

2001

2001-2 Time Use and College Outcomes

Todd R. Stinebrickner

Ralph Stinebrickner

Follow this and additional works at: <https://ir.lib.uwo.ca/economicscibc>



Part of the [Economics Commons](#)

Citation of this paper:

Stinebrickner, Todd R., Ralph Stinebrickner. "Time Use and College Outcomes." CIBC Centre for Human Capital and Productivity. CIBC Working Papers, 2001-2. London, ON: Department of Economics, University of Western Ontario (2001).

ISSN: 1702-0085
ISBN: 0-7714-2354-3

Time-Use and College Outcomes

by

Todd R. Stinebrickner and Ralph Stinebrickner

Working Paper # 2001-2

December 2001



CIBC Working Paper Series

Department of Economics
Social Science Centre
The University of Western Ontario
London, Ontario, N6A 5C2
Canada

This working paper is available as a downloadable pdf file on our website
<http://www.ssc.uwo.ca/economics/centres/cibc/>

Time-Use and College Outcomes

Todd R. Stinebrickner
Dept. of Economics
Social Science Centre
The University of Western Ontario
London, Ontario Canada n6a 5c2
trstineb@julian.uwo.ca
519 662-2111 ext. 85293
fax: 519 661-3666

Ralph Stinebrickner
Berea College

Despite an increased awareness of the policy importance of understanding the determinants of education outcomes, knowledge of the relationship between education outcomes and perhaps the most basic input in the education production process - students' study time and effort - has remained virtually non-existent. This paper examines this issue using unique longitudinal data from Berea College.

I. Introduction

Despite an increased awareness of the policy importance of understanding the determinants of education outcomes, knowledge of the relationship between education outcomes and perhaps the most basic input in the education production process - 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.

The current lack of knowledge about the relationship between time inputs and higher education outcomes is almost certainly a result of the cost and difficulty of collecting appropriate data. In most cases, researchers will be examining college outcomes such as grades or drop-out decisions that are measured at semester or yearly intervals and will be interested in time-use variables that are measured on the same time interval. Unfortunately, providing accurate measures such as the average number of hours that a person spends studying during an academic year or semester is difficult in standard cases in which survey collectors have contact with respondents only once a year.

One obvious approach in this case is to collect responses to a retrospective question of the sort, "In a typical/average week during the last academic year (semester), how many hours did you spend studying?" Unfortunately, as has been well-documented, the reporting error from retrospective questions of this sort is likely to be substantial. An alternative approach in this case is to collect information about a single time period (perhaps 24 hours) using narrowly defined questions or a time-diary. However, while it seems likely that this method will produce an accurate view of a particular time period, the collected information will represent only a noisy proxy for the desired yearly (or semester) measure of study hours given the certain presence of variation in study-time across days in the year. Correcting the estimator bias and/or incorrect measures of precision that arise when these approaches are used is problematic when there exists

no obvious means of characterizing the nature of the reporting error that is present in the retrospective question or the amount of inter-day variation that leads to measurement error in variables constructed from the single time-diary.

In this paper, we examine a unique data set that we have created by merging detailed data about the time-use patterns of first-year students at a liberal arts college in Kentucky (Berea College) with administrative data on the demographic characteristics, educational backgrounds, and college outcomes of these students. At six different times during the year we conducted “primary” time-use surveys which asked students in our sample to tell us how much time they had spent studying (and how much time they had spent on other activities) in both the twenty-four hour period and the seven day period that immediately preceded the time of the survey. We use these questions to derive an estimator of the relationship between time-use and student outcomes. The approach addresses the problems of estimator bias and incorrect measures of precision by utilizing simulation methods that both explicitly account for the fact that, despite our multiple surveys, we have only recorded time-use information for a subset of the total period of interest and are flexible enough to correctly deal with the additional missing data problem that is present because some respondents have only answered a subset of the six time-use surveys.

Although we initially assume that students’ answers to our “primary” time-use questions are accurate, we are also able to examine the robustness of our results to this assumption. This is possible because, for one twenty-four hour period, we collected both a primary time-use survey and a detailed time-diary. If we are willing to assume that the time-diary reflects a person’s true time-use, we can estimate the stochastic relationship between a respondent’s true time-use and his answer to our “primary” time-use survey and modify our simulation estimator to take into account the assumed noisy nature of the six primary time-use surveys.¹

¹This requires that we make an assumption that the stochastic relationship between the retrospective time-use surveys and the “truth” remains the same over the course of the year.

The paper makes both a substantive and methodological contribution. From a substantive standpoint, the paper provides one of the first explorations into the importance of effort, as measured by study-time, in the production of education. A statistically significant and quantitatively large non-linear relationship is found between a person's study-time and his/her first-year cumulative grade point average. From the standpoint of predicting first year grade point average, study-effort is found to be more important than college entrance exam scores which have traditionally been found to be perhaps the best predictor of college outcomes. It is important to mention the obvious difficulty of trying to determine causality in this context given the reality that study-effort is endogenously determined. As a result, we view our results as largely descriptive in nature and discuss some of the possible sources of correlation between study-time and the unobservable in our outcome equation. Nonetheless, especially after taking into account the possible presence of reporting error in our time-use surveys, our results bear a striking resemblance to responses to separate survey questions that elicited students' beliefs about the causal relationship between their study-effort and their grades. In addition, the types of estimates in this paper are consistent with results in Stinebrickner and Stinebrickner (2001) that established that increasing a student's paid employment during school has a large, negative causal effect on the student's grade performance. Thus, all evidence in this paper points to our substantive conclusion that future work which leads to a better understanding of how individuals make decisions about time-use or a better understanding of the impact of these decisions is extremely important from the standpoint of understanding outcomes in education.

From a methodological standpoint, this paper provides insight into the type of estimators that might be appropriate if one had access to detailed time-use information, and provides information about the types of errors a researcher might make if he/she uses estimators that are less appropriate (but perhaps easier to implement). However, given the current absence of data of this sort, the more important methodological contribution may be to provide survey administrators with guidance in thinking about incorporating time-use information into future surveys. For example, by comparing our results with results obtained using only

end-of-the-year retrospective questions we are able to provide some information about the type of mistakes that might be made by researchers if only this retrospective information were available. Similarly, by comparing our full results to results obtained using various subsets of our six time-use surveys we can provide some information about the benefit of collecting various numbers of time-use surveys. Our results indicate that it is essential to have more than one time-use survey, but that our time-use information would have been useful even if we had collected less than six time-use surveys.

Section II provides a brief review of the literature on time-use and educational outcomes. Section III discusses Berea College and the longitudinal survey that we have conducted at the school. Section IV describes the estimators and substantive results from Berea. Section V examines the methodological issues discussed above and Section VI concludes.

II. Literature Review

Literature that examines the relationship between the time-use of students and educational outcomes is scarce. 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) and, therefore, is not repeated in full here. Ehrenberg and Sherman (1992), which represents some of the earliest work involving college outcomes, found that employment in off-campus jobs during college led to a decline in academic performance, but found no negative effect of working in on-campus jobs. More recently, Stinebrickner and Stinebrickner (2001) analyzed eight years of data from a mandatory labor program at Berea College. By taking advantage of institutional details of the labor program, the authors found that it is extremely important to account for the reality that the number of hours that a person works is endogenously chosen and that working during school can have a quantitatively large and statistically significant negative impact on academic performance.²

²The paper takes advantage of the the fact that students are randomly assigned to a variety of jobs in their first year that are similar in nature and that the job to which a person is assigned has an

Employment during school is not an input into the education production function per se. Rather, it can influence academic performance because it enters into a person's budget constraint for time along with inputs such as time devoted to studying. This paper differs from virtually all previous work in its goal of examining study-inputs directly. Perhaps most similar in spirit to the objectives of this work is that of Betts (1997) who finds that the amount of homework assigned by teachers between grades seven and eleven has a quantitatively large effect on student achievement as measured by test scores.

