The Causal Effect of Studying on Academic Performance

Todd R. Stinebrickner

Department of Economics
Social Science Centre
The University of Western Ontario
London, Ontario Canada N6A 5C2
email: trstineb@uwo.ca

Ralph Stinebrickner

Berea College

Despite an increased awareness of the policy importance of understanding the determinants of educational outcomes, knowledge of the importance of perhaps the most fundamental input in the education production function - students' study time and effort - has remained virtually non-existent. In this paper, we examine the causal effect of studying on grade performance using an Instrumental Variable approach . Our approach takes advantage of particular institutional details at Berea College. It is made possible by new data from the Berea Panel Study whose unique features provide us with information about student time-use, allow us to construct the instrumental variable, and allow us to provide evidence about the validity of the instrument. We find that studying has a very important causal effect on student grade performance and that the Instrumental Variable estimate is much larger than the Ordinary Least Squares estimate. We design a test that provides evidence that the difference between the OLS and IV estimates arises primarily because individuals react to grade shocks in a particular semester by changing their study effort in that semester, and discuss why the findings sound a cautionary alarm about the use of Fixed Effects estimators in cases where behavioral responses to information may be present. The paper is perhaps the first to identify a specific underlying avenue through which peer effects are transmitted in educational contexts and also provides evidence that video games can have a large causal effect on educational outcomes.

JEL codes: Education I2, Labor and Demographic Economics J0

Acknowledgments: The work was made possible by generous funding from The Mellon Foundation, The National Science Foundation, The Social Science Humanities Research Council and support from Berea College. We are very thankful to Anne Kee, Lori Scafidi, Dianne Stinebrickner, Pam Thomas, and Albert Conley who have played invaluable roles in the collection and organization of the data from the Berea Panel Study. The authors would also like to thank John Bound and Lance Lochner for comments.

I. Introduction

Despite an increased awareness of the policy importance of understanding the determinants of educational outcomes, knowledge of the importance of perhaps the most fundamental input in the education production function - students' study time and effort - has remained virtually non-existent. In the context of higher education, this void in our understanding is important because designing sensible and cost-effective education policies requires an understanding of the extent to which college outcomes of interest are driven by decisions that take place after students arrive at college rather than by background factors that influence students before they arrive at college.

In perhaps the only recent research that directly examines the relationship between study effort and outcomes, Stinebrickner and Stinebrickner (2004) used multiple time-use surveys that were collected for freshmen students at Berea College as part of the Berea Panel Study. The descriptive relationship that was estimated between a student's first semester grade performance and his/her average daily study hours during the first semester dealt with the errors-in-variables problem that arises because average daily study hours cannot be observed directly.[1] A test of the null hypothesis that study time plays no explanatory role in the grade equation was rejected with a t-statistic of approximately 6.0 and the estimated coefficient associated with study time was found to be quantitatively large.

The bias associated with viewing the descriptive estimator in Stinebrickner and Stinebrickner (2004) as an estimator of the causal role that studying plays in the grade production process arises, in part, because

[1]A number of authors have studied the relationship between employment during school and academic performance. A complete summary of this work is available in Ruhm (1997) and Stern and Nakata (1991). In more recent work, Stinebrickner and Stinebrickner (2003) took advantage of the institutional details of a mandatory labor program at Berea College to establish that working during school can have a quantitatively large and statistically significant negative causal impact on academic performance. Perhaps most similar in spirit to the objectives of examining inputs into the education production function directly is work of Betts (1997) who finds that the amount of homework assigned by teachers between grades seven and eleven has a quantitatively important relationship with student achievement as measured by test scores.

students who spend more time studying may be different in ways related to, say, ability than those who spend less time studying.[2] Theory alone is not sufficient to determine whether students who put more effort into studying tend to be of higher or lower ability than other students. On one hand, high ability students may enjoy studying more than other students. On the other hand, given that high ability students may achieve the maximum grade in a class at lower amounts of studying, an additional hour of studying may lead to higher grade and future benefits for the lower ability student(s), and, in addition, low ability students may be forced to study more just to "stay afloat." Further confounding the endogeneity problem is the possibility that individuals who receive bad grade shocks during a particular semester may react by increasing their effort during that semester.[3] Thus, not only is it not possible to know the size of the bias that would be present if one were to view the descriptive estimator as an estimator of the causal effect, but it is also not possible to know the direction of the bias. While the strong descriptive relationship between study effort and grades may be suggestive, it does not guarantee that effort plays a large role in college grade performance.

In this paper, we examine the causal effect of studying on grade performance using an Instrumental Variable approach (IV). Our approach is made possible by new data from the Berea Panel Study whose unique features provide us with information about student time-use, allow us to construct the instrument that we use, and allow us to provide evidence about the validity of the instrument. In section II we describe the Berea Panel Study in general terms. In Section III we describe our IV approach with a detailed examination of whether the instrument is likely to satisfy the two conditions necessary to be valid. Section IV contains

[2]While in this paper, for ease of discussion, we often talk about unobserved ability as a source of omitted variables bias, it is also possible that people who study more may differ in other behavioral ways. For example, people who study more may have different class attendance habits, sleep habits, or may study differently than other students.

[3]S&S (2004) discuss the relationship between the endogeneity issues in this context and the difficulties of establishing causality between education and earnings which is discussed in detail in Card (1999).

results. Our IV estimate indicates that studying has a very important causal effect on student grade performance and is much larger than the Ordinary Least Squares (OLS) estimate. We design a test which provides strong evidence that the difference between the OLS and IV estimates arises primarily because individuals react to grade shocks in a particular semester by changing their study effort in that semester and discuss why this finding sounds a cautionary alarm about the use of Fixed Effects estimators. In Section V we discuss, in more detail, the importance of this work and conclude.

