(.2963)	(.2783)	(.2424)	(.2504)	(.2462)
Expectations=4	**.9762****	**.9932****	**.8934***	**.2486**	**.2447**	**.2485**
	(.2942)	(.3119)	(.3338)	(.2941)	(.2651)	(.3523)
South Australian residence	**-1.4326***	**-1.4475***	**-1.3598***	**.2822**	**.2761**	**.2820**
	(.5988)	(.7082)	(.6112)	(.3077)	(.3702)	(.3522)
Australian citizenship	**1.6709****	**1.6836***	**1.6067****	**-.4382**	**-.4113**	**-.4376**
	(.5130)	(.6879)	(.6774)	(.4366)	(.4276)	(.4071)
Female	**-.0128**	**-.0210**	**.0270**	**.0425**	**.0627**	**.0429**
	(.1346)	(.1088)	(.1290)	(.0990)	(.1225)	(.1545)
N	904	904	904	886	886	886
Adjusted R-squared	.1586	.1584	.1471	.0540	.0578	.0557

[a]Each column shows the estimates from a separate regression; standard errors have been bootstrapped for Columns 2, 3, 5, and 6. See text for variable definitions. The "Expectations" dummy variable array codes the student's response to the question, "How well do you expect to do in this subject overall?" on a scale of 0 ("I think I'll probably fail") to 4 ("Excellently—near the top of all students"); the excluded category is the superset of 0 or 1. Additional explanatory variables included in these regressions whose estimates are not shown in this table are as follows: tutor; whether the tutorial fell on a Monday and whether it fell in the morning; dummies for whether the student reported on the initial Reflections on Learning Inventory liking math, being the first in his family to attend university, to have taken economics elsewhere, to have attended a public school, to have English as a first language, and to enjoy working alone; TER score (and a dummy for its absence, if the student entered university through nontraditional means); the number of courses the student is taking in the current semester; and dummies University entry type (TAFE/year-12/STAT/tertiary transfer), part-time student status, and each of six different major programs of study.

Table 4: Testing for heteroskedasticity in response to peers as a function of selection proclivities

Dep. Var: Tutorial Quiz Performance[a]	Selected groups		Randomized groups	
FE (small group ability)	**1.0405**	**.8560***	**.9890**	**.6648**
	(.5756)	(.4102)	(.6674)	(.3613)
Unpredictable small group ability	**.7687**	**.7791***	**-.3773**	**-.3839**
	(.5130)	(.3943)	(.3664)	(.4090)
Mechanical small group ability	**4.3322****	**4.1775***	**2.0707**	**2.0980**
	(1.1070)	(1.6207)	(2.2817)	(2.5007)
(FE>0) x small group ability	**.0157**	–	**-.0123**	–
	(.0325)		(.0381)	
FE x FE	–	**3.8910****	–	**-2.0343**
		(.9821)		(1.2649)
N	904	904	886	886
Adjusted R-squared	.1577	.1649	.0569	.0585

[a]Each column shows the estimates from a separate regression; standard errors have been bootstrapped for all columns. See text for variable definitions. Additional explanatory variables included in these regressions whose estimates are not shown in this table are as follows: tutor; whether the tutorial fell on a Monday and whether it fell in the morning; the student's score on the initial Test of Economic Literacy; dummies for whether the student reported on the initial Reflections on Learning Inventory liking math, being the first in his family to attend university, to have taken economics elsewhere, to have attended a public school, to have English as a first language, and to enjoy working alone; TER score (and a dummy for its absence, if the student entered university through nontraditional means); the student's reported expectations of success in the course, coded 0 to 4 (with students answering 0 or 1 forming the excluded category); the number of courses the student is taking in the current semester; and dummies for South Australian residence, Australia as the student's home country, female, University entry type (TAFE/year-12/STAT/tertiary transfer), part-time student status, and each of six different major programs of study.

Table 5: Average quiz scores for well- versus poorly-prepared students

Student Level[a]	Selected groups		Randomized groups	
Ability measure:	Below Average	Above Average	Below Average	Above Average
Ability	**6.4497**	**7.0204**	**6.5833**	**6.8979**
Test of Economic Literacy score	**6.6775**	**6.7685**	**6.8104**	**6.7197**
Expectations of doing well	**6.7575**	**6.6995**	**6.7245**	**6.7923**

[a]Means of quiz scores are first calculated at the level of the student, across all quizzes, and then averaged across all students who meet the given criteria for each cell. Columns labelled "Above average" display average tutorial quiz scores of students who are observed to have values on the background characteristics shown in the leftmost column that are equal to or above the average value of those characteristics, individually, across the student-level sample within the corresponding tutorial type (randomized or selected). Columns labelled "Below average" display average tutorial quiz scores of students who are observed to have values on the background characteristics shown in the leftmost column that are below the average value of those characteristics, individually, across the student-level sample within the corresponding tutorial type.

Table 6: Attributes of chosen versus randomized groups

Group Level[a]	Selected groups	Randomized groups
Average size of group	**2.75** (.83)	**2.51** (.65)
Average Test of Economic Literacy score	**24.49** (4.22)	**24.49** (4.49)
Average percent female	**.54** (.42)	**.52** (.34)
Average percent Australian	**.89** (.23)	**.83** (.27)
Average level of expectations	**2.51** (.46)	**2.46** (.57)
Average of own abilities	**6.74** (.20)	**6.82** (.20)
Within-group stddev of ability	**.22** (.16)	**.23** (.16)
N	329	353

[a]Means (in bold) and standard deviations (in parentheses) are calculated at the level of the small study group, and therefore the correct interpretation of the mean statistics is as averages of averages calculated across all individuals within each group. For example, on average, a small group in a selected tutorial is about 54% female. The "Within-group stddev of ability" row displays the mean, across all groups in tutorials of the given type, of the standard deviation of own ability within the small group.

Table 7: Correlations of estimated selection proclivities with observables

Student-Level Correlations of Selection Proclivities ($\hat{\eta}_i$) with:[a]	Selected groups	Randomized groups
Own ability	**.0601** (.3102)	**.0630** (.2848)
Australian	**-.0709** (.2314)	**-.0935** (.1123)
Female	**.1418** (.0162)	**-.0783** (.1833)
Test of Economic Literacy score	**-.0553** (.3506)	**.0356** (.5459)
Expectations	**-.1774** (.0026)	**.0884** (.1332)
N	287	290

[a]Correlations are calculated at the student level; p-values are in parentheses.