III. Data

Berea College is located in the foothills of the Appalachian mountains. The school operates with a mission of providing an education to those who "have great promise, but limited economic resources" and provides a full tuition subsidy to all entering students (and large room and board subsidies) regardless of family income. Stinebrickner and Stinebrickner (1998) discovered that many students at Berea do not graduate despite the full tuition subsidy. The time-use data described here is part of a larger survey effort designed to provide an understanding of this finding.

The students who entered Berea in the fall of 2000 were asked to complete nine surveys over the course of their freshman year. The first survey was a written survey that took place as soon as students arrived at Berea. Approximately 81% (343 of 426) of all freshmen completed this survey which was necessary for involvement in future surveys. For purposes of this paper, we limit our attention to domestic, dependent students. There are a total of 306 students who meet this condition and completed our first survey. With respect to time-use issues, the first survey asked students questions about how many hours they had studied during their senior year in high school and how many hours they expected to study during their first year of college. Descriptive statistics for these questions are shown in Table 1.

Six of the remaining surveys were aimed primarily at eliciting information about how students were

important effect on the number of hours that a person works.

using their time. These surveys were sent to students via campus mail with strict completion deadlines. The proportion of our first-wave respondents who answered each of these time-use surveys was .90, .82, .83, .77, .75, and .75.³ With respect to study time, we asked each student to think carefully about how he/she had been spending his/her time and to report the amount of time he/she had spent studying and doing homework (outside of class time) in the following three periods: the immediately preceding twenty-four hour period (subsequently referred to as the “24-hour” report), the most recent weekend day, and the previous seven days (subsequently referred to as the “7-day” report). The timing of the delivery and deadlines of the surveys implied that all time-use surveys were completed on a weekday. Thus, the answer to the 24-hour question represented hours of study in a twenty-four hour weekday period. The weekend day question was included largely in an effort to encourage respondents to think more carefully when answering the weekly hours question. In the analysis below we concentrate primarily on results using the 24-hour reports but also briefly examine the robustness of our results to the use of the 7-day reports.

The first six rows of Table 1 show descriptive statistics associated with the weekday and weekly study questions for each of the six primary time-use surveys. The standard deviations of the entries indicate that there is a large amount of variation in study hours across individuals within a week. The seventh row of Table 1 shows descriptive statistics when each person’s reported hours are averaged over all of the time-use surveys that the person completed. The large standard deviation of this entry suggests that some students may have substantially different study habits than other students.

Additional information about the students in our sample was obtained by merging our survey data with administrative data obtained from Berea. Table 1 includes a descriptive look at the observable characteristics $X = \{\text{sex, race, family income, American College Test (ACT) scores, and parental education}\}$

³Approximately six percent of our respondents left school before the end of the first year. This explains the decline in response rates over the course of the year.

that are used in the paper.⁴ The administrative data are very complete and only eleven of the 306 dependent, domestic students who answered our first survey are missing any of the observable characteristics in X_i . Thus, our final sample consists of 295 individuals.

With respect to outcomes, we focus on a person's cumulative college grade point average during his first year (GPA). The descriptive statistics in Table 1 show that the average first year cumulative GPA is 2.93.

IV. Model and Berea Results

IV.A. Model

Our primary objective is to derive an estimator for the unknown parameters in the equation

$$(1) \quad O_i = \alpha_s \text{STUDY}_i + \alpha_x X_i + \epsilon_i$$

where O_i is the GPA outcome for person i and we have a particular interest in α_s , the relationship between studying and the grade outcome. Later we discuss some of the possible sources of correlation between STUDY_i and ϵ_i . At this point we note that dealing with this endogeneity problem is a difficult task, and, as a result, we view our primary task as providing an estimator for equation (1) which we view as being descriptive in nature.

Given this descriptive view of the problem, the primary obstacle in the estimation of equation (1) involves the measurement of STUDY_i . Because our outcome, O_i , of interest pertains to a full academic year, it is important that our measure of a person's study time, STUDY_i , also be measured over a year interval. Therefore, we define STUDY_i to be the average study amount per "period" over the course of the entire academic year.⁵ When we use the 7-day (weekly) reports of study-time, the relevant period is a week

⁴Most students at Berea take ACT exams. For students who took only SAT exams, standard conversion tables were used to convert their scores to ACT equivalents. The parental education variable is a dummy variable that is equal to one if person i has at least one parent who graduated from college.

⁵ STUDY_i could also be defined to be total yearly hours rather than average yearly hours.

and $STUDY_i$ represents the average study amount per week over the course of a year. When we use the twenty-four (weekday) reports of study-time, the relevant period is a day and $STUDY_i$ represents the average study amount per day over the course of a year. The total number of relevant periods, T , during the year also depends on whether we are using the 7-day or 24-hour data. We assume that there are 24 weeks and 120 weekdays during the academic year.⁶ Thus, when we use the 7-day reports $T=24$, and when we use the 24-hour reports $T=120$.

$STUDY_i$ is a function of the person's study-time in each of the periods of the year.

$$(2) \quad STUDY_i = (s_{i,1} + s_{i,2} + s_{i,3} + \dots + s_{i,T}) / T$$

where $s_{i,t}$ represents the amount that person i studied in week t . Equation (2) implies that we do not observe $STUDY_i$ for any of the respondents in our sample. Rather, for each person i we observe weekly study-time, $s_{i,t}$, for only some number of weeks $N_i \leq 6$. For ease of exposition, consider ordering the weekly study amounts so that $s_{i,1}, \dots, s_{i,N_i}$ are person i 's study amounts for the N_i weeks that he completed surveys and $s_{i,N_i+1}, \dots, s_{i,T}$ are person i 's study amounts for the $T - N_i$ weeks for which he did not complete surveys. As discussed in more detail in the next section, simply using the average of the observed weeks, (i.e., setting $STUDY_i = \frac{1}{N_i} \sum_{t=1}^{N_i} s_{i,t}$) in equation (1) would be straightforward but potentially problematic.

The approach taken here involves specifying the stochastic process associated with $s_{i,1}, \dots, s_{i,T}$ and using this stochastic process to explicitly account for our uncertainty about the value of $STUDY_i$ during the estimation of equation (1). More specifically, we assume that person i 's study-time in week t can be written as

$$(3) \quad s_{i,t} = C_t + \beta X_i + \mu_i + v_{i,t} \quad t=1,2,\dots,T.$$

C_t is a constant that takes into account that the average amount of time students spend studying may vary by week. For example, students may tend to have more exams at certain times during the year than at other

⁶The results in the paper are not sensitive to varying the number of weeks or number of days for the academic year in a reasonable way.

times. βX_i captures the reality that individuals with certain observable characteristics may study more on average than individuals with other types of observable characteristics. For example, students from certain type of family backgrounds may study more or less than students from different types of backgrounds. The permanent, unobserved component μ_i represents a heterogeneity term that allows for the possibility that some individuals may have a permanent propensity to study more than other individuals that is not captured by the observable characteristics in the vector X_i . The transitory component $v_{i,t}$ allows for randomness in a person's study-time in a particular week. We begin by assuming that the distributions of μ_i and $v_{i,t}$ are $N(0, \sigma_\mu^2)$ and $N(0, \sigma_v^2)$ respectively where the size of the two variance terms reflects the relative importance of unobservable permanent and random determinants of study-time. However, because the relative importance of these two unobservables has extremely important implications for our estimation, we later examine the robustness of our results to these assumptions.

Because we have time-use observations for only six of the T periods that are assumed to compose the academic year, it is not possible to estimate C_i separately for each of the T periods. To deal with this situation when we are using the 7-day reports, we make the assumption that each of the six weeks for which we observe time-use is "similar" to three other weeks during the year. That is, we assume that for each $t \leq 6$, $C_t = C_k = C_l = C_m$ for some k, l , and $m > 6$. Similarly, when we are using the 24-hour reports, we make the assumption that each of the six days for which we observe time-use is "similar" to nineteen other days during the year. Given this assumption, the parameters of equation (3) that need to be estimated are $C = \{C_1, C_2, C_3, C_4, C_5, C_6\}$, β , σ_μ^2 , and σ_v^2 .