II. A general overview of the Berea Panel Study

Todd Stinebrickner and Ralph Stinebrickner (hereafter referred to as S&S) began the Berea Panel Study (BPS) with the explicit objective of collecting the type of detailed information that is necessary to provide a comprehensive view of the decision-making process of college students. Two cohorts were chosen with baseline surveys being administered to the first BPS cohort (the 2000 cohort) immediately before it began its freshman year in the fall of 2000 and baseline surveys being administered to the second BPS cohort (the 2001 cohort) immediately before it began its freshman year in the fall of 2001. In addition to collecting detailed background information about students and their families, the baseline surveys were designed to take advantage of recent advances in survey methodology in order to collect information about students' preferences and expectations towards uncertain future events and outcomes (e.g., academic performance, labor market outcomes, non-pecuniary benefits of school, marriage and children) that could influence decisions. Substantial follow-up surveys that are administered at the beginning and end of each subsequent semester have been designed to document the experiences of students and provide information about how various factors that might influence decisions change over time.

Of direct relevance to the analysis in this paper, a sequence of time-use surveys are administered at multiple times during each academic year. Also of relevance, the baseline and follow-up surveys collect substantial information about friends, roommates and other information related to study time. Student identifiers allow the survey data to be merged with Berea College's administrative data.

Section III. The Instrumental Variable estimator

Section III.1 The equation of interest

Our equation of interest is

(1) $GPA_i = \alpha_0 STUDY^*_i + \alpha_1 X_i + u_i$.

The dependent variable is the first semester grade point average (GPA) of student i. $STUDY^*_i$ is the average number of hours that a person studies per day over all of the days in the first semester. X_i contains a constant, and exogenous characteristics of student i, including a MALE indicator variable, a student's score on the American College Test (ACT), and an indicator of whether the student is BLACK. u_i represents unobserved individual determinants of the grade performance of person i. It contains, for example, information about other behaviors such as class attendance that influence grade performance, unobserved measures of ability, and whether the person has good or bad "luck" in a particular semester.

Two problems are potentially present in the estimation of equation (1). First, an errors-in-variables problem arises because $STUDY^*_i$ is not observed in the data. What is observed is $STUDY_i$, a noisy proxy for $STUDY^*_i$ which is created by averaging the number of hours that a person studies per day over the subset of days during the semester that his/her study effort is observed. During the first semester, daily study effort was collected on four different weekdays using the time diaries that are shown at the end of the Appendix. Response rates were relatively high on these surveys; the median person in our sample described below answered all four surveys and the average number of responses was 3.11. Second, STUDY* is potentially correlated with the unobservable u because decisions about how much to study in a particular semester may depend on permanent factors such as a student's unobserved ability or may depend on information that the student receives about his/her luck in that semester.

III.2 The Instrumental Variable estimation strategy

In theory, IV estimation represents a desirable way to deal with the two issues above. However, in practice, finding a valid instrument that influences a student's study effort and also satisfies the exogeneity condition of being uncorrelated with the unobservable in equation (1) is a very difficult task. The instrumental variable that we use here is inspired, in part, by recent research that uses randomly assigned (or conditionally randomly assigned) roommates to study peer effects and is made possible by specific questions that were added to the BPS for this purpose.[4]

Specifically, we instrument for STUDY* in equation (1) with a variable, which we refer to generically as TREATMENT, that indicates whether a student's roommate brought any type of video game with him/her at the beginning of the school year. The survey question which asked whether a student's roommate brought video game(s) to school appeared for the first time in our surveys in the fall of 2001. As a result, we focus on the BPS cohort that entered Berea as freshmen in 2001. As discussed later, the validity of our instrument takes advantage of the fact that students at Berea who do not request roommates are unconditionally randomly assigned roommates.[5] Slightly more than 1/3 of students at Berea either live off campus or request a roommate. The sample used in this paper contains information about 210 students who live on campus and were assigned roommates. Table 1 shows that 53% of males and 24% of females have

---

[4] See Sacerdote (2001), Zimmerman (2003), S&S (2000,2004), Kremer and Levy (2003), and Foster (2003).

[5] Freshmen at Berea are not asked to complete a housing preference questionnaire and are simply placed in available rooms without reference to preferences, backgrounds, or academic ability. There seems to be a belief at Berea that housing preference questions are limited in usefulness due to misreporting of behaviors such as smoking. As evidence of the school's intention to randomly assign rooms, in at least some years roommates were assigned using a random assignment software that exists on the campus computer system.

roommates that bring some sort of video game(s) to school.

It is worth noting that our sample size is small given the decrease in precision (relative to OLS) that can be expected to accompany the IV estimator. As a concession to the small sample size, we combine males and females when we apply the IV estimator. While this is perhaps less than ideal, we present information in the following sections that it is a generally reasonable thing to do.

In the next two subsections we examine the conditions that are necessary for our instrument to be valid.

Does the Instrument Influence Study Decisions?

With respect to the first condition that a valid instrument must satisfy, the descriptive statistics in the second row of Table 1 show that, for both males and females, study effort differs in a quantitatively important manner between students in the sample whose roommates bring video games to school and students in the sample whose roommates do not bring video games to school. Specifically, the sample average of STUDY is .667 lower (2.924 vs. 3.591) for males who receive the video game treatment than for males who do not receive the treatment. The sample average of STUDY is .467 lower (3.226 vs. 3.693) for females who receive the video game treatment than for females who do not receive the treatment. It is not possible to reject the null hypothesis that the effect of the treatment is the same for males and females.

