ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT

Journal of Econometrics ▮▮▮ (▮▮▮▮) ▮▮▮–▮▮▮

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Time-use and college outcomes

Ralph Stinebrickner[a], Todd R. Stinebrickner[b,*]

[a] *CPO 1971, Berea College, Berea, KY 40404, USA*

[b] *Department of Economics, Social Science Centre, The University of Western Ontario, London, Ont., Canada N6A 5C2*

Abstract

Despite an increased awareness of the policy importance of understanding the determinants of educational outcomes, knowledge of the relationship between educational 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 new data from the Berea Panel Study.

JEL classification: I2; J0; C5

Keywords: Time-use; Higher education; Grade performance; Measurement error; Study-time

1. Introduction

Despite an increased awareness of the policy importance of understanding the determinants of educational outcomes, knowledge of the relationship between educational 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

* Corresponding author.

E-mail addresses: ralph_stinebrickner@berea.edu (R. Stinebrickner), trstineb@uwo.ca (T.R. Stinebrickner).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2003.10.013

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 to is to collect information about a single time period (perhaps 24 h) 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 unique data that have been created by merging detailed information 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 24-h period and the 7-day period that immediately preceded the time of the survey. We use these questions to derive Maximum Likelihood Estimators (MLEs) of the relationship between time-use and student outcomes. The MLE 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 collected time-use information for a subset of the total number of time-use periods during the academic year 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. While the flexibility associated with the MLE is very desirable in this context, it comes at the cost of distributional assumptions. To provide some evidence that these distributional assumptions are not driving our results, we describe the relationship between our MLE and a standard distributional-free measurement error estimator and compare estimates from the two estimators in a case where it is feasible to compute both.

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 24-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 possible reporting error in the six primary time-use surveys.[1]

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 at least as important as 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 importance of studying found in this paper is consistent in spirit with results in Stinebrickner and Stinebrickner (2003a) 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 a 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. 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, after comparing our results with results obtained using only a single time-use survey or a single end-of-the-year retrospective question, we conclude that it is very important for survey administrators to collect time-use information at more than one time during the year. A small simulation study provides additional information about the benefit of having more than two time-use surveys during the year.

Section 2 provides a brief review of the literature on time-use and educational outcomes. Section 3 discusses Berea College and the longitudinal survey that we are conducting at the school. Section 4 describes the estimators and substantive results from

[1] Because we collect the detailed time-diary only once during the year, this requires that we make an assumption that the stochastic relationship between the retrospective time-use surveys and the "truth" (as measured by the time-diary) remains the same over the course of the year.

Berea. Section 5 examines the methodological issues discussed above and Section 6 concludes.

2. 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 (1987), 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 (2003a) analyzed 8 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]

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.

3. Data

Berea College is located in the foothills of the Appalachian mountains in Berea Kentucky. 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 (2003b) discovered that many students at Berea do not graduate despite the full tuition subsidy.

The time-use data described here come from the Berea Panel Study (BPS) which is being conducted by Todd Stinebrickner and Ralph Stinebrickner in an effort to obtain a better understanding of attrition and other outcomes in higher education. The BPS design involves collecting a wide array of detailed information about two cohorts of students from the time of college entrance through the time when early post-college

[2] The paper takes advantage of 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 important effect on the number of hours that a person works.

experiences take place.[3] As of June 2003, the 2000 cohort (which entered college in the fall semester of 2000) had been followed for 3 years and the 2001 cohort (which entered college in the fall semester of 2001) had been followed for 2 years.

This paper focuses on the 2000 cohort in its first year in college. Students were asked to complete nine surveys over the course of their first year. The initial 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.[4] For the purposes of this paper, we limit our attention to domestic, dependent students. There are a total of 306 freshmen 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.

Six of the remaining surveys were aimed primarily at eliciting information about how students were 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 0.90, 0.82, 0.83, 0.77, 0.75, and 0.75.[5] With respect to time spent studying, 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 24-h period (subsequently referred to as the "24-h" report), the most recent weekend day, and the previous 7 days (subsequently referred to as the "7-day" report). The timing of the delivery and deadlines of the surveys implies that all time-use surveys were completed on a weekday. Thus, the answer to the 24-h question represented hours of study in a 24-h 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.

The first six rows of Table 1(A) show descriptive statistics associated with the weekday and weekly study questions for each of the six primary time-use surveys. The seventh row of Table 1(A) shows descriptive statistics when each person's reported hours are averaged over all of the time-use surveys that the person completed. The sample distributions related to the 24-h reports from row 1 and row 7 of Table 1(A) are shown in the remainder of Table 1(A).

[3] In addition to collecting detailed background information about students and their families, the initial BPS survey that took place at the time of college entrance elicited information about students' 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 to those that remain at Berea, document the experiences of students during college and how the various factors that might influence decisions change over time. Shorter surveys, that are administered at multiple times each year to those that remain at Berea, provide information about how students use their time. Exit surveys, that are administered to those who have left college, document, among other things, students' work and education activities after leaving Berea.

[4] The initial participation rate for the 2001 cohort is approximately 0.90.

[5] Approximately 6% 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.