Given the specification of $STUDY_i$ in equation (2) and the reality that s_t is not observed for all t , it is clear that estimation of equation (1) will also depend on the parameters of equation (3). Specifically, $STUDY_i$ in equation (1) can be written as a function of s_1, s_2, \dots, s_{N_i} , X_i , and the unknown parameters in equation (3). For sake of illustration, consider a person who answered all six of our time-use surveys. For this person, $N_i = 6$ and average study hours per period can be written as,

$$(4) \text{ STUDY}_i = \frac{1}{T} \sum_{t=1}^T s_{i,t} = \frac{\sum_{t=1}^6 s_{i,t} + \sum_{t=7}^T s_{i,t}}{T} = \frac{\sum_{t=1}^6 s_{i,t} + \frac{T-6}{6} \sum_{t=1}^6 C_t + (T-6)\beta X_i + (T-6)\mu_i + \sum_{t=7}^T v_{i,t}}{T}.$$

It is straightforward to adjust equation (4) to reflect STUDY_i for someone who completed only a subset of the six time-use surveys during the year.

Intuitively, the precision with which we will be able to estimate the relationship between study-time and a particular outcome depends on the amount of variation that exists in STUDY_i . We note from equation (4) that differences in the observed $s_{i,t}$'s across people will create some variation in STUDY_i across people. However, because $s_{i,t}$'s are observed for a relatively small subset of the periods of the year, the extent to which STUDY_i varies across individuals depends to a large extent on the other terms in equation (4). In particular, because the C 's are constant for all people and because the variance of $\frac{1}{T} \sum_{t=7}^T v_{i,t}$ is small, the variation that exists in STUDY_i will to a large extent depend on the importance of the observable permanent differences in study-time βX_i and the unobservable permanent differences in study-time μ_i .

To the normality assumptions for μ_i and $v_{i,t}$ described earlier, we add the additional assumption that $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ and estimate equation (1) and equation (3) jointly. Given that the outcome O_i is the continuous variable cumulative GPA, the likelihood contribution for person i is

$$(5) \quad L_i = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\sigma_\epsilon^2 + \alpha_s^2 \sigma_{\bar{v}}^2}} \phi\left(\frac{O_i - \alpha_s \text{STUDY}_i^*(\mu_i, C, \beta) - \alpha_x X_i}{\sqrt{\sigma_\epsilon^2 + \alpha_s^2 \sigma_{\bar{v}}^2}}\right) \prod_{t=1}^{N_i} \frac{1}{\sigma_v} \phi\left(\frac{s_{i,t} - C_t - \beta X_i - \mu_i}{\sigma_v}\right) f(\mu_i) d\mu_i$$

where $\bar{v} = \frac{1}{T} \sum_{t=7}^T v_{i,t}$, $\sigma_{\bar{v}}^2 = \text{var}(\bar{v})$, and $\text{STUDY}_i^* = \text{STUDY}_i - \bar{v}$, and ϕ is the standard normal probability density function.

The likelihood contribution in equation (5) can be simulated as

$$(6) \quad L_i = \frac{1}{D} \sum_{d=1}^D \frac{1}{\sqrt{\sigma_\epsilon^2 + \alpha_s^2 \sigma_{\bar{v}}^2}} \phi\left(\frac{O_i - \alpha_s \text{STUDY}_i^*(\mu_i^d, C, \beta) - \alpha_x X_i}{\sqrt{\sigma_\epsilon^2 + \alpha_s^2 \sigma_{\bar{v}}^2}}\right) \prod_{t=1}^{N_i} \frac{1}{\sigma_v} \phi\left(\frac{s_{i,t} - C_t - \beta X_i - \mu_i^d}{\sigma_v}\right)$$

where μ_i^d is the d th of D draws from the distribution of μ_i .

Equation (6) reveals the importance of estimating equations (1) and (3) jointly. The first term in equation (6) represents the likelihood/probability that person i would have outcome O_i if his study-time is $\text{STUDY}_i(\mu_i^d, \cdot)$. The distribution, $N(0, \sigma_\mu^2)$, from which μ_i^d will be drawn will be the same for all i . However, it is important that our estimation approach take into account that a value of μ_i^d that may seem quite likely given the values of $s_{i,1}, \dots, s_{i,N_i}$ reported by one person may seem quite unlikely for a person with a different set of observed study times. When equations (1) and (3) are estimated jointly as in equation (6), the second (product) term in equation (6) simultaneously serves the purpose of allowing the estimation of the parameters in the study equation (3) and, for each person, serving as appropriate weights for the different values of μ_i^d that are needed for the calculation of the outcome portion of equation (6).

IV.B. Results

The difficulties of collecting time-use data are well known from work such as Juster and Stafford (1985) and Juster and Stafford (1991). In particular, retrospective questions that require individuals to recall time-use information over time intervals of substantial length have been found to be particularly problematic. As a result, there is an intuitive appeal to using questions that cover narrowly defined time-periods. This motivates us to primarily utilize the 24-hour reports that we have collected rather than the 7-day reports, but as we discuss below there are also reasons to examine the 7-day reports. We have found that students at Berea tend to have fairly fixed daily schedules with structure added by both regularly scheduled classes and also the labor program at Berea in which students work at least ten hours a week. Given this added structure, it is our hope that students answer questions of the sort “how much time did you spend studying in the last 24 hours” accurately.⁷ However, below we provide some evidence that answers

⁷Evidence in Stafford and Juster (1985) suggests that time-use reports of activities such as work that are regularly scheduled tend to be more accurate than reports of activities that take place less regularly.

to these questions contain at least some degree of reporting error and we compare estimates using the twenty-four hour reports to estimates using the weekly reports and also to estimates from a procedure that explicitly takes into account the reporting error that may be present in the twenty-four hour reports.

Column one and two of Table 2 show the estimation results obtained using the 24-hour reports where the two columns differ only in the manner in which $STUDY_i$ enters the outcome equation. The first group of numbers in each column are estimates associated with the study equation, (3). Observable characteristics do not explain much of the variation in study-time. Males, on average, study approximately .5 hours less per day than females, but all other included observable characteristics are statistically insignificant at conventional levels. However, this does not imply that all variation in study-time is transitory in nature. While the standard deviation of the transitory portion of study-time, σ_v , is substantial, the permanent portion of the unobserved portion of study-time is also important with a standard deviation of $\sigma_\mu=1.33$ and an associated t-statistic of over 16.0 in both columns one and two. In practice, this implies that variation in $STUDY_i$ will remain between individuals after using equation (3) to account for study-time in all periods where study-time is not observed.

The results suggest that some people have very different study habits than others, although the majority of these differences are not explained by the observable characteristics that have been included here. Given the importance that we will find for study-time in the determination of outcomes, an important question is whether these study-habits are determined before or after the person arrives at college. Although the results are not shown in Table 2, we found a strong correlation between students' college study habits and both the amount of time that they reported studying in high school in high school and the amount they expected to study in college as reported on our first survey that took place before the start of classes.⁸ Thus, it appears that some of the permanent differences in study-habits originate before college.

⁸The estimated effect (t-statistic) of high school weekly study hours on daily (24-hour) study hours in equation (1) was .046 (4.777). The estimated effect (t-statistic) of expected college weekly

The second group of estimates in each column of Table 2 are estimates pertaining to the relationship between college grades and study time as specified in equation (1). In the first column, $STUDY_i$ enters in a simple linear fashion. The estimated effect of study-time is both quantitatively and statistically significant with a t-statistic of 5.737. The estimated effect implies that an additional hour of studying on each weekday is associated with a GPA that is higher by .182. Roughly speaking, the predictive nature of study-time is at least as important as that of college entrance exam scores that have been found in past work to be one of the best available predictors of academic performance. A two standard deviation change in ACT is associated with a .457 higher college GPA.⁹ A two standard deviation change in $STUDY_i$ is associated with a .589 higher college GPA. In the second column of Table 2, $STUDY_i$ enters in a quadratic fashion. The estimated effects of both the linear and quadratic terms are statistically significant at conventional levels and the estimates imply that the predicted marginal grade increase associated with an additional hour of studying per day decreases with the number of hours that are studied. The predicted increase in GPA associated with changing weekday study amounts from one to two hours, two to three hours, three to four hours, and four to five hours is shown in the first column of Table 3.