Pooling the male and female observations we estimate a first stage regression of the form

(2) $STUDY_i = \beta_0 TREATMENT_i + \beta_1 X_i + \nu_i$.

and show the results in the first column of Table 2. As expected given the random assignment of the treatment, for both males and females the sample means of ACT and BLACK in Table 1 are very similar for students who receive the treatment and those that do not receive the treatment.[6] Thus, it is not surprising that the estimated effect of a roommate bringing a video game in equation (2) falls between the differences in sample means for males (.667) and the difference in sample means for females (.467) described in the previous paragraph. Specifically, we find an estimate (std. error) of −.565 (.241) which indicates that the treatment reduces study time by over half an hour per day. Given that students in the sample study 3.48 hours per day on average, the estimated effect is quantitatively important, and a test of the null hypothesis that the treatment has no effect on study effort is rejected at all levels of significance greater than .02.

Does the Instrument Satisfy the Exogeneity Requirement?

With respect to the second condition that a valid instrument must satisfy, in this context it must be the case that the instrument's only influence on a student's grade performance comes through its effect on the student's study effort. There are two avenues through which this exogeneity requirement could be violated. First, it would be violated if the treatment itself contains information about a student's unobserved characteristics. Second, it would be violated if the treatment influences how well a person performs academically at a particular study level. It is worth noting at this point that roommates who bring video games to school may be different in observable and unobservable ways than those who do not. As a result, in thinking about the two avenues above through which the second condition could be violated, it is necessary to take into account that the treatment involves both the physical presence of the video game(s) and the presence of whatever type of roommate accompanies the game(s).

---

[6] The null hypothesis that ACT is the same in the population for males (females) who receive treatment and males (females) who do not receive treatment cannot be rejected for any significance levels less than .46 (.37). The null hypotheses that the proportion of students that are BLACK is the same in the population for males (females) that receive treatment and males (females) that do not receive treatment cannot be rejected for any significance levels less than .25 (.37). The proportion of males in the population who receive the treatment is not expected to be the same as the proportion of females in the population who receive the treatment because males and females are not assigned to the same rooms.

The random assignment of roommates in our sample plays the key role in ensuring that the exogeneity condition is not violated by the first avenue described in the previous paragraph. If students were choosing roommates, they would also (perhaps quite indirectly) be choosing whether roommates bring video games. In this case, the amount that a student intends to study and other factors such as the student's ability could be related to whether his roommate brings a video game. The random assignment of roommates guarantees that, conditional on a student's sex, students in the sample who receive the treatment come from the same population distribution as students in the sample who do not receive the treatment.[7]

With respect to whether the exogeneity condition could be violated through the second avenue described above, there seem to be two general possibilities. One possibility is that, in addition to reducing the amount of time spent studying, students who receive the treatment also reduce the time spent in other activities, such as class attendance and sleeping, that potentially influence grade performance directly. A second possibility is that, in addition to reducing the amount of time spent studying, the treatment causes students to study less efficiently than other students. The BPS includes questions that were designed to allow the examination of these two possibilities. We do this in the remainder of this subsection.

In the next three paragraphs we examine the possibility that students who have reduced study effort as a result of the treatment have also reduced the time they spend on other activities, such as class attendance and sleep, that may have direct effects on grade performance. With respect to class attendance, our knowledge of institutional details at Berea suggests that the treatment would have little effect at Berea. Unlike many other schools, class attendance is to a large degree mandatory at Berea and this expectation is made very clear to students. Many faculty members impose strict attendance policies and faculty

[7]This assumes that randomly assigned roommates do not coordinate on what to bring to school before the year begins. This does not seem like a particularly problematic assumption. These roommates do not know each other before the school year begins. In addition, it seems likely that students will want to bring their own video games (and their own computers in the case where the video game is on a computer) regardless of what their roommates are bringing to school. Further, in our data we do not reject the null hypothesis that there is no relationship between whether a student brings a video game and whether his roommate brings a video game..

typically either formally or informally keep track of attendance of individual students. Thus, we expect that attendance would be very high for both students who receive the treatment and those who do not. We can check this empirically. At four times during the first semester, we used Question A in the Appendix to elicit information about the number of times in the previous seven days that a student's classes were schedule to meet and the number of these classes that the student attended. For each student we compute the proportion of classes that he/she attended across all time-use surveys that he/she completed. In column 1 of Table 3 we regress this proportion, PATTEND, on TREATMENT and SEX. The estimated effect (std. error) of TREATMENT is -.014 (.009). Thus, the estimated effect is not significant at .10 and is quantitatively very small; the treatment decreases attendance by only 1.4 percentage points or just slightly more than 1.4 percent given an overall average attendance rate of approximately .96. We can also provide information about whether the treatment affects class attendance by using information from our time diaries. For each student we construct a CLASSHOURS variable in a manner that is analogous to how the STUDY variable is calculated - by averaging the number of daily hours a person reports being in class over all of the time-use diaries. The regression of CLASSHOURS on TREATMENT and SEX in column 2 of Table 3 indicates that students spend approximately three and one-half hours per day in class and that the treatment has a quantitatively small and statistically insignificant effect on class attendance.[8]

With respect to the number of hours of sleep, we did not have a strong prior about what to expect. Using our time diaries we construct the variable SLEEP in a way that is directly analogous to the way that the variable STUDY is constructed. The third column of Table 3 shows the results from a regression of SLEEP on TREATMENT and MALE. The estimated effect (std. error) of TREATMENT is .275 (.208).

---

[8]It seems reasonable to assume that the treated and non-treated students have similar numbers of classes and this assumption is supported by evidence from the first part of Question A in the Appendix. On average, students who receive the treatment report that their classes were scheduled to meet 14.40 hours in the previous seven days. On average, students who do not receive the treatment report that their classes were scheduled to meet 14.10 hours in the previous seven days. A test that the number of scheduled classes is the same in the population for treated and non-treated students cannot be rejected at significance levels less than .44.