Table 1

(A) Descriptive statistics/hours

Reported study time

	Mean (standard deviation) Study hours in previous day	Mean (standard deviation) 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)

Sample distributions

Hour category	Survey 1 Study hours in previous day ($n=273$)	Average study across surveys ($n=300$)
[0, 1) Hour	0.055	0.03
[1, 2) Hour	0.095	0.123
[2, 3) Hour	0.227	0.27
[3, 4) Hour	0.208	0.256
[4, 5) Hour	0.186	0.173
[5, 6) Hour	0.098	0.07
[6, 7) Hour	0.073	0.036
[7+) Hour	0.054	0.04

(B) Correlation matrix—24-h study reports

	Survey 1	Survey 2	Survey 3	Survey 4	Survey 5	Survey 6
Survey 1		0.345	0.345	0.358	0.291	0.42
Survey 2			0.448	0.462	0.31	0.429
Survey 3				0.332	0.207	0.394
Survey 4					0.475	0.434

(C) Correlation matrix—7-day study reports

	Survey 1	Survey 2	Survey 3	Survey 4	Survey 5	Survey 6
Survey 1		0.488	0.388	0.457	0.333	0.464
Survey 2			0.583	0.549	0.49	0.504
Survey 3				0.441	0.441	0.393
Survey 4					0.532	0.576

(D) Other descriptive statistics

	Mean (standard deviation)
Observable char's X_i	
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 (0.686)
Leave school before 2nd year	0.18

The standard deviations of the entries in the first six rows indicate that there is a large amount of variation in study hours across individuals within a 24-h (7-day) period. The standard deviations of the entries in the seventh row indicate that permanent differences exist in study habits. However, the fact that the standard deviations in row 7 are smaller than those in rows 1–6 suggests that the stochastic process that determines 24-h or 7-day study-times also contains a non-trivial non-permanent component. This motivates our decision in Section 4 to model the stochastic process associated with 24-h and 7-day reports as the sum of a permanent and transitory component. It is convenient to assume that the transitory component is independent across time for a particular person. An alternative possibility is that serial correlation remains in the transitory component because, for example, some students study more and more as the academic year progresses. Table 1(B) and (C) show the correlation matrices of the six 24-h reports and the six 7-day reports respectively and provide some rough information about whether serial correlation exists in the transitory component. If the transitory component of study-times are serially correlated, then survey responses that are closer together should have higher correlations than survey responses that are further apart. Table 1(B) and (C) do not seem to reveal this pattern and a formal test of autocorrelation reaches the same conclusion.[6] As a result, it seems reasonable to model study-times in Section 4 using a permanent/transitory specification in which the transitory component is independent across time for a particular student.

Additional information about the students in our sample was obtained by merging our survey data with administrative data obtained from Berea College. Table 1(D) includes a descriptive look at the observable characteristics $X = \{$sex, race, family income, American College (ACT) scores, and parental education$\}$ that are used in the paper.[7] 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(D) show that the average first-year cumulative GPA is 2.93.

4. Estimators and 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

[6] The test involves differencing the data and is based on the fact that, under the null hypothesis of independence of $v_{i,t}$, $\mathrm{E}(v_{i,t+2} - v_{i,t+1})(v_{i,t+1} - v_{i,t}) = -\mathrm{E}(v_{i,t+1})(v_{i,t+1}) = -\mathrm{E}(v_{i,t+1} - v_{i,t})(v_{i,t+1} - v_{i,t})/2$. The test of the null hypothesis produces a t-statistic of 0.475 so the null hypothesis cannot be rejected at conventional levels.

[7] Most 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.

appeal to using questions that cover narrowly defined time-periods. This motivates us to primarily concentrate on the 24-h reports in the subsequent model description and analyses, but we also briefly examine the robustness of our results to the use of 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 10 h 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.[8] However, evidence from a survey design experiment we conducted suggests that these questions may contain a non-trivial amount of reporting error. As a result, we also discuss estimates from a procedure that explicitly takes into account the reporting error that may be present in the 24-h reports.

4.1. Description and comparison of MLE and measurement error estimator

We consider the outcome equation

$$O_i = \alpha_S \ \mathrm{STUDY}_i + \alpha_X X_i + \varepsilon_i. \tag{1}$$

The unobservable ε_i is assumed to have mean zero and variance σ^2_ε, O_i is the yearly freshman GPA outcome for person i, STUDY_i is the average amount that person i studies per day during the academic year, and X_i is a vector of observable characteristics that might influence grade performance. 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 Eq. (1) which is descriptive in nature.

Given this descriptive view of the problem, the primary obstacle in the estimation of Eq. (1) involves the measurement of STUDY_i. Letting T denote the total number of possible study days during the year and letting $s_{i,t}$ denote the number of hours that person i studied on day t,

$$\mathrm{STUDY}_i = \frac{1}{T}\sum_{t=1}^{T} s_{i,t}. \tag{2}$$

STUDY_i is not observed because $s_{i,t}$ is observed for only a subset of the days $t = 1, \ldots, T$. Instead, letting N_i denote the total number of days that study time is measured for person i, what is observed is a noisy measure of STUDY_i

$$\widehat{\mathrm{STUDY}}_i = \frac{1}{N_i}\sum_{t=1}^{N_i} s_{i,t}, \tag{3}$$

where we have reordered the T days so that the days $1, \ldots, N_i$ are the days for which study-time is observed for person i and days $N_{i+1}, \ldots, T$ are the days for which study-time is not observed for person i.

[8] Evidence in Juster and Stafford (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.

Using $\widehat{\text{STUDY}}_i$ as if it is the true STUDY_i when estimating Eq. (1) will typically imply that the OLS estimator of α_S will be biased. To characterize the nature and extent of this bias it is necessary to understand the manner in which $\widehat{\text{STUDY}}_i$ differs from STUDY_i. We assume a permanent/transitory process

$$s_{i,t} = C + \beta X_i + \mu_i + v_{i,t}, \quad t = 1, 2, \ldots, T. \tag{4}$$

The permanent component $C + \beta X_i + \mu_i$ represents the average amount that person i studies per day. This average is specified to vary with X_i to allow for the possibility that average study-hours may depend on observable characteristics such as family background or academic test scores. μ_i captures the reality that average study-time may depend on characteristics that are unobservable. The transitory component $v_{i,t}$ represents a daily deviation which has mean zero and variance σ_v^2. Substituting Eq. (4) into Eqs. (2) and (3), respectively, shows that

$$\text{STUDY}_i = C + \beta X_i + \mu_i + \frac{1}{T}\sum_{t=1}^{T} v_{i,t}$$

and

$$\widehat{\text{STUDY}}_i = C + \beta X_i + \mu_i + \frac{1}{N_i}\sum_{t=1}^{N_i} v_{i,t}. \tag{5}$$