The relative importance of the permanent unobserved component, μ_i , and the transitory component, $v_{i,t}$, play a crucial role in determining the extent to which variation remains in $STUDY_i$ after “filling in” study-time for all days for which study-time is not observed. As a result, it is important to check the sensitivity of our results to assumptions about these unobservables.

With respect to the permanent component, we relax our normality assumption for μ_i and estimate a model in which μ_i is of the semi-parametric type proposed by Heckman and Singer (1984). Specifically,

study hours on daily (weekday) study hours in equation (1) was .041 (4.908). One reason for not including these variables in Table 2 was simply that they are missing for some students. Regardless, including them did not have a substantial impact on the estimated effect of $STUDY_i$ on outcomes O_i .

⁹For this calculation, the standard deviation of study-time was taken from row seven of Table 1.

we allow μ_i to be a discrete random variable with four possible values $\mu_i^1, \mu_i^2, \mu_i^3$, and μ_i^4 and we estimate both these values and the probabilities $P(\mu_i^j) j=1,2,3,4$ associated with each of these values.¹⁰ The results are shown in column 3 of table 2. Consistent with our previous results we find that permanent heterogeneity plays a very important role in determining study amounts. We find that the estimated effect of studying on outcomes, .168, is very similar to that in column 1 of Table 2, .182, where we assumed that μ_i was normal.

With respect to the transitory unobservable, it seems possible that $\text{var}(v_{i,t})$ might vary with the average amount of hours that a person studies. Given the inability of observable characteristics to predict average study-hours, we relax the homoskedasticity assumption for $v_{i,t}$ by specifying the variance of $v_{i,t}$ to be a function of μ_i :

$$(7) \quad \sigma_{i,v} = \rho\mu_i + \sigma_v$$

The results of this heteroskedastic model are shown in column 4 of Table 2. The large positive influence of ρ indicates that individuals with higher propensities to study also have a higher transitory variation in hours. As a result, in this specification, parameters in the study equation (1) are estimated with more precision than they are in the homoskedastic case. However, the estimated effect that STUDY_i has on the grade outcome remains virtually unchanged.

At several points during the year, we asked each respondent what he thought his GPA would be if he studied a variety of hours. Responses to these questions can be used to construct the marginal grade benefit that each person believes would arise from an additional hour of studying. Averaging across both people and survey dates, these marginal returns to an additional hour of studying are shown in column 2

¹⁰Given the presence of the constants C , it is necessary to normalize one of the μ_i^j 's. We set $\mu_i^1=0$.

The likelihood contributions are similar to equation (5). The integral over the density of the continuous μ_i is replaced by a weighted average over the four possible values of the discrete μ_i where the weights are the probabilities of the particular values.

of Table 3. Consistent with our estimated relationship, students believe that there is a diminishing marginal benefit to study-time. However, a comparison of columns 1 and 2 of Table 3 indicates that the marginal effects reported by students are somewhat larger than those found in our analysis. There are several reasons that this could be the case. First, the numbers reported by the students have a causal interpretation, whereas this is not necessarily the case with our estimated effects. As discussed earlier, despite the fact that our grade equation contains a measure of ability (college entrance ACT scores), we are not comfortable making a causal interpretation of our results because our estimation does not take into account the reality that study-hours are endogenously determined. Further, it is difficult to know the direction of the bias of our descriptive estimator if the population parameter of interest is the causal relationship between study-time and grades. On one hand, students who have higher unobserved academic ability or have higher returns to studying may choose to study more. This would tend to lead our results to overstate the true causal relationship. However, it is certainly possible that individuals who are worse academically in unobservable ways may decide that they need to study more to “stay afloat” than other students who are doing better academically due to higher unobserved academic ability. In this case, our results, while quantitatively large, would understate the true causal effect. This would be consistent with the larger causal effects described by students. The competing nature of the effects that ability may have on study-time is a potential explanation for the finding in Table 2 that study-time does not vary systematically with observed ability as measured by ACT scores. If a similar situation exists for unobserved ability, it would be reasonable to view our results as being causal in nature.

A second possible explanation relates to the measurement of our time-use information. Our estimation approach described above accounts explicitly for variation in study-hours across periods. However, to this point, our estimation has proceeded under the assumption that students’ 24-hour study reports are accurate. If reporting error exists and is classical in nature our estimator is likely to suffer from attenuation bias. A simple look at the data reveals evidence that there is at least some degree of reporting

error in the daily reports. Despite instructions that allowed students to report hours in partial-hour increments, more than 85% of all 24-hour study reports are integers. Given an average study amount of approximately three hours per day, rounding error could be very substantial in percentage terms. Our 7-day reports may not be as likely to suffer from this type of rounding error.¹¹ The results of estimation using the 7-day measure are shown in column 5 of Table 2. The effect of study-time is still statistically important with a level of significance that is roughly the same as in column 1 where the 24-hour measure is used. In addition, the estimated effect of $STUDY_i$ in column 5 is very close to that in column 1 after adjusting for the fact that the $STUDY$ variable in column 5 is a weekly amount and the $STUDY$ variable in column 1 is a daily amount.

Unfortunately, rounding error is not the only reason that measurement error may be present in our reports. Individuals may also make random mistakes because they do not think carefully enough about (or are not able to recall) exactly how they spent their time in the last 24 hours or in the last 7 days. While the weekly measure is likely to have less of a problem with rounding error, it may have more of a problem with these other types of reporting error. As a result, the fact that the results are not sensitive to whether we use the 7-day or 24-day reports cannot be taken to imply that measurement error is not problematic.

Further, there are many conceivable avenues through which reporting error could arise that are not classical in nature. That is, referring to $s^*_{i,t}$ as the true study time for person i at time t , $s_{i,t}$ as person i 's report at time t , and $m_{i,t}$ as the reporting error for i at t , there are several plausible reasons that the relationship between true study-time and reported study-time may deviate from a classical relationship in which

$$(8) \quad s^*_{i,t} = s_{i,t} + m_{i,t},$$

¹¹Of course, rounding error could exist if individuals round weekly hours to the nearest five or ten hour amount. However, we found much less evidence of rounding error than in the daily case; only about .25 and .40 of the 7-day reports were multiples of ten and five respectively.

with $\text{var}(m_{i,t})$ constant for all i , and $m_{i,t}$ uncorrelated with $s_{i,t}$, $X_{i,t}$, and the unobservable in the outcome equation, ϵ_i . One possibility is that the mean of the reporting error may be the same for all i but the variance may be larger for some types of individuals than other types. This might occur, for example, if individuals with certain X 's tend to be more conscientious or more able to recall their time-use than other individuals. Perhaps of more concern is the possibility that $m_{i,t}$ may be correlated with $s_{i,t}$, $X_{i,t}$, or ϵ_i . For example, with respect to the latter, the discussion in Bound (1991), which takes place in the context of self-reporting of health status, raises the possibility that individuals who are underperforming in school may rationalize their grade outcomes (conscientiously or unconscientiously) in terms of study-behavior. However, it is unclear in our context whether individuals who are underperforming academically would tend to over report study-time or under report study-time¹². Similarly, one can construct plausible reasons that $m_{i,t}$ may be correlated with $s_{i,t}$ and $X_{i,t}$ but it is difficult to be entirely confident about the likely signs of the correlations. What does seem clear is that, for reasons described in detail in Bound et. al (forthcoming), without further information it is very difficult to know what the likely direction of existing bias would be for our estimator of α_s , the relationship between studying and outcomes.