Thus, the effect is not statistically significant and indicates that students in the sample who receive the treatment sleep approximately fifteen minutes more per night than students in the sample who do not receive the treatment. We also use our time-diaries to construct a variable BEDTIME that indicates the time at which a student goes to bed. This variable is created such that positive values indicate the number of hours after midnight and negative values indicate the number of hours before midnight. Column 4 of Table 3 shows a regression of BEDTIME on TREATMENT and MALE. We find that, on average, students go to bed between 12:45 and 1:00, and we find no evidence that the treatment influences BEDTIME.

Overall, these results imply that, while the treatment leads to substantial decreases in study effort, it has very little effect on other activities that might influence grade outcomes. There is an additional survey question that can help support this conclusion. At the end of the first semester, we asked each student how much time he/she spent playing video games in an average week during the semester. On average, students in the treatment group reported playing 4.06 hours a week and non-treated students reported playing only .79 hours per week. Given that the treatment reduces study time by approximately .5 hours per day, these numbers are remarkably consistent with the notion that the treatment is having little effect on other activities. In addition, this information provides direct evidence that study time is lower for the treatment group because students are playing games. A test that there is no difference in game playing between students who receive the treatment and students who do not receive the treatment yields a t−statistic of 3.54 and is rejected at all traditional significance levels.

In the remainder of this subsection we examine the possibility that the treatment not only causes students to reduce their study effort but also somehow causes studying to be less efficient or class attendance to be less worthwhile. This possibility could be of relevance if the presence of video games in rooms implies that the student may not be able to study in the room when he/she want to because, for example, the room has become a place where others congregate. We can examine this possibility using question B in the Appendix which asked students about the physical locations where they studied. We find no difference in study locations for those who received the treatment and those who did not. In column 1 of Table 3b we regress the percentage of study time that takes place in the dorm room on TREATMENT and MALE. The estimated effect of TREATMENT is not statistically significant. Since some video games are played on televisions, treated students may be more likely to have a television in their room and one might worry that treated students may spend a higher percentage of time studying with the television on. We do not find any evidence that this is the case in column 2 of Table 3b where we regress the percentage of time spent studying with television on TREATMENT and MALE. Similarly, since some video games are played on computers, treated students may be more likely to have a computer in their room and this could represent an academic advantage for treated students. In column 3 of Table 3b we regress the number of hours per week that a student uses a computer for academic reasons on TREATMENT and MALE. Students in the sample whose roommates bring video games report that they use the computer for academic reasons about one extra hour per week than non-treated students in the sample, but the estimated effect of TREATMENT is not statistically significant.

The possibility that students who receive the treatment are studying less efficiently could also be of relevance if treated students have roommates who are less able or less willing to help them directly with their coursework. However, S&S (2004) discuss in depth the avenues through which roommates could transmit peer effects and conclude that, in the short-run, peer effects are much more likely to be transmitted by good role models influencing the time-use decisions of their roommates than by high ability students helping their roommates understand their coursework.[9] Further, at least in terms of college entrance exam

[9]There are many reasons for this conclusion. One issue is that it may be quite costly for students to help each other given that may not be taking the same classes with the same faculty members. We find empirical evidence that, while roommates spend considerable amounts of time together, they spend little time "studying or discussing course material."This is also the conclusion of Kremer and Levy (2003) who conclude that "Overall, these findings are more consistent with models in which peers change preferences than models in which they change endowments."

scores, there is no evidence that treated roommates have higher ability roommates than non-treated roommates.[10] In short, it seems highly unlikely that grade differences between treated and non-treated students are being driven in a non-trivial manner by differences in help with coursework from roommates.

Finally, students who receive the treatment might also perform less efficiently at a particular level of study effort if the treatment makes them more likely to drink alcohol and this affects cognitive skills. While the prevalence of drinking is quite low at Berea, it is worth examining this issue. This is possible because our time diaries contain a category "partying." Column 4 of Table 3 shows a regression of the number of hours spent partying on MALE and the TREATMENT. On average, students spend only about ten minutes a day partying, and we find no evidence of a relationship between the number of hours spent partying and whether a person receives the treatment. Approximately 85% of all students do not report any partying on any of the time-use surveys and this percentage also does not vary in a meaningful way with whether a person's roommate brought a video game. While we were not surprised by the low prevalence of weekday drinking, it is at least possible that some students are wary of reporting this information on their time diaries. Nonetheless, our intuition is that, if substantial differences in drinking behavior exist between the treated and non-treated students, these differences would reveal themselves in, for example, the variable BEDTIME. Further, while Kremer and Levy (2003) find that a student is more likely to drink if he/she is assigned a roommate that drinks, there is no strong reason to think that students who bring video games to school are more likely to drink and there is no evidence in the time diaries that this is the case.[11]

[10]When we estimate a linear regression of a dependent variable which indicates whether a person brought a video game to school on ACT and MALE the estimated effect (std. error) on ACT is .526 (.534). Thus, holding sex constant, students in the sample who bring video games have average ACT scores that are one-half of a point higher than students who do not bring video games.

[11]Including a variable which indicates whether a person brought a video game is found to have no effect in column 4 of Table 3b. The proportion of people who bring video games who report drinking on at least one time-use survey, .854, is virtually identical to the proportion of students who do not bring video games, .851.

While it is never possible to empirically establish that an instrument satisfies the condition of being exogenous, the unique features of the BPS data allow us to examine the reasons that this condition might be violated, and we find no evidence that this is the case. Thus, it seems very likely to us that the instrument satisfies the exogeneity condition, and we assume that this is the case in the remainder of the paper.