Thus, the difference between the true and noisy measures of study time is given by

$$\text{STUDY}_i - \widehat{\text{STUDY}}_i = \frac{1}{T}\sum_{t=1}^{T} v_{i,t} - \frac{1}{N_i}\sum_{t=1}^{N_i} v_{i,t}. \tag{6}$$

Given the specification in Eq. (4), Eq. (6) characterizes the measurement error in $\widehat{\text{STUDY}}_i$. We note that in this application the total number of study days in a year, T, is large relative to the total number of observed days, N_i. With T large, the variance of $(1/T)\sum_{t=1}^{T} v_{i,t}$ becomes small, the measurement error shown in Eq. (6) depends on only the term $(1/N)\sum_{t=1}^{N_i} v_{i,t}$, and Eq. (1) can be rewritten

$$\begin{aligned} O_i &= \alpha_S\left(\widehat{\text{STUDY}}_i - \frac{1}{N_i}\sum_{t=1}^{N_i} v_{i,t}\right) + \alpha_X X_i + \varepsilon_i \\ &= \alpha_S\,\widehat{\text{STUDY}}_i + \alpha_X X_i + \left[\varepsilon_i - \alpha_S\,\frac{1}{N_i}\sum_{t=1}^{N_i} v_{i,t}\right]. \end{aligned} \tag{7}$$

The correlation between $\widehat{\text{STUDY}}_i$ and $[\varepsilon_i - \alpha_S\,(1/N_i)\sum_{t=1}^{N_i} v_{i,t}]$ implies that the OLS estimator $\hat{\alpha}_{S,\text{OLS}}$ from a regression of O_i on $\widehat{\text{STUDY}}_i$ will suffer from attenuation bias.[9] However, if we assume that N_i is constant for all i, an unbiased estimator can be constructed using standard textbook methods that depend on the measurement of the

[9] As is well-known, the bias in the OLS estimator remains even in large samples for $N_i < T$.

signal to noise ratio. Specifically, in the simplified case where $\alpha_X = 0$, the unbiased measurement error estimator of α_S is given by

$$\hat{\alpha}_{S,\text{measurement}} = \hat{\alpha}_{S,\text{OLS}} \frac{\text{Var}(\widehat{\text{STUDY}})}{\text{Var}(\widehat{\text{STUDY}}) - \text{Var}((1/N)\sum_{i=1}^{N} v_{i,t})}$$

$$= \hat{\alpha}_{S,\text{OLS}} \frac{\text{Var}(\widehat{\text{STUDY}})}{\text{Var}(\widehat{\text{STUDY}}) - \sigma_v^2/N}. \tag{8}$$

An estimate of σ_v^2 can be constructed without distributional assumptions by differencing the observations $s_{i1}, \ldots, s_{i,N}$.

This measurement error estimator is appealing because it requires no distributional assumptions. Unfortunately, it becomes difficult to implement an analogous approach when the model deviates from the simple framework described above or when data are missing for some people as is the case in our data where the majority of individuals answer only some subset of our six time-use surveys. This motivates our construction of a MLE which, at the cost of making distributional assumptions for ε_i, μ_i and $v_{i,t}$, is applicable in very general modeling situations and provides a natural way to accommodate missing data.

For ease of exposition, our description here focuses on the case where the number of periods in the year, T, is large relative to the number of observed periods N. The MLE estimator is analogous to the MLEs derived in the missing data literature.[10] The likelihood contribution for person i is given by

$$L_i = \int g(s_{i,1}, \ldots, s_{i,Ni}, O_i, \mu_i | X_i)\, \mathrm{d}\mu_i, \tag{9}$$

where g is the joint density of its arguments. Recall that, with T large, $\text{STUDY}_i = C + \beta X_i + \mu_i$ is determined by μ_i (and some guess of C and β). Thus, Eq. (9) is consistent in spirit with MLEs in the missing data literature which recognize that the appropriate likelihood contribution for a person is found by integrating the joint density of the dependent variables and any unobserved independent variables (given the set of observed independent variables) with respect to the unobserved variables (in this case STUDY_i).

Conditional on μ_i and X_i, the variables $s_{i,1}, \ldots, s_{i,Ni}$ and O_i are independent so Eq. (9) can be rewritten

$$L_i = \int g_1(s_{i,1}|\mu_i, X_i) \cdots g_1(s_{i,Ni}|\mu_i, X_i) g_2(O_i|\mu_i, X_i) h(\mu_i)\, \mathrm{d}\mu_i. \tag{10}$$

Eq. (4) indicates that the density g_1 has mean $C + \beta X_i + \mu_i$, variance σ_v^2, and a shape that is determined by distributional assumptions about $v_{i,t}$. Eq. (1) indicates that the density function g_2 has mean $\alpha_S\,\text{STUDY}_i + \alpha_X X_i = \alpha_S(C + \beta X_i + \mu_i) + \alpha_X X_i$, variance σ_ε^2, and a shape that is determined by distributional assumptions about ε_i. Our MLEs assume that both $v_{i,t}$ and ε_i are normally distributed.

[10] See for example, Little and Rubin (1987), Lavy et al. (1998), and Stinebrickner (1999).

h represents the density of μ_i in the population. In the majority of what follows, we assume that $\mu_i \sim N(0, \sigma_\mu^2)$. If the term $g_2(O_i|\mu_i, X_i)$ was removed from Eq. (10) what would be left would be a standard random effects estimator of the study Eq. (4). The intuition underlying our MLE estimator of α_S is that the product $g_1(s_{i,1}|\mu_i, X_i) \cdots g_1(s_{i,Ni}|\mu_i, X_i)$ provides information about how likely a particular value of μ_i is for person i given his/her observed study-times $s_{i,1}, \ldots, s_{i,Ni}$. In essence, Eq. (10) simply involves integrating $g_2(O_i|\mu_i, X_i)$ over the distribution of μ_i that is appropriate for person i in light of his observed study-times.[11]

Eq. (10) can be evaluated by simulation

$$L_i^s = \frac{1}{D} \sum_{d=1}^{D} g_1(s_{i,1}|\mu_i^d, X_i) \cdots g_1(s_{i,Ni}|\mu_i^d, X_i) g_2(O_i|\mu_i^d, X_i), \tag{11}$$

where μ_i^d is the dth of D draws from the distribution $h(\mu_i)$. Model estimates are found by maximizing the simulated likelihood function $L = \prod_i L_i^s$.