Fortunately, we do have an independent means for attempting to determine the extent to which measurement error may be influencing the results that we have presented to this point. Although we were aware that it is generally accepted that time-diaries are the most accurate way to collect time-use information, we decided on our approach (narrowly defined questions about the preceding 24-hour period) because we felt that students would answer these questions quite accurately given the existence of relatively structured schedules discussed above and because we felt that we would obtain lower response rates and

¹²On one hand, students might feel that it is not desirable to be viewed (or may not wish to view themselves) as performing poorly due to "laziness." This might lead to over reports of study-effort for those performing poorly. On the other hand, students might feel that it is not desirable to be viewed (or may not wish to view themselves) as a low ability person who performs poorly even when his effort is high. This might lead to under reports of study-effort for those performing poorly.

would be able to collect less observations per person during the year if we used the time-diary approach. However, at one point during the year, we conducted an experiment that was designed specifically to allow us to gauge the quality of the answers to our time-use questions. Approximately 200 students were randomly selected and came to a classroom on campus. Each student first completed a survey that included our standard time-use component and then completed either a written time-diary or an oral time-diary. Our intentions were primarily to compare study-time answers across our standard survey and the time-diary. However, we designed the two surveys in a manner such that it was very unlikely that individuals would know this was our intention. As a result, we feel comfortable that this experiment was a good test of our standard survey question.

The correlation between students' diary amounts and the standard survey reports was .72. Thus, the survey amount is a strong predictor of the diary amount, but only explains slightly more than half of the variation in the diary amount. As mentioned above it seems to be commonly accepted that time-diaries are the most accurate way to collect time-use information. In addition, in some limited debriefing sessions we noticed that many of the discrepancies between the time-diary and the standard survey arose because some students had not thought carefully about their previous 24-hour period when filling out the standard survey, and, as a result, had overlooked or overstated some of their study hours.¹³ Thus, although we stress that the time-diaries themselves undoubtedly contain some reporting error, we succumb to the convenience of thinking of the time-diary responses as students' true time-use for this 24-hour period.

Under this assumption, the first column of Table 4 shows a regression of the reporting error, m_i , (defined as the person's diary report minus the person's standard survey report) on X_i and the standard survey report, $s_{i,t}$. The estimated standard deviation of the unobservable in this regression is 1.523 which

¹³A few individuals also told us that their time-diary number conflicted with their standard survey number because they had adjusted the standard survey report because the last twenty-four hour was not typical for them. In some cases, upon debriefing, individuals also realized that they had incorrectly reported their activities during the oral time diary.

indicates that a non-trivial amount of reporting error exists. Consistent with the findings of previous work in other contexts that is detailed in Juster and Stafford (1985), the average person tends to over report hours; the unconditional mean of m_i is $-.417$ and a test of the null hypothesis that the mean reporting error is zero is rejected with a t-statistic of 3.636 . In addition, Table 4 indicates that over-reporting increases with the number of hours that are reported $s_{i,t}$ in a statistically significant and quantitatively important manner. However, with the exception of some evidence that blacks tend to over report, no evidence is found that reporting error varies systematically with the observable characteristics X_i .

Determining whether reporting error is correlated with the unobservable in the outcome equation ϵ_i is more difficult because the true residuals in the model cannot be computed without knowledge of the reporting error. Nonetheless, given the amount of variation in grades that remains unexplained by observable characteristics, it seems that a reasonable (if imperfect) view of the importance of this issue can be obtained by examining the relationship between reporting error and grades (rather than only the unobservable portion of grades, ϵ_i), especially if we control for the possibility that observable characteristics and $s_{i,t}$ might be correlated with both reporting error and grades. A regression of this form is shown in column 2 of Table 4. No evidence is found that reporting error is related to grades.

Thus, our results indicate that the average person over reports their study-time by approximately $\frac{1}{2}$ hour a day and that over reporting increases substantially with the amount of hours that a person reports but, by and large, is not related in any systematic manner to either grades or observable characteristics. We use these findings to motivate a simplified reporting error equation of the form

$$(9) \quad m_{i,t} = \delta_1 + \delta_2 s_{i,t} + \kappa_{i,t} \text{ where } \kappa_{i,t} \sim N(0, \sigma_\kappa^2) \text{ and is uncorrelated with } X_i \text{ and } \epsilon_i.$$

Estimates of the unknown parameters of equation (9) are shown in column 3 of Table 4. Under the assumption that this estimated relationship reflects the true reporting error process for all days that we collected time-use information, it is possible to adjust equation (5) to reflect the presence of reporting error in our 24-hour reports. The likelihood contribution in this case can be written as

(10)

$$L_i = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{\sigma_\epsilon^2 + \alpha_s^2 \sigma_v^2}} \phi\left(\frac{O_i - \alpha_s \text{STUDY}_i(\mu_i, C, \beta) - \alpha_x X_i}{\sqrt{\sigma_\epsilon^2 + \alpha_s^2 \sigma_v^2}}\right) \prod_{t=1}^{N_i} \frac{1}{\sigma_v} \phi\left(\frac{s_{i,t} + m_{i,t} - C_t - \beta X_i - \mu_i}{\sigma_v}\right) f(\mu_i) d\mu_i g(m_{i,t} | S_{i,t}) dm_{i,t}$$

where g , the normal density of the reporting error in day t , is as described in equation (9). The simulator in equation (6) can be modified for the changes present in equation (10). In particular, each of the d draws in equation (6) now entails drawing both a value of μ_i and values of $m_{i,1}, m_{i,2}, \dots, m_{i,N_i}$.

The results of this estimation are shown in Table (5).¹⁴ As compared to Table 2, the parameters are now estimated with less precision because the new estimator recognizes a new source of data uncertainty. The estimated effect of STUDY_i increases by approximately fifty percent in the linear case and even larger increases in the point estimates are seen in the quadratic specification. We can reject the null hypotheses that the effects of studying are zero at a 5% level of significance for both the linear and quadratic specification. Further, column 3 of Table 3 shows that the quadratic case has predicted marginal effects of studying that are very close in size to the causal beliefs expressed by students.

V. Implications for survey administrators and users of time-use data

The previous section suggests that understanding issues related to time-use may be very important for researchers who are interested in understanding a wide array of educational issues. As a result, surveys that focus on individuals of schooling age may want to seriously consider providing some type of time-use information. One of the methodological goals of this paper is to provide some information about the potential benefits of providing researchers with different types or different amounts of time-use data. Hopefully, this will assist survey administrators, who are likely to face difficult cost and logistical issues

¹⁴We note that, to some extent, our standard errors will be understated because we have not taken into account existing uncertainty about the parameters in equation (9). While this is not ideal, it does not seem especially problematic given that our intention is simply to provide some rough information about the nature of the biases that might be present in our initial estimator that assumes that no reporting error is present.

related to the collection of time-use information. A second methodological goal of this paper is to provide some information to potential users about the benefits of incorporating the type of estimators described above that explicitly account for sampling variation in each student's time-use. Understanding the extent to which these estimators provide different inference/insight is important because implementation of these estimators is likely to be more costly than simply assuming that a person's observed time-use behavior in some subset of the days/weeks of a year (semester) perfectly depicts his time-use over the course of an entire academic year (semester). Note that the two methodological issues are likely to be interrelated. For example, the extent to which it is problematic to assume that a person's observed time-use in some subset of the year perfectly depicts his time-use over the course of the entire year is likely to depend on how many time-use observations are available.

As noted earlier, not all of the individuals in our sample responded to all of our time-use surveys. For this section, given our desire to compare results based on different number of survey responses, we concentrate exclusively on the 139 individuals who answered all six of our time use surveys. We also ignore the possible presence of measurement error in the reports $s_{i,1}, \dots, s_{i,6}$ that was discussed in the previous section. To facilitate comparisons, we focus on the model in which study-time enters in a linear fashion and concentrate almost exclusively on α_s , the relationship between $STUDY_i$ and grade outcomes.