IV. Results

Column 1 of Table 4 shows Ordinary Least Squares estimates of equation (1) which ignore both the endogeneity and errors-in-variables problems discussed earlier. While a test of the null hypothesis that studying has no effect on grades is rejected at significance levels of greater than .05, the estimated effect is quantitatively quite small with a one hour increase in daily study-time increasing first semester GPA by only .049.

The intuition about how the IV estimator achieves identification is straightforward with the binary instrument. The validity of the instrument implies that, conditional on sex, all factors other than study-effort that influence grade performance are identical for treated and non-treated students in the population. Thus, if studying has no effect on grade performance, grade performance would be identical (conditional on sex) for the treated and untreated groups even though study-effort is different between the groups. As can be seen in the second row of Table 1, males in the sample who receive the treatment have grades that are .239 lower than males who do not receive the treatment and females in the sample who receive the treatment have grades that are .128 lower than females who do not receive the treatment. The size of the IV estimate takes into account the differences in average study-effort that led to these differences in average grades. So, for example, given that the treatment reduces study-effort by .667 of an hour for males, a Wald estimate of the effect of studying on GPA obtained from the sample of males would be .239/.667=.358.

Similarly, a Wald estimate of the effect of studying on GPA obtained from the sample of females would be .128/.467=.274.

Formal IV estimates are shown in column 2 of Table 4. As noted earlier, while it would perhaps be desirable to estimate the model separately for males and females, this is problematic given our small sample. However, the earlier evidence that it is not possible to reject the null hypothesis that the treatment has the same effect on the study-effort of males and females along with the evidence in the previous paragraph that Wald estimates are similar for males and females suggests that pooling males and females is generally reasonable. For the pooled IV estimation it is important to include MALE as a regressor that is included in X because students are randomly assigned conditional on sex. We also include the variables ACT and BLACK in X both because understanding the importance of these variables is useful for interpreting the estimated effect of studying and because, even with random assignment of the treatment, the values of these variables could vary to a small degree between the treated and untreated groups due to sampling variation associated with our small sample.

The IV estimate indicates that an additional hour of studying per day causes first semester grade point average to increase by .356.[12] Although, as expected the effect is estimated with much less precision under IV than under OLS, a test of the null hypothesis that studying has no effect on grade performance produces a t-statistic of 1.748 and the test is rejected at significance levels greater than .08.[13]

The IV estimate, .356, is much larger than the OLS estimate, .049. Part of this difference arises because of the errors-in-variables problem from using STUDY instead of STUDY* in equation (1). As

---

[12]In Table 2 we found that, conditional on the other included covariates, the treatment decreases average study hours by .564. The second column of Table 2 shows that, conditional on the other covariates, students in the treatment group receive grades that are .201 lower than students in the untreated group (and a test of the null hypothesis that the treatment has no effect on grades is rejected at significance levels greater than .02). Thus, the IV estimate is .201/.564.

[13]We also estimated a model which added as regressors all of the dependent variables in Table 3a and Table 3b. The estimated effect (std. error) in this specification was .375 (.23).

discussed in S&S (2004), the OLS estimator would need to be multiplied by a factor of

$$(3)\quad \frac{\mathrm{Var(STUDY)}}{\mathrm{Var(STUDY)} - \frac{\sigma_v^2}{N}}.$$

to correct for this problem, where σ_v^2 is the variance of the unobservable in equation (2) and N is the number of time-use surveys. It is difficult in our case to know exactly what the bias factor is since N is not constant across people. However, using equation (3) we ascertain that the bias factor is between 1.40 and 1.94.[14] Thus, the difference between the IV and OLS estimates remains after accounting for the errors-in-variables problem is between .260 and .287.

As discussed earlier, the direction of the bias due to the endogeneity problem is uncertain from a theoretical standpoint. However, the fact that the IV estimate is much larger than the OLS estimate suggests that either students who study more hours are likely to be of lower ability or that students increase their effort in a particular semester in response to grade shocks that are received in that semester. While the potential importance of unobserved ability makes it impossible to provide conclusive evidence about the first possibility, one gets a sense that this might not be the driving influence from examining the results in the first column of Table 2 which reveal no evidence of a relationship between our observable measure of ability (ACT) and study effort.

This suggests that the difference between the IV and OLS estimates might arise because students adjust their effort in a particular semester in response to information about grade shocks that is received in that semester. The presence of a second semester of grade and study-effort information presents us with an opportunity to examine whether there is evidence in the data of this type of behavior. We design a test

[14] An estimate of σ_v^2 can be constructed by differencing the individual daily study reports for a particular person. Estimates of VAR(STUDY) can be computed conditional on N from the sample. 1.40 is an estimate of the factor by which the OLS estimator would be biased if all students answered four time-use surveys. 1.94 is an estimate of the factor by which the OLS estimator would be biased if all students answered only one time-use survey.

that takes advantage of the fact that, while study effort in the first semester may be correlated with the transitory component of grades in the first semester, it should be uncorrelated with the transitory component of grades in the second semester under the assumption that the transitory portion of grades is uncorrelated across time. This implies that the grade difference between the second and first semesters, averaged over all people who studied a particular amount in the first semester, will be larger if this group experienced bad luck on average in the first semester.

To be more specific about this test, it is worthwhile to disaggregate the unobservable in equation (1) into a person-specific, permanent component μ_i and a transitory component ε_{ti} that is assumed to be serially uncorrelated

(4) $u_{ti} = \mu_i + \varepsilon_{ti}$.