The MLE estimator provides a way to accommodate missing data and allows a large amount of modeling flexibility. As a result, it is very useful in this application. However, before turning exclusively to the MLE estimator it is desirable to provide some evidence that the distributional assumptions that are required for the MLE are not driving the results that we obtain later. To do this, we focus on the 139 individuals in our sample who answered all six time-use surveys. Because no data are missing for this group, $\hat{\alpha}_{S,\text{measurement}}$ can be computed using Eq. (8) and can be compared to $\hat{\alpha}_{S,\text{MLE}}$. For simplicity, we constrain $\alpha_X = 0$.[12] The OLS estimator $\hat{\alpha}_{S,\text{OLS}}$ produces an estimate (std. error) of 0.124 (0.036). We estimate $\text{Var}(\widehat{\text{STUDY}}) = 2.038$ and $\text{Var}(v_{i,t}) = 2.763$. Therefore, from Eq. (8), $\hat{\alpha}_{S,\text{measurement}} = 0.160$. The estimate (std. error) from the MLE $\hat{\alpha}_{S,\text{MLE}}$ is 0.171 (0.057). The measurement error estimator and MLE produce similar estimates for the parameter of primary interest α_S.

4.2. *Results using full-sample*

4.2.1. *Primary results*

We now turn to results involving our full sample. The MLE used in this section differs from the MLE in Section 4.1 in two minor ways that are detailed in Appendix A. First, we allow the constant in Eq. (4) to vary with each of the 6 days that time-use was measured to take into account that students may have more academic work on some days during the year than on other days.[13] Second, the estimator involves a slight

[11] This is perhaps most evident if we rewrite Eq. (10) as $\int g_2(O_i|\mu_i, X_i) h_1(\mu_i, s_{i,1}, \ldots, s_{i,Ni}|X_i)\, \mathrm{d}\mu_i$.

[12] This simplification is not likely to influence the conclusions from our comparison. As we show later, although the observable characteristics X_i are important in determining a student's GPA they are not important in determining how much a person studies.

[13] This is useful primarily because some students do not answer all time-use surveys. For a person who answers only a subset of time-use surveys it is useful to take into account the amount he studied relative to others on *those* days (or weeks). We assume throughout the paper that the decision of whether to participate in a particular time-use survey is exogenously determined. This seems somewhat reasonable given the very small amount of time that is required to complete a time-use survey.

modification that relaxes the assumption that T is large. In this application where T is large, we found that as expected this modification had virtually no effect on model estimates.[14]

Columns 1 and 2 of Table 2 show the estimation results obtained using the 24-h 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 Eq. (4). Observable characteristics do not explain much of the variation in study-time. Males, on average, study approximately 0.5 h 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 1 and 2.

These 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 and the amount they expected to study in college as reported on our first survey that took place before the start of classes.[15] Thus, it appears that some of the permanent differences in study-habits originate before college.

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 Eq. (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 0.182. 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 0.457 higher college GPA. A two standard deviation change in STUDY is associated with a 0.589 higher college GPA.[16] 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

[14] When we consider the 24-h reports, T represents the total number of study days in an academic year. When we consider the 7-day reports, T represents the total number of weeks in an academic year. In this case, T is not as large.

[15] The estimated effect (t-statistic) of high school *weekly* study hours on *daily* (24-h) study hours in Eq. (1) was 0.046 (4.777). The estimated effect (t-statistic) of expected college weekly study hours on daily (weekday) study hours in Eq. (1) was 0.041 (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.

[16] For this calculation, the standard deviation of study-time was taken from row 7 of Table 1.

Table 2
Estimates of full model

	I	II	III	IV	V
	24-h reports	24-h reports quadratic	Non-parametric μ_i	Heteroskedasticity	7-day reports
STUDY					
C_1	4.438 (0.751)*	4.554 (0.750)*	2.562 (1.175)*	4.512 (0.552)*	23.254 (3.978)*
C_2	4.505 (0.753)*	4.635 (0.750)*	2.634 (1.175)*	4.538 (0.552)*	23.414 (3.985)*
C_3	4.479 (0.753)*	4.589 (0.750)*	2.617 (1.176)*	4.559 (0.558)*	20.459 (3.982)*
C_4	4.451 (0.753)*	4.568 (0.750)*	2.573 (1.176)*	4.517 (0.547)*	24.202 (3.986)*
C_5	4.390 (0.753)*	4.494 (0.750)*	2.525 (1.175)*	4.333 (0.555)*	24.423 (3.985)*
C_6	4.523 (0.753)*	4.627 (0.750)*	2.667 (1.174)*	4.426 (0.551)*	23.997 (3.985)*
Male	−0.398 (0.193)*	−0.472 (0.194)*	−0.530 (0.179)*	−0.310 (0.148)*	−1.964 (1.032)
Black	0.347 (0.293)	0.341 (0.293)	0.047 (0.257)	−0.412 (0.223)	1.798 (1.616)
ACT	−0.034 (0.029)	−0.038 (0.029)	−0.030 (0.029)	−0.032 (0.021)	−0.172 (0.155)
Fam. income	−0.015 (0.059)	−0.018 (0.059)	−0.026 (0.053)	−0.020 (0.039)	0.052 (0.311)
Parental educ.	−0.103 (0.185)	−0.068 (0.186)	−0.078 (0.179)	−0.093 (0.135)	−0.540 (1.004)
σ_μ-permanent	1.329 (0.079)*	1.329 (0.081)*		1.254 (0.067)*	7.80 (0.394)*
μ_2			5.462 (0.776)*		
μ_3			2.787 (0.709)*		
μ_4			1.160 (0.570)*		
P(μ_2)			0.065		
P(μ_3)			0.335		
P(μ_4)			0.512		
σ_v-transitory	1.740 (0.036)*	1.753 (0.037)*	0.564 (0.036)*	1.659 (0.045)*	8.102 (0.170)*
ρ				0.393 (0.030)*	
GPA					
Constant	0.835 (0.309)*	0.158 (0.380)	0.903 (0.308)*	0.896 (0.306)*	1.06 (0.299)*
Male	−0.305 (0.073)*	−0.244 (0.075)*	−0.316 (0.074)*	−0.312 (0.073)*	−0.329 (0.073)*
Black	−0.194 (0.110)*	−0.177 (0.105)	−0.191 (0.111)	−0.173 (0.110)	−0.191 (0.111)
ACT	0.065 (0.010)*	0.068 (0.010)*	0.064 (0.010)*	0.065 (0.010)*	0.063 (0.010)*
Fam. income	0.002 (0.021)	0.006 (0.020)	0.002 (0.021)	0.003 (0.021)	−0.002 (0.021)
Parental educ.	0.141 (0.069)*	0.125 (0.067)	0.141 (0.069)*	0.138 (0.069)*	0.136 (0.069)
STUDY PER WEEKDAY	0.182 (0.031)*	0.517 (0.126)*	0.168 (0.030)*	0.169 (0.030)*	
STUDY*STUDY WEEKDAY		−0.041 (0.015)*			
STUDY PER WEEK					0.025 (0.005)*
σ_ε	0.556 (0.025)*	0.514 (0.030)*	0.564 (0.024)*	0.567 (0.024)*	0.573 (0.024)*
Log likelihood	−3247.46	−3245.26	−3235.87	−3167.1	−5496.47