Estimates that are analogous to column 1 of Table 2 for our 139 person subsample are shown in Table 6. Comparing Table 6 to the first column of Table 2 shows that the estimated effect of α_s is, .182 for the full sample and is .153 for the subsample.

V. A. Analysis using Time-Use Information from a Single Contact with Respondent

We begin by considering the situation in which it is only possible for a survey to have contact with its respondents once a year. In this case, the data choices are limited to either the collection of a single time-use report (e.g., a time diary) detailing student's activities in some short period (perhaps a 24-hour period) or a retrospective question about time-use during the previous academic year (semester). In either

case, the user has few options other than to assume that the information provided by the survey is a good proxy for the yearly (or semester) time-use measure that is desired.

Consider first the case where the survey provides a researcher with a single time-use report for a single 24-hour period. Referring to this time period as k , the stochastic process of studying described in equation (3), implies that person i 's study time at k would be generated by $s_{i,k} = C_k + \beta X_i + \mu_i + v_{i,k}$ whereas the true average yearly study-time would be given by $\frac{1}{120} \sum_{t=1}^{120} s_{i,t} = \frac{20}{120} \sum_{t=1}^6 C_t + \beta X_i + \mu_i + \frac{1}{120} \sum_{t=1}^{120} v_{i,t}$. The difference between these two amounts represents the measurement error that is introduced by using the single measure $s_{i,k}$ as if it was the true average yearly measure. Assuming for the sake of simplicity that $C_1=C_2=C_3=\dots=C_6$, the measurement error is given by $v_{i,k} - \frac{1}{120} \sum_{t=1}^{120} v_{i,t}$. The sum in this expression is a random variable with a mean of zero and a variance of $\sigma^2/120$. Thus, the importance of measurement error and the degree to which attenuation bias for the estimator of α_s may be problematic depends on the importance of the transitory error $v_{i,k}$. From Table (6) we see that, as observed earlier, σ_v is substantial. As a result, it is not surprising that the estimated effects of $STUDY_i$ when a single 24-hour report is used as the true amount are much smaller than that shown in Table 6. In particular, for each $t=1,2,3,4,5,6$ we estimated equation (1) under the assumption that $STUDY_i=s_{i,t}$ and found estimates (std. errors) of .043 (.020), .054 (.023), .038 (.019), .070 (.020), .052 (.019), and .052 (.018) for $t=1,2,3,4,5,6$ respectively.

The standard errors reveal that the estimates of precision in case where a single 24-hour report is used as the true amount understate the amount of sampling variation present in the estimator of the relationship between yearly study-time and outcomes as seen in Table 6. This is the case because, unlike the estimator used for Table 6, the simple single-report estimator assumes that $STUDY_i$ is known with certainty. Using simulation methods we can use a particular person's six time-use observations, the assumed study-time model in equation (3), and the estimates in Table 6 to approximate the mean and variance of the distribution of $STUDY_i$ for each person i . We found that, on average, the $STUDY_i$ distribution for a person has a standard deviation of .566. Ignoring this uncertainty leads to the overstatement of precision.

At the end of the academic year, we asked individuals a retrospective question of the form “how much did you study on a typical day in the last academic year.” Using the answers to this question in place of $STUDY_i$ in equation (1), we found an estimated effect (std. error) of α_s to be .086 (.027). The fact that the estimated effect using the retrospective data is smaller than that in Table 6 is consistent with previous research that suggests that a substantial amount of measurement error is likely to exist in retrospective questions of this form.

In short, it seems that researchers will have difficulty understanding the impacts and importance of time-use if surveys provide time-use information based on a single contact with respondents. Further, retrospective data on time-use may tend to be worse than what we found in our data given the reality that most interviews will not take place at exactly the end of the academic year when a person’s recollections of his study habits during the most recent academic year are presumably most accurate.

V. B. Analysis using multiple time-use responses

A large marginal benefit to researchers may arise from increasing the number of time-use observations from one to two. The presence of multiple time-use observations provides researchers with an opportunity to identify the relative importance of the permanent and transitory unobservable components in the determination of study-time. As a result, our estimators, which explicitly take into account sampling variation in time-use, become feasible with two observations. However, the usefulness of two time-use observations will depend on the extent to which this is sufficient to allow a reasonably precise determination of each person’s permanent heterogeneity type μ_i . This is the case because greater uncertainty about μ_i translates directly into greater uncertainty about $STUDY_i$, and, ultimately, into less precise estimators of α_s . Earlier, we found that, on average the standard deviation of $STUDY_i$ across individuals is .566 when all six time-use reports are used in estimation (i.e., when $N_i=6$). On average, the standard deviation of $STUDY_i$ across individuals is .835 when only two time-use reports are used (i.e., when $N_i=2$). To see the extent to which this additional uncertainty in $STUDY_i$ influences estimates for α_s

we first estimated our model for our 139 person subsample using only the two observations $s_{i,1}$ and $s_{i,2}$. We then repeated this estimation using $s_{i,j}$ and $s_{i,k}$ for each of the other fourteen combinations of j and k that can be formed from our six time-use reports. The fifteen estimated values of α_s are shown in column 1 of Table 7. The average of the fifteen estimated standard errors is .062 which is approximately sixty-three percent larger than the standard error when all six time-use observations were used.

A benefit of our approach is that it explicitly accounts for sampling variation in time-use, and, as a result, allows valid inference. However, implementation is more costly than simply assuming that $STUDY_i$ is equal to the average of the observed study times, in which case equation (1) could be estimated by Ordinary Least Squares (OLS). To provide a sense of the benefits of implementing our estimator, column 2 of Table 7 shows estimation of equation (1) under the assumption that $STUDY_i = (s_{i,j} + s_{i,k})/2$ for each of the fifteen j,k combinations.¹⁵ The amount of measurement error is now approximately $\frac{v_{i,j} + v_{i,k}}{2} - \frac{1}{120} \sum_{t=1}^{120} v_{i,t}$. As before, because the sum has mean zero and a small variance, the amount of measurement that is present depends on the size of $\frac{v_{i,j} + v_{i,k}}{2}$ which has a mean of zero but a non-trivial variance of $\sigma^2/2$. Thus, it is not surprising that ignoring the sampling variation in the reports produces estimates in column 2 that are smaller than their counterparts in column 1 where the amount of measurement error is explicitly account for during estimation. In addition, as in the $N_i=1$ case, inference is difficult because this approach assumes that $STUDY_i$ is known with certainty.

The $N_i=2$ numbers in columns one and two of Table 7 are recreated in Table 8 for $N_i=4$. In column 1, the estimated standard errors decline substantially from the $N_i=2$ case to an average of .043 which is only thirteen percent larger than the estimated standard error associated with the $N_i=6$ case. This arises because the uncertainty in the $STUDY_i$ measures declines from the $N_i=2$ case; our approximations indicate that the average standard deviation of the $STUDY_i$ distribution declines from .835 to .686 when one moves from

¹⁵Although these estimates could have been estimated by OLS, we estimated them by maximum likelihood to maintain consistency with other results.

$N_i=2$ to $N_i=4$. Column 2 shows the estimated effects when, analogous to column 2 of Table 7, the model is estimated assuming $STUDY_i=(s_{i,j}+s_{i,k}+s_{i,l}+s_{i,m})/4$. The estimated variance of the measurement error is now approximately $\sigma^2/4$. As a result, the estimates in Column 2 are larger than those in Column 2 of Table 7, but clearly problems with the size of the standard errors remain. This problem also remains when the model is estimated with $STUDY_i=(s_{i,1}+s_{i,2}+s_{i,3}+s_{i,4}+s_{i,6}+s_{i,6})/6$. In this case, the estimated effect (std. error) of α_s is .111 (.028).