Equation (1) represents a model in which grades are generated by a a study component, $\alpha STUDY^*_i$, a permanent ability component, $\beta X_i + \mu_i$, and a transitory or luck component, ε_{it}. At this point we rename variables slightly to differentiate between the first and second semesters.The grade equation for semesters one and two are given by equations (5) and (6) respectively

(5) $GPA_{1i} = \alpha_0 STUDY_1{*}_i + \alpha_1 X_i + \mu_i + \varepsilon_{1i}$

(6) $GPA_{2i} = \alpha_0 STUDY_2{*}_i + \alpha_1 X_i + \mu_i + \varepsilon_{2i}$

Differencing equation (6) from equation (5) and rearranging yields

(7) $GPA_{1i} - GPA_{2i} - \alpha_0(STUDY_1{*}_i - STUDY_2{*}_i) = \varepsilon_{1i} - \varepsilon_{2i}$.

Thus, the left hand side of equation (7) represents the difference in a person's transitory component or "luck" between the two semesters.[15] For illustrative purposes, consider a case where there are only two

[15]Of course, we do not intend to imply that all variation in the transitory components should necessarily be interpreted literally as "luck." For example, while students at Berea have rather limited flexibility about the classes they take during the first year due to a large number of required "general studies" courses, it is possible that some of the changes in the transitory component across semesters could reflect differences in the difficulty of classes across semesters.

study levels in the population: $STUDY_1$*= high or $STUDY_1$*=low. Averaging the left hand side of equation (7) over all individuals who have $STUDY_1$*= high yields $E(\varepsilon_1|STUDY_1^* = high)$ since the assumption that the transitory components are uncorrelated implies that $E(\varepsilon_2 \, |STUDY_1^*=high)=0$. Similarly, averaging the left hand side of equation (7) over all individuals who have $STUDY_1$*= low yields $E(\varepsilon_1|STUDY_1^* = low)$. Comparing $E(\varepsilon_1|STUDY_1^* = high)$ to $E(\varepsilon_1|STUDY_1^* = low)$ indicates how the transitory component of grades varies, on average, across the two $STUDY_1$* amounts.

This discussion motivates our estimation of an equation of the form

(8) $GPA_{1i} - GPA_{2i} - .359\,(STUDY_{1i} - STUDY_{2i}) = constant + \delta\, STUDY_{1i} + \eta_i$.

We find an OLS estimate (std. error) for δ of -.276 (.040). This implies that students who study an extra hour per day have an average realization of the transitory component ε_1 that is .276 lower than otherwise similar students. Identification for the OLS estimator involves comparing the GPA of students who study an extra hour to the GPA of students who do not study an extra hour. Earlier we found that a difference of between .260 and .287 remains between the IV and OLS estimates remains after accounting for the errors-in-variables problem. The results here indicate that this remaining difference can be attributed to the finding that the average GPA of students who study an extra hour per day would be .276 lower than the average GPA of students who do not study the extra hour under the counterfactual in which both groups study the same amount.

These results sound a cautionary alarm about the use of fixed effects estimators. In this application, a fixed effects estimator would achieve identification using the within person variation in study effort across the two semesters. However, our results indicate that assuming that this variation is exogenous is extremely problematic. In addition, the evidence that ACT scores are unrelated to study effort suggests that the variation in study effort across people, which is discarded by the fixed effect estimator, may be less likely to suffer from problems of endogeneity. As a result, not only is the use of a Fixed Effects estimator unlikely to satisfactorily deal with the endogeneity problems, but the Fixed Effects estimator may perform worse than the OLS estimator. Striking evidence that this is the case is shown in column 3 of Table 4. The estimated effect of studying, -.043, is negative, and a test of the null hypothesis that studying has harmful effect on grades cannot be rejected at levels of significance greater than .10.

Section V. Conclusion

To the best of our knowledge, this study represents the only evidence about the causal relationship between study effort and grade production. The results are consistent with recent literature such as S&S (2003) that found that working an extra hour per day in paid employment during college has a large causal effect on grade performance. Importantly, the results suggest that human capital accumulation is far from predetermined at the time of college entrance. For example, an increase in study effort of one hour per day (an increase of approximately 2/3 of a standard deviation in our sample) has the same effect on grades as a 5.74 point increase in ACT scores (an increase of 1.54 standard deviations in our sample).

The IV estimate is much larger than the OLS estimate. Evidence in the paper indicates that this difference arises primarily because students increase their effort during semesters in which the transitory portion of grades is low rather than because of individual fixed effects. As a result, not only does a Fixed Effects estimator not solve the endogeneity problems, but it makes matters worse; the estimated effect of studying is negative and the null hypothesis that studying has no effect is rejected at significance levels greater than .10. Thus, the paper suggests that significant caution should be taken when considering the use of Fixed Effects estimator in cases where behavioral responses to information may be present.

Finally, while not the focus of this paper, this paper also makes an important contribution to the peer effects literature in general and to the peer effects literature that achieves identification by using college roommates in particular. The goal of the empirical peer effects literature has been to look for empirical evidence which documents that peer effects can matter. This paper provides depth to that literature by not only providing evidence that peer effects can matter, but by providing perhaps the first direct evidence about the avenues through which peer effects operate in a particular educational context.

References

Betts, 1997, The role of homework in improving school quality, discussion paper 96-16, University of California at San Diego.

Card, David. 1999. "The Causal Effect of Education on Earnings." In Handbook of Labor Economics, volume 3A.

Orley Ashenfelter and David Card, eds. New York: Elsevier Science.

Foster, Ruth (2003): “Peerless Performers: The Absence of Robust Peer Effects at a Large, Heterogeneous University, working paper.

Kremer and Levy (2003): “Peer Effects and Alcohol Use Among College Students,” working paper.

Ruhm, Chris, 1997, Is high school employment consumption or investment, Journal of Labor Economics 15(4) (October), 735-776.

Sacerdote, Bruce (2001): “Peer Effects with Random Assignment: Results for Dartmouth Roommates,” Quarterly Journal of Economics, volume 116.