Columns I shows estimates of full model using 24-h reports with likelihood contributions as in Eq. (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 2 Column 2	Increase in grades model prediction (reporting error) Table 5 Column 2	Increase in grades individual beliefs Survey report
1–2 h	0.397	0.523	0.653
2–3 h	0.317	0.375	0.497
3–4 h	0.237	0.229	0.403
4–5 h	0.157	0.083	0.222

Table shows increase in grades from an additional hour of study. For example, 0.397 is the predicted increase in grade point average from studying 2 h per day instead of 1 h per day.

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 1 to 2, 2 to 3, 3 to 4, and 4 to 5 h is shown in the first column of Table 3.

4.2.2. Robustness checks

The comparison of the measurement error estimator and MLE estimator in Section 4.1 provided some evidence that our MLE results are not being driven by distributional assumptions. However, given the important role that the unobserved permanent component μ_i plays in our estimator it seems worthwhile to specifically examine the sensitivity of our results to distributional assumptions involving μ_i. 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, we allow μ_i to be a discrete random variable with four possible values μ_i^1, μ_i^2, μ_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.[17] 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, 0.168, is very similar to that in column 1 of Table 2, 0.182, 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

[17] Given 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 Eq. (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.

assumption for $v_{i,t}$ by specifying the variance of $v_{i,t}$ to be a function of μ_i:

$$\sigma_{i,v} = \rho\mu_i + \sigma_v. \tag{12}$$

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 higher transitory variation in hours. As a result, in this specification, parameters in the study Eq. (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.

Finally, we examine the robustness of our results to a change in the length of the time-use reporting period by estimating the model using the 7-day reports rather than the 24-h reports. The results of estimation using the weekly hours measure are shown in column 5 of Table 2. The effect of study-time is statistically important with a level of significance that is roughly the same as in column 1 where the 24-h measure is used. In addition, the estimated effect of STUDY_i in column 5 is similar 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.

4.3. *Reporting error in time-use reports*

Our estimation approach described above accounts explicitly for true variation in study-hours across days (weeks), but has proceeded under the assumption that students' 24-h (7-day) reports are accurate. In reality, it is likely that the reports of $s_{i,t}$ suffer to some degree from reporting error. If the reporting error is classical in nature than our previous methods will continue to be appropriate. To see this, let s^*_{it} denote true study-time, and, as before, assume that true study-time is the sum of a permanent component and a transitory component which we now denote $u_{i,t}$

$$s^*_{i,t} = C + \beta X_i + \mu_i + u_{i,t}. \tag{13}$$

The assumption that reporting error is classical in nature implies that

$$s_{i,t} = s^*_{i,t} + e_{i,t} \tag{14}$$

where the reporting error $e_{i,t}$ is uncorrelated with all other random variables in the model. Substituting Eq. (13) into Eq. (14) produces

$$s_{i,t} = C + \beta X_i + \mu_i + v_{i,t}, \quad \text{where } v_{i,t} = e_{i,t} + u_{i,t}. \tag{15}$$

Eq. (15) indicates that the presence of classical reporting error implies that $v_{i,t}$ in Eq. (4) contains both the daily variation in true study time $u_{i,t}$ and the reporting error $e_{i,t}$. However, because the assumption that the reporting error is classical implies that $v_{i,t}$ retains the same properties as it did when it was assumed to contain only day-to-day variation in study-time, the previous estimators remain appropriate.

However, as described in detail in Bound et al. (2001), there is no guarantee that the reporting error is classical in nature, and if this is the case, it is difficult to know without further information what the likely direction of existing bias would be for our MLE estimator of α_S. Fortunately, we do have an independent means for attempting to characterize the nature of the reporting error. 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-h period) because we felt that students would answer these questions quite accurately given the existence of relatively structured schedules discussed earlier and because we felt that we would obtain lower response rates and would be able to collect less observations per person during the year if we used the time-diary approach. However, as part of our sixth time-use survey, we conducted a survey design experiment that specifically allowed us to gauge the quality of the answers to our time-use questions. Approximately 200 of our survey participants were randomly selected and came to a classroom on campus. Each student first completed a time-use survey that included our standard time-use component and then completed either a written time-diary or an oral time-diary. We designed the surveys in a manner such that it was very unlikely that individuals would know that our primary intention was to compare study-time measures across our standard survey and the time-diaries. As a result, we feel comfortable that this experiment was a good test of our standard survey question.