VI. Conclusion

By taking advantage of unique new longitudinal survey data, this paper provides perhaps the first examination of the importance of effort in the production of education. We find a quantitatively and statistically large relationship between study-time and first-year college grades. Although we view our estimator as descriptive in nature, there are reasons to believe that our estimated relationship may be an indicator of a strong causal effect of effort on performance. Among these reasons, the predictions of our models are consistent with both individuals' reported beliefs regarding the causal relationship between studying and outcomes and earlier work in Stinebrickner and Stinebrickner (2001) that shows that employment during school has a large, negative, causal impact on academic performance.

Our results suggest the value of future work that examines both the manner in which students make time-use decisions and the consequences of these decisions. However, for this to occur, researchers must be provided with time-use information. The paper suggests that it may be difficult for survey administrators to provide useful time-use data if they are constrained to a single contact each year with respondents. However, the results obtained here with estimators that explicitly take into account sampling variation in time-use suggest that large gains in the usefulness of data can be achieved with a relatively small number of time-use observations. With respect to the type of questions that survey administrators should use, evidence of an survey experiment that we conducted suggests that non-trivial measurement error may exist even in narrowly defined questions of the form "How much did you study in the last twenty-four hours."

As a result of these findings, we are utilizing time-diaries for the current (second) year of our ongoing survey.

References

- Betts (1997): The Role of Homework in Improving School Quality, unpublished manuscript.
- Ehrenberg, Ronald, and Sherman, Daniel. "Employment While in College, Academic Achievement and Postcollege Outcomes." *Journal of Human Resources* 22(1) (Winter 1987m): 1-24.
- Bound, John. "Self-Reported Objective Measures of Health in Retirement Models," *The Journal of Human Resources*, Volume 26 (1), (1991): 106-138.
- Bound, John, Brown, Charles, and Mathiowetz, Nancy, "Measurement Error in Survey Data," forthcoming, *Journal of Econometrics*.
- Juster, Thomas, and Stafford, Frank. "Time, Goods, and Well-Being." Survey Research Center, Institute for Social Research, The University of Michigan, 1985.
- Juster, F. Thomas and Frank Stafford. "The Allocation of Time: Empirical Findings, Behavioral Models, and Problems of Measurement." *Journal of Economic Literature* 26 (1991): 471-522.
- Heckman, J. and Singer, B. "The Identifiability of the Competing Risks Model" *Biometrika*, vol. 76, (1984): 325-330.

Ruhm, Chris. "Is High School Employment Consumption or Investment." *Journal of Labor Economics* 15(4) (October 1997): 735-776.

David Stern, Yoshi-Fumi Nakata. "Paid Employment among U.S. College Students: Trends, Effects, and Possible Causes." *Journal of Higher Education*, Vol. 62, No. 1. (Jan. - Feb., 1991), pp. 25-43.

Stinebrickner, Ralph and Stinebrickner, Todd. "Understanding Outcomes of Students from Low Income Families: Evidence from a School with a Full-Tuition Subsidy," unpublished manuscript (1998).

Stinebrickner, Ralph and Stinebrickner, Todd. "Working During School and Academic Performance." Forthcoming, *Journal of Labor Economics* (2001).

White, Halbert. 1982. "Maximum Likelihood Estimation of Misspecified Models." *Econometrica* 50: 1-25.

Table 1: Descriptive Statistics

REPORTED STUDY TIME		
	mean (standard deviation)	mean (standard deviation)
	study hours in previous day	study hours in previous 7 days
survey 1	3.41 (2.11)	18.37 (11.28)
survey 2	3.50 (2.16)	18.75 (11.37)
survey 3	3.52 (2.18)	16.11 (10.24)
survey 4	3.48 (2.16)	19.66 (10.91)
survey 5	3.39 (2.22)	19.85 (12.09)
survey 6	3.48 (2.20)	19.49 (12.51)
average study across surveys	3.42 (1.62)	18.20 (8.75)
OBSERVABLE CHAR'S X_i	mean (standard deviation)	
Male	0.4	
Black	0.14	
Family Income	27,000 (16,379)	
Combined ACT	23.58 (3.52)	
Parent With College Degree	0.47	
Hours Week Study High School	13.23 (10.21)	
Hours Week Expected Study	24.72 (13.00)	
OUTCOMES O_i		
Grade Point Average	2.93 (.686)	
Leave School Before 2 nd Year	0.18	

Table 2: Estimates of full model

	I.	II.	III.	IV.	V.
STUDY	24-hour reports	24-hour reports quadratic	Non-Parametric μ_1	Hetero-skedasticity.	7-day reports
C_1	4.438 (.751)*	4.554 (.750)*	2.562 (1.175)*	4.512 (.552)*	23.254 (3.978)*
C_2	4.505 (.753)*	4.635 (.750)*	2.634 (1.175)*	4.538 (.552)*	23.414 (3.985)*
C_3	4.479 (.753)*	4.589 (.750)*	2.617 (1.176)*	4.559 (.558)*	20.459 (3.982)*
C_4	4.451 (.753)*	4.568 (.750)*	2.573 (1.176)*	4.517 (.547)*	24.202 (3.986)*
C_5	4.390 (.753)*	4.494 (.750)*	2.525 (1.175)*	4.333 (.555)*	24.423 (3.985)*
C_6	4.523 (.753)*	4.627 (.750)*	2.667 (1.174)*	4.426 (.551)*	23.997 (3.985)*
Male	-.398 (.193)*	-.472 (.194)*	-.530 (.179)*	-.310 (.148)*	-1.964 (1.032)
Black	.347 (.293)	.341 (.293)	.047 (.257)	-.412 (.223)	1.798 (1.616)
ACT	-.034 (.029)	-.038 (.029)	-.030 (.029)	-.032 (.021)	-.172 (.155)
Fam. Income	-.015 (.059)	-.018 (.059)	-.026 (.053)	-.020 (.039)	.052 (.311)
Parental Educ.	-.103 (.185)	-.068 (.186)	-.078 (.179)	-.093 (.135)	-.540 (1.004)
σ_μ -permanent	1.329 (.079)*	1.329 (.081)*		1.254 (.067)*	7.80 (.394)*
μ_2			5.462 (.776)*		
μ_3			2.787 (.709)*		
μ_4			1.160 (.570)*		
$P(\mu_2)$.065		
$P(\mu_3)$.335		
$P(\mu_4)$.512		
σ_v -transitory	1.740 (.036)*	1.753 (.037)*	.564 (.036)*	1.659 (.045)*	8.102 (.170)*
ρ				.393 (.030)*	
GPA					
Constant	.835 (.309)*	.158 (.380)	.903 (.308)*	.896 (.306)*	1.06 (.299)*
Male	-.305 (.073)*	-.244 (.075)*	-.316 (.074)*	-.312 (.073)*	-.329 (.073)*
Black	-.194 (.110)*	-.177 (.105)	-.191 (.111)	-.173 (.110)	-.191 (.111)
ACT	.065 (.010)*	.068 (.010)*	.064 (.010)*	.065 (.010)*	.063 (.010)*

Fam. Income	.002 (.021)	.006 (.020)	.002 (.021)	.003 (.021)	-.002 (.021)
Parental Educ.	.141 (.069)*	.125 (.067)	.141 (.069)*	.138 (.069)*	.136 (.069)
STUDY PER WEEKDAY	.182 (.031)*	.517 (.126)*	.168 (.030)*	.169 (.030)*	
STUDY^2 WEEKDAY		-.041 (.015)*			
STUDY PER WEEK					.025 (.005)*
σ_ϵ	.556 (.025)*	.514 (.030)*	.564 (.024)*	.567 (.024)*	.573 (.024)*
log likelihood	-3247.46	-3245.26	-3235.87	-3167.10	-5496.47

Columns I shows estimates of full model using 24-hour reports with likelihood contributions as in equation (6). Column II is same as I but adds quadratic STUDY term. Column III specifies μ_i non-parametrically. Column IV allows heteroskedasticity for σ_v^2 . Column V uses 7-day reports. Standard errors are computed using robust methods of White (1982) and are in parentheses. * represents a t-statistic of greater than two.