Stern, David and Yoshi-Fumi Nakata, 1991, Paid employment among U.S. college students: Trends, effects, and possible causes, Journal of Higher Education 62, No. 1. (January - February), 25-43.

Stinebrickner and Stinebrickner (2000): “Peer Effects Among Students From Disadvantaged Backgrounds,” unpublished manuscript.

Stinebrickner, Ralph and Todd Stinebrickner, 2003, Working during school and academic performance, Journal of Labor Economics 21(2) (April), 473-491.

Stinebrickner, Todd R. and Ralph Stinebrickner (2004a). “Time-Use and College Outcomes,” *Journal of Econometrics*, 121(1-2), 243-269.

Stinebrickner & Stinebrickner (2004b): “What can we learn about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds,” unpublished manuscript.

Zimmerman, David (2003): “Peer Effects in Academic Outcomes: Evidence from a Natural Experiment,” *The Review of Economics and Statistics,* 85(1), 9-23

Table 1
Descriptive Statistics

	Male All n=95	Male treatment =0 n=45	Male treatment =1 n=50	Female All n=115	Female treatment =0 n=88	Female treatment =1 n=27
TREATMENT - Roommate brought a video game to school	.526			.235		
STUDY	3.240 (1.688)	3.591 (1.748)	2.924 (1.583)	3.583 (1.573)	3.693 (1.595)	3.226 (1.473)
GPA - First semester Grade Point Average	2.853 (.677)	2.979 (.663)	2.740 (.677)	3.129 (.605)	3.159 (.598)	3.031 (.628)
ACT	22.463 (3.842)	22.155 (3.931)	22.740 (3.779)	24.139 (3.431)	24.205 (3.527)	23.925 (3.149)
BLACK	.189	.200	.180	.157	.159	.148

Table 2
The direct effect of treatment on study hours (column 1) and grades (column 2)

Independent Variable	**Dependent Variable STUDY hours per day estimate (std error)**	**Dependent Variable GPA first semester grades estimate (std error)**
CONSTANT	3.912 (.241)*	1.717 (.313)*
TREATMENT	-.564 (.241)*	-.201 (.087)*
MALE	-.211 (.239)	-1.07 (.086)
ACT	-.011 (.034)	.062 (.012)
BLACK	-443 (.329)	-.196 (.119)
	R^2=.051	R^2=.226

*significant at .10

Table 3a
The effect of treatment on other behaviors

Independent Variable	**Dependent Variable PATTEND proportion of classes attended** **estimate (std. error)**	**Dependent Variable CLASSHOURS daily hours in class** **estimate (std. error)**	**Dependent Variable SLEEP daily sleep hours** **estimate (std. error)**	**Dependent Variable BEDTIME time student went to sleep**** **estimate (std. error)**
TREATMENT	-.014 (.009)	-.114 (.188)	.275 (.208)	.143 (.199)
MALE	.003 (.009)	.059 (.182)	.209 (.202)	-.276 (.192)
CONSTANT	.962 (.006)	3.444 (.25)*	7.089 (.138)*	.833 (.130)*
	R^2=.012	R^2=.0016	R^2=.019	R^2=.011

*significant at .10
** dependent variable is created so that it is zero at 12:00 midnight. Positive numbers represent hours after midnight. Negative numbers represent hours before midnight.

Table 3b
The effect of treatment on additional behaviors

Independent Variable	**Dependent Variable percentage of study time that takes place in dorm room** **estimate (std. error)**	**Dependent Variable percentage of study time that takes place in dorm room with tv on** **estimate (std. error)**	**Dependent Variable hours per week using computer for academic purposes** **estimate (std.error)**	**Dependent Variable daily hours partying** **estimate (std. error)**
TREATMENT	-2.111 (4.670)	3.515 (2.933)	.963 (1.069)	.007 (.050)
MALE	-4.677 (4.498)	-3.812 (2.825)	-.254 (1.032)	-.015 (.048)
CONSTANT	61.456 (3.058)*	12.756 (1.921)*	6.820 (.699)*	.125 (.033)*
	R^2=.008	R^2=.008	R^2=.012	R^2=0.011

*significant at .10

Table 4
Estimates of the effect of studying on grade performance

Independent Variable	**Ordinary Least Squares**	**Instrumental Variables**	**Fixed Effects**
	estimate (std. error)	**estimate (std. error)**	**estimate (std. error)**
CONSTANT	1.494 (.025)*	.322 (.880)	-.050 (.047)
STUDY	.049 (.025)*	.356 (.203)*	-.043 (.027)*
SEX	-.148 (.083)*	-.031 (.134)	
ACT	.062 (.012)*	.065 (.017)*	
BLACK	-.216 (.120)*	-.354 (.182)*	
	R^2=.221		R^2=.014

*significant at .10

Appendix: Survey questions

Survey Question A.

In the last 7 days (one week), how many times were your classes scheduled to meet?_____
Please count up carefully the number of scheduled class meeting for each one of the seven days and add them together. (If your schedule for a particular day included one math class meeting, one GST class, a biology lab, and a music class you would count 4 for that day. Add together these numbers for each day to get a total for the week.

How many of these classes did you actually attend? ________

Survey Question B.

We are interested in where you studied. For a typical week during the Fall semester, tell us the percentage of your study time that took place in each of the following places.
Note: Numbers on the five lines should add up to 100

In dorm room (or at home if live off campus) with TV on	_______
In dorm room (or at home if live off campus) without TV on	_______
In library, empty classroom, quiet study lounge, or other quite place	_______
In TV lounge, other (non-quiet) lounges	_______
Other places	_______

Question A on the survey asks that you carefully fill out a time diary which is a list of activities during the past **24 hours**. In order to complete the time diary on the actual survey form on **page 3**, do the following:

1) Please put an arrow (-->) next to the time that it is right now. Label this arrow with the words **YESTERDAY** and **START**.