As mentioned earlier, 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-h period when filling out the standard survey, and, as a result, had overlooked or overstated some of their study hours.[18] Thus, although we stress that the time-diaries themselves undoubtedly contain some reporting error, we give in to the convenience of thinking of the time-diary responses as students' true time-use for this 24-h period.[19] Under this assumption, we find evidence of reporting error. The correlation between students' diary amounts and the standard survey reports was 0.72. 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. Consistent with the findings of previous work in other contexts that is detailed in Juster and Stafford (1985), the average person tends to overreport hours; the average person overreports study-time by 0.417 h and a test of the null hypothesis that the mean reporting error is zero is rejected with a t-statistic of 3.636.

As earlier discussion suggests, what is primarily important is whether the reporting error is classical in nature. If this is the case, reporting error should be related to only the transitory component of study-time and our previous estimators are

[18] A 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 24-h 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.

[19] Bound et al. (2001) discuss more generally the possibility that validation data may be imperfect.

Table 4
Determinants of reporting error (standard 24-h report minus 24-h report)

Constant	−0.778 (0.538)	−1.180 (0.737)	−0.158 (0.383)
$\widehat{\text{STUDY}}_i$	0.178 (0.078)*	0.166 (0.079)*	0.176 (0.088)*
$(s_{i,6} - \widehat{\text{STUDY}}_i)$	0.320 (0.080)*	0.317 (0.080)*	0.346 (0.095)
Male	0.171 (0.234)	0.216 (0.241)	
Black	0.698 (0.299)*	0.768 (0.312)*	
ACT	0.017 (0.018)	0.013 (0.018)	
Family income	0.0006 (0.008)	0.0009 (0.008)	
Parental education	0.033 (0.069)	0.031 (0.069)	
GPA		0.169 (0.212)	
Estimated std. $\kappa_{i,t}$	1.524	1.524	1.544 (0.076)*
Log likelihood			−303.989

Table shows regression of reporting error on permanent component of study-time, transitory component of study-time, and observable characteristics. Reporting error is defined to be difference between report on 24-h time-use question and detailed time-diary report. * represents t-statistic greater than two.

appropriate. However, it is also possible that reporting error is systematically related to the permanent component of study time. In this case, our previous estimators will be biased. The existence of our multiple time-use reports and the data from the survey design experiment allow us to examine the relative importance of the two possibilities. Defining reporting error $m_{i,t}$ to be the difference between the noisy measure of time-use (as measured by the standard time-use survey) and true time use, we regress $m_{i,6}$ on an estimate $\widehat{\text{STUDY}}_i$ of the permanent component of study-time, an estimate $(s_{i,6} - \widehat{\text{STUDY}}_i)$ of the transitory component of study-time, and the observable characteristics X_i using the 204 students who participated in our survey design experiment. The results are shown in column 1 of Table 4. The estimated standard deviation of the unobservable in this regression is 1.524 which indicates that, as noted before, a non-trivial amount of reporting error exists. With the exception of some evidence that black students tend to overreport, no evidence is found that reporting error varies systematically with the observable characteristics. However, overreporting increases with both the permanent component $\widehat{\text{STUDY}}_i$ and the transitory component $(s_{i,6} - \widehat{\text{STUDY}}_i)$ in a statistically significant manner with the relationship between reporting error and the latter being more pronounced.[20]

Given that reporting error is not entirely related to the transitory component, it seems worthwhile to examine adjustments to our estimators which explicitly take into account

[20] The effects of the permanent and transitory components are very similar if X_i is not included in the specification and the effects of X_i are very similar if the permanent and transitory components are not included in the specification.

Using the rationale that permanent income changes are often related to more important events, Pishke (1995) characterizes measurement error in earnings using a model in which individuals underreport the transitory component of earnings but in which measurement error is unrelated to permanent earnings. Here we find that, although errors related to transitory component of study-time are more pronounced than the errors related to the permanent component, both matter.

the stochastic process by which reporting errors arise. It is worth noting that biases can also arise if reporting error is correlated with the unobservable ε_i in the outcome equation (1).[21] Determining whether this is the case is 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 overreports his study-time by approximately $\frac{1}{2}$ hour a day and that overreporting increases with both the permanent and transitory components of study-time 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

$$m_{i,t} = \delta_1 + \delta_2\,\widehat{\mathrm{STUDY}}_i + \delta_2(s_{i,6} - \widehat{\mathrm{STUDY}}_i) + \kappa_{i,t}, \tag{16}$$

where $\kappa_{i,t} \sim N(0, \sigma_\kappa^2)$ and is uncorrelated with X_i and ε_i. Eq. (16) is estimated using the sample from our survey design experiment and the results are shown in column 3 of Table 4. The reality that we need to correct for measurement error in each of the six time-use surveys but can only estimate Eq. (16) for the participants in our survey design experiment which took place during the sixth time-use survey has two implications. First, we must assume that the relationship between the variables in Eq. (16) is the same for the sixth time-use survey as it is for all other time-use surveys. Second, because we cannot observe information about the joint distribution of $\kappa_{i,t},\ t = 1, \ldots, 6$, our simulation estimator proceeds under the assumption that conditional on $s_{i,t}$ reporting error is independent across time, $\mathrm{E}(\kappa_{i,j}\kappa_{i,k}) = 0\ \forall j \neq k$.[22]

The estimation problem necessitated by the existence of reporting error in the 24-h reports bears strong similarities to the work of Brownstone and Valletta (1996) who also deal with reporting error in a dependent variable.[23] In their application, self-reported earnings, which are assumed to be reported with error, are observed for all individuals. However, true earnings, as measured by administrative records, are observed for only a relatively small subset of their sample. Their multiple imputation approach

[21] For example, 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 (consciously or unconsciously) in terms of study-behavior. It is unclear whether individuals who are underperforming academically would tend to overreport or under report study-time. On one hand, students might feel that it is not desirable to be viewed (or view themselves) as lazy. On the other hand, students might feel that it is not desirable to be viewed as a low ability person who performs poorly even when effort is high.