Table 3 Marginal Effect of Study - Model Prediction and Individual Beliefs

Hours	Increase in Grades Model Prediction (no reporting error) Table 1 Column 2	Increase in Grades Individual Beliefs survey report	Increase in Grades Model Prediction (reporting error) Table 4 Column 2
1 Hour to 2 Hours	0.397	0.653	0.760
2 Hours to 3 Hours	0.317	0.497	0.535
3 Hours to 4 Hours	0.237	0.403	0.305
4 Hours to 5 Hours	0.157	0.222	0.075

Table 4
 Determinants of reporting error (time-diary - standard 24-hour report)

Constant	1.013 (.505)	1.363 (.726)	.445 (.316)
Report $s^*_{i,t}$	-.247 (.055)*	-.241 (.056)*	-.260 (.067)
Male	-.207 (.233)	-.246 (.240)	
Black	-.635 (.295)*	-.691 (.307)*	
ACT	-.017 (.018)	-.014 (.018)	
Fam. Income	.0004 (.008)	.001 (.008)	
Parental Educ.	-.042 (.069)	-.041 (.214)	
GPA		-.149 (.211)	
estimated std. $\kappa_{i,t}$	1.523	1.526	1.552 (.077)*
log likelihood			-304.90

Table shows regression of reporting error on 24-hour time-use report as a function of observable characteristics. Reporting error is assumed to be difference between detailed time-diary report and report on 24-hour time-use question.

Table 5: Estimates of full model: correction for reporting error in 24-hour reports

STUDY	I.	II.
C ₁	3.736 (.593)*	3.696 (.583)*
C ₂	3.771 (.590)*	3.773 (.572)*
C ₃	3.868 (.579)*	3.708 (.570)*
C ₄	3.719 (.579)*	3.807 (.572)*
C ₅	3.719 (.592)*	3.676 (.579)*
C ₆	3.833 (.599)*	3.771 (.586)*
Male	-.125 (.216)	-.214 (.241)
Black	.271 (.281)	.286 (.256)
ACT	-.026 (.022)	-.023 (.022)
Fam. Income	.013 (.047)	.004 (.016)
Parental Educ.	-.008 (.217)	.031 (.229)
σ_{μ} -permanent	.973 (.124)*	.923 (.141)*
σ_{ν} -transitory	.883 (.051)*	.909 (.052)*
GPA		
Constant	.569 (.346)	-.957 (.785)
Male	-.310 (.084)*	-.199 (.105)
Black	-.208 (.104)*	-.200 (.109)
ACT	.067 (.011)*	.071 (.010)*
Fam. Income	.001 (.017)	.007 (.019)
Parental Educ.	.139 (.072)	.106 (.071)
STUDY PER WEEKDAY	.277 (.055)*	1.115 (.394)*
STUDY ² WEEKDAY		-.115 (.053)*
σ_{ϵ}	.548 (.033)*	.466 (.045)*
log likelihood	-2995.992	-2989.48

Columns I shows estimates of full model using 24-hour reports with likelihood contributions corrected for possibility of reporting error as in equation (10).

Standard errors are computed using robust methods of White (1982) and are in parentheses.

* represents a t-statistic of greater than two.

Table 6 Estimates of full model: restricted subsample of students who answered all 6 time-use surveys.

STUDY	
C ₁	2.109 (1.009)*
C ₂	2.287 (1.009)*
C ₃	2.296 (1.009)*
C ₄	2.133 (1.009)*
C ₅	2.222 (1.009)*
C ₆	2.405 (1.009)*
Male	.168 (.260)
Black	.496 (.382)
ACT	.053 (.039)
Fam. Income	-.046 (.083)
Parental Educ.	.138 (.246)
σ_{μ} -permanent	1.289 (.103)*
σ_{ν} -transitory	1.712 (.046)*
GPA	
Constant	.750 (.345)*
Male	-.214 (.088)*
Black	-.318 (.128)*
ACT	.076 (.013)*
Fam. Income	.022 (.028)
Parental Educ.	.086 (.083)
STUDY PER WEEKDAY	.153 (.038)*
σ_{ϵ}	.464 (.029)*
log likelihood	-1807.56

Table shows same results as column 1 of Table 2 but uses restricted sample of students who answered all six of our time-use surveys.

Standard errors are computed using robust methods of White (1982) and are in parentheses.

* represents a t-statistic of greater than two.

Table 7: Estimates of α_s for model with $N_i=2$.

$s_{i,j} s_{i,k}$		
j=1, k=2	.188 (.079)*	.074 (.026)*
j=1, k=3	.113 (.047)*	.059 (.023)*
j=1, k=4	.132 (.041)*	.079 (.023)*
j=1, k=5	.184 (.073)*	.076 (.024)*
j=1, k=6	.149 (.054)*	.072 (.023)*
j=2, k=3	.149 (.062)*	.065 (.025)*
j=2, k=4	.150 (.044)*	.086 (.024)*
j=2, k=5	.230 (.093)*	.083 (.026)*
j=2, k=6	.243 (.096)*	.083 (.025)*
j=3, k=4	.172 (.057)*	.079 (.023)*
j=3, k=5	.219 (.095)*	.074 (.024)*
j=3, k=6	.143 (.053)*	.066 (.022)*
j=4, k=5	.122 (.035)*	.080 (.022)*
j=4, k=6	.162 (.047)*	.086 (.022)*
j=5, k=6	.157 (.053)*	.076 (.023)*

Each row in first column shows estimate (std. error) of α_s from the estimation of a specification and subsample identical to that in table 6 but restricting $N_i=2$ and using only the two 24-hour reports $s_{i,j}$ and $s_{i,k}$.

Each row in the second column shows an estimate of equation (1) where $STUDY_i$ is assumed to be $(s_{i,j}+s_{i,k})/2$. Standard errors are computed using robust methods of White (1982) and are in parentheses.

* represents a t-statistic of greater than two.

Table 8: Estimates of α_s for model with $N_i=4$

$s_{ij} s_{i,k}, s_{i,m} s_{i,n}$		
j=3, k=4, m=5, n=6	.153 (.039)*	.099 (.026)*
j=2, k=4, m=5, n=6	.160 (.040)*	.105 (.026)*
j=2, k=3, m=5, n=6	.177 (.049)*	.100 (.028)*
j=2, k=3, m=4, n=6	.160 (.042)*	.100 (.027)*
j=2, k=3, m=4, n=5	.159 (.042)*	.101 (.027)*
j=1, k=4, m=5, n=6	.145 (.038)*	.100 (.026)*
j=1, k=3, m=5, n=6	.153 (.044)*	.095 (.027)*
j=1, k=3, m=4, n=6	.141 (.038)*	.095 (.027)*
j=1, k=3, m=4, n=5	.146 (.040)*	.097 (.027)*
j=1, k=2, m=5, n=6	.178 (.049)*	.105 (.028)*
j=1, k=2, m=4, n=6	.158 (.041)*	.104 (.027)*
j=1, k=2, m=4, n=5	.152 (.041)*	.103 (.027)*
j=1, k=2, m=3, n=6	.151 (.044)*	.092 (.027)*
j=1, k=2, m=3, n=5	.167 (.050)*	.098 (.029)*
j=1, k=2, m=3, n=4	.142 (.040)*	.095 (.027)*

Each row in first column shows estimate (std. error) of α_s from the estimation of a specification and subsample identical to that in table 6 but restricting $N_i=4$ and using only the four 24-hour reports s_{ij} , $s_{i,k}$, $s_{i,n}$, and $s_{i,m}$. Each row in the second column shows an estimate of equation (1) where $STUDY_i$ is assumed to be $(s_{ij} + s_{i,k} + s_{i,n} + s_{i,m})/4$.

Standard errors are computed using robust methods of White (1982) and are in parentheses.

* represents a t-statistic of greater than two.