2) Now start with the box next to which you put the arrow (-->). Place in this box the activity you were doing during that time period yesterday.

For example, if it is now 7 p.m., you would put an arrow (-->) next to the box labeled "7:00PM".

Next to 7:00PM, you should write what you were doing from 7:00 to 7:20 **yesterday**.

Next to 7:20PM, you should write what you were doing from 7:20 to 7:40 **yesterday**.

Next to 7:40PM, you should write what you were doing from 7:40 to 8:00, and so forth.

As you proceed, you should work down the column below your arrow (-->) and then move to the top of the other column. Complete this other column and then move back to the top of the column where you started and finish filling in until you reach the arrow(-->).

When you begin to fill in the time period boxes, you will be writing your activities from yesterday until you reach the box labeled 12:00 midnight. From then on, you will be writing about your activities earlier today.

A sample completed time diary

Time Period	What were you doing?	Time Period	What were you doing?
MORNING		EVENING	
6:00 AM	SLEEPING	6:00 PM	
6:20 AM		6:20 PM	EATING
6:40 AM	PERSONAL	6:40 PM	
7:00 AM	EATING	7:00 PM	
7:20 AM		7:20 PM	SHOPPING
7:40 AM		7:40 PM	
8:00 AM	IN	8:00 PM	
8:20 AM	CLASS	8:20 PM	
8:40 AM		8:40 PM	
9:00 AM	WORKING (Labor)	9:00 PM	
9:20 AM		9:20 PM	STUDYING
9:40 AM		9:40 PM	
10:00 AM	IN	10:00 PM	
10:20 AM	CLASS	10:20 PM	RECREATION AND
10:40 AM		10:40 PM	STUDYING
11:00 AM		11:00 PM	
11:20 AM	WORKING (Labor)	11:20 PM	
11:40 AM		11:40 PM	STUDYING
AFTERNOON		NIGHT	
12:00 noon	EATING	12:00 midnight	
12:20 PM		12:20 AM	
12:40 PM		12:40 AM	
1:00 PM	IN	1:00 AM	
1:20 PM	CLASS	1:20 AM	
1:40 PM		1:40 AM	
2:00 PM		2:00 AM	
2:20 PM	EXERCISING	2:20 AM	SLEEPING
2:40 PM		2:40 AM	
3:00 PM		3:00 AM	
3:20 PM	STUDYING	3:20 AM	
3:40 PM		3:40 AM	
4:00 PM	IN	4:00 AM	
4:20 PM		4:20 AM	
4:40 PM	CLASS	4:40 AM	
5:00 PM		5:00 AM	
5:20 PM		5:20 AM	
5:40 PM		5:40 AM	

START ← YESTERDAY

Note(1): The activities will be chosen from the **13 words in BOLD** which are listed on **page 3** to the right of the time diary form that you will complete

Note(2): Notice in the example that the brace symbol (}) is used when an activity continues through several time periods.

Note(3): If you are involved in two activities during the same time period(s), please list both activities and circle the activity you spent more time on.

Studying (outside of class)

includes studying for your classes, preparation for class, studying for an exam, doing take-home exams, **homework**, writing essays and papers, optional study sessions, any other work done outside of class time for your classes.

Survey #5 (Please complete both sides of this sheet) CPO 1971 (3)

Question A.

Reminders: Be sure to put an arrow (-->) next to the time that it is right now. And label this arrow with the words **YESTERDAY** and **START**.

Beginning with the **What were you doing** box next to the arrow, fill in your activities starting 24 hours ago (yesterday) and ending right before you began completing this survey.

Please use the 13 words listed in BOLD on the right of this page to describe your activities.

Time Period	What were you doing?	Time Period	What were you doing?
MORNING		EVENING	
6:00 AM		6:00 PM	
6:20 AM		6:20 PM	
6:40 AM		6:40 PM	
7:00 AM		7:00 PM	
7:20 AM		7:20 PM	
7:40 AM		7:40 PM	
8:00 AM		8:00 PM	
8:20 AM		8:20 PM	
8:40 AM		8:40 PM	
9:00 AM		9:00 PM	
9:20 AM		9:20 PM	
9:40 AM		9:40 PM	
10:00 AM		10:00 PM	
10:20 AM		10:20 PM	
10:40 AM		10:40 PM	
11:00 AM		11:00 PM	
11:20 AM		11:20 PM	
11:40 AM		11:40 PM	
AFTERNOON		NIGHT	
12:00 noon		12:00 midnight	
12:20 PM		12:20 AM	
12:40 PM		12:40 AM	
1:00 PM		1:00 AM	
1:20 PM		1:20 AM	
1:40 PM		1:40 AM	
2:00 PM		2:00 AM	
2:20 PM		2:20 AM	
2:40 PM		2:40 AM	
3:00 PM		3:00 AM	
3:20 PM		3:20 AM	
3:40 PM		3:40 AM	
4:00 PM		4:00 AM	
4:20 PM		4:20 AM	
4:40 PM		4:40 AM	
5:00 PM		5:00 AM	
5:20 PM		5:20 AM	
5:40 PM		5:40 AM	

LIST OF WORDS in bold

In Class
Attending class, attending labs, attending required class sessions

Studying (Outside of class time)
(refer to **pg 2** for more details)

Athletics
(Intercollegiate or Intramural - games or practice)

Clubs

Exercising

Recreation
(reading which is unrelated to courses, listening to music, watching movie, spending time with friends, etc.)

Shopping

Eating

Sleeping

Partying

Personal

Working (in Labor position)

Other
(Please describe on your sheet)