[22] In Eq. (15) this implies that the draws of $m_{i,1}, \ldots, m_{i,Ni}$ are independent conditional on $s_{i,1}, \ldots, s_{i,Ni}$.

[23] See also Lee and Sepanski (1995). Bound et al. (2001) discuss the use of validation data to address problems issues of measurement error more generally.

Note that the 24-h study-times are dependent variables with respect to the study Eq. (4) but are also closely tied to the independent variable STUDY_i that enters the outcome Eq. (1).

proceeds by first specifying and estimating the distribution of true earnings conditional on self-reported earnings and other observable characteristics and then using this distribution to simulate true earnings for the subset of the sample for which true earnings are not observed.[24]

Our MLE described earlier can be adjusted to accommodate reporting error in a manner that is similar in spirit to the multiple imputation approach of Brownstone and Valletta (1996).[25] As in Eq. (9), the MLE is found by integrating the joint density of the dependent and unobserved independent variables (given the observed independent variables) with respect to any variables that are not fully observed. However, with reporting error in reported study-times, the set of variables for which integration must take place now includes the true 24-h study times $s^*_{i,1}, \ldots, s^*_{i,Ni}$,

$$L_i = \int \int g(s^*_{i,1}, \ldots, s^*_{i,Ni}, O_i, \mu_i | X_i)\, \mathrm{d}\mu_i\, \mathrm{d}s^*_{i,1}, \ldots, \mathrm{d}s^*_{i,Ni} \tag{17}$$

Given our definition of reporting error $m_{i,t}$ as the difference between the noisy measure of time-use (as measured by the standard time-use survey) and true time use. Eq. (17) can be rewritten as

$$L_i = \int \int g(s_{i,1} - m_{i,1}, \ldots, s_{i,Ni} - m_{i,Ni}, O_i, \mu_i | X_i)\, \mathrm{d}\mu_i\, \mathrm{d}m_{i,1}, \ldots, \mathrm{d}m_{i,Ni} \tag{18}$$

Rewriting Eq. (18) analogously to Eq. (10) leads to the simulator

$$L^s_i = \frac{1}{D} \sum_{d=1}^{D} g_1(s_{i,1} - m^d_{i,1} | \mu^d_i, X_i) \cdots g_1(s_{i,Ni} - m^d_{i,Ni} | \mu^d_i, X_i) g_2(O_i | \mu^d_i, X_i), \tag{19}$$

where μ^d_i is the dth of D draws from the distribution of μ_i and $m^d_{i,1}, \ldots, m^d_{i,Ni}$ represents the dth of D draws from the distribution of $m_{i,1}, \ldots, m_{i,Ni}$.

The results of the estimation which takes into account reporting error are shown in Table 5.[26] As compared to Table 2, the parameters are now estimated with somewhat less precision because the new estimator recognizes a new source of data uncertainty. The estimated effect of STUDY_i increases by approximately 20% in the linear case (column 1) and somewhat larger changes in point estimates are seen in the quadratic specification (column 2). The predicted increase in GPA associated with changing

[24] Brownstone and Valletta (1996) note that it is desirable to specify a model with true earnings as an explanatory variable in a measurement error equation but stress that their multiple imputation technique requires that observed interview earnings be used as an explanatory variable. An analogous situation exists here.

[25] Brownstone and Valletta (1996) suggest that their model could be estimated by maximum likelihood at the cost of “custom programming”.

[26] We note that, to some extent, our standard errors will be understated because we have not taken into account existing uncertainty about the parameters in Eq. (16) when we estimate the model given by the likelihood contribution in Eq. (14). 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.

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

	I	II
STUDY		
C_1	3.637 (1.088)*	3.907 (1.139)*
C_2	3.625 (1.056)*	3.912 (1.122)*
C_3	3.724 (1.047)*	3.981 (1.133)*
C_4	3.658 (1.061)*	3.923 (1.110)*
C_5	3.646 (0.990)*	3.884 (1.103)*
C_6	3.821 (1.120)*	4.079 (1.143)*
Male	−0.077 (0.394)	−0.228 (0.424)
Black	0.109 (0.361)	0.059 (0.363)
ACT	−0.033 (0.041)	−0.041 (0.044)
Family income	0.065 (0.090)	0.032 (0.075)
Parental education	−0.002 (0.632)	0.132 (0.296)
σ_μ-permanent	1.158 (0.090)*	1.151 (0.127)*
σ_μ-transitory	0.678 (0.036)*	0.697 (0.042)*
GPA		
Constant	0.817 (0.324)*	−0.133 (0.539)
Male	−0.326 (0.085)*	−0.240 (0.094)*
Black	−0.189 (0.096)*	−0.160 (0.101)
ACT	0.065 (0.011)*	0.070 (0.010)*
Family income	−0.002 (0.021)	0.003 (0.021)
Parental education	0.131 (0.075)	0.096 (0.070)
STUDY PER WEEKDAY	0.221 (0.043)*	0.740 (0.252)*
STUDY^*STUDY WEEKDAY		−0.073 (0.035)*
σ_ε	0.556 (0.032)*	0.501 (0.038)*
Log likelihood	−2885.09	−2881.15

Columns I shows estimates of full model using 24-h reports with likelihood contributions corrected for reporting error using Eqs. (16) and (19).
Standard errors are computed using robust methods of White (1982) and are in parentheses.

weekday study amounts from 1 to 2, 2 to 3, 3 to 4, and 4 to 5 h is shown in the second column of Table 3.

4.4. Causal interpretations

We stress that, due to the endogeneity of the study decision, we feel most comfortable viewing our results as being descriptive in nature. Nonetheless, it is worth discussing the issue of causality because this is the relationship of ultimate interest. The issue of causality between study-time and grades bears a strong resemblance to the much-studied issue of causality between education and earnings which is discussed in detail in Card (1999). Taken at face value in this context, the arguments in Card (1999) would suggest that our estimator will likely overstate the true causal effect that studying has on grades if students with high unobserved ability tend to study more because they find studying more enjoyable, rewarding, or otherwise less “costly” than other students or if students who have a higher return to studying choose to study more.

However, while the analogs to these conditions are likely satisfied in the education/earnings context, an important difference exists between the education/earnings context and the context studied here.[27] In the education/earnings context, individuals choose education to maximize discounted expected lifetime earnings. On the other hand, in the study/grade context it is likely to be lifetime earnings rather than grades that influences an individual's decision of how much to study. As a result, decisions regarding how much to study depend in a complex manner on the both the relationship between study time and grades and the relationship between grades and future earnings outcomes. To see that this makes the direction of likely bias uncertain, suppose that students believe that future earnings depend on only the amount of completed schooling and that grades only affect future earnings by determining whether a person is allowed to continue in school after each semester. Roughly speaking, in this scenario students would choose the minimum study effort that produces the grades necessary to remain in school. Thus, students with higher unobserved ability and higher grade returns to studying who need less study-time to reach this minimum would study less than other types of students and our estimator would tend to understate the causal effect of effort on grades. While the scenario described above is perhaps extreme, the general notion that students with lower unobserved ability and lower grade returns to studying might have to study harder in order to "stay afloat" seems very plausible.

Thus, the existence of competing factors makes it difficult to know what types of students will tend to study more. On one hand, higher non-pecuniary benefits and higher grade benefits are likely to have a partial effect of encouraging high ability students to study more. On the other hand, high ability students may not need to study as much to maintain the type of grade performance that they feel is reasonable given future earning considerations. This competing/offsetting situation provides a possible explanation for why observable measures such as ACT scores are found to be unrelated to actual study-time. If a similar offsetting situation exists for unobserved ability and returns, it would be reasonable to view our results as being causal in nature. At least some intriguing evidence that this could possibly be true is presented in column 3 of Table 3. 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 causal marginal returns to an additional hour of studying are shown in column 3 of Table 3.

It is unclear how accurate students views are about the causal relationship between studying and grades. Another possible approach for understanding the relationship between our estimator and the true causal effect is to compare our results to the estimate of the causal relationship between employment and grade performance in Stinebrickner and Stinebrickner (2001). Using a sample of freshmen who entered Berea between 1990 and 1997, Stinebrickner and Stinebrickner (2001) found that an additional hour

[27] In the education/earnings context an upward bias is likely to arise because students with high unobserved ability (e.g., students from privileged backgrounds) are likely to have lower costs of education and because individuals with higher returns to education are likely to obtain more education.

of work per week causes grades to decrease by 0.16. Ignoring possible differences in study-time on the weekends, our estimates from models where STUDY enters in a linear fashion suggests that an additional hour of studying per week has smaller effects (0.182/5 in column 1 of Table 2 and 0.221/5 in column 1 of Table 5).

At first glance this would seem to suggest that our estimator presents a downwardly biased view of the causal relationship between study-time and grades due to either an imperfect correction for reporting error or because worse students need to study more. However, there are other possible explanations as well. One possible alternative explanation is that the difference between the estimates arises because the students in our current sample are somewhat different than the students who were in the 1990–1997 sample used in Stinebrickner and Stinebrickner (2003a). As a result of a change in admissions at Berea, average total ACT scores are more than three points higher and average first-year grade point averages are approximately 0.5 higher for students in our current sample. This could matter for several reasons. For example, it is possible that current students at Berea study substantially more than the students who were at Berea during the earlier sample period. This would make comparing estimates of models in which STUDY enters linearly very difficult if STUDY indeed enters non-linearly as suggested by this paper.[28] Another possible alternative explanation is that study-time may decrease on average by more than one hour for each additional hour of work that takes place. Work-shifts at Berea are typically quite short so that additional hours of work will often imply additional shifts. This matters because jobs in the first year at Berea are typically service-type jobs which require not only commuting to and from work but also a substantial amount of transition time before and after work. Examples include changing clothes, taking showers, and perhaps resting after tiring physical labor. Unfortunately, the presence of these latter explanations imply that is difficult to learn much about the bias in our current estimator by comparing our current results with the causal results in Stinebrickner and Stinebrickner (2003a).

5. 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. Here we briefly examine what our work might suggest to survey administrators about the benefits of providing researchers with different types or different amounts of time-use data.

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-h period) or a retrospective question about

[28] Students in the earlier sample period may have been on a steeper part of the study/grade profile relative to current students. Estimates of the importance of a linear relationship between study-time and grades would then be higher for the previous group even if the true non-linear relationship is the same between groups.

time-use during the previous academic year (semester). Our work suggests that both of these options will tend to be quite problematic.

The problems associated with the former stem from the fact that it is not possible to characterize the importance of the transitory component $\mathrm{Var}(v_{i,t})$ given a single time-use survey. Since this is essential to both the measurement error and MLE estimators, the researcher may be forced to assume that the single time-use survey is a good proxy for the time-use variable of interest.[29] Our estimates at the end of Section 4.1 and in Table 1 together with Eq. (8) suggest that the OLS estimator of α_S based on the use of a single 24-hour time-use report as if it is the truth would be biased by a factor of approximately 2.38.[30]

Retrospective questions of the form "how much did you study on an average day during the last academic year" have the obvious appeal that they do not suffer from true day-to-day sampling variation of the type that is present in the daily time-diaries. At the end of the academic year, we asked individuals in our survey this question. Using the answers to this question as STUDY_i in Eq. (1), we found an estimated effect (std. error) of α_S of 0.086 (0.027). The fact that the estimated effect using the retrospective data is smaller than the results found throughout the paper is consistent with previous research that suggests that a substantial amount of recall error is likely to exist in retrospective questions of this form.[31]

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. Adding a second time-use survey which provides the possibility of discerning the importance of the permanent and transitory components of $s_{i,t}$ is extremely valuable.

To get a sense of the value of having more than two time-use surveys we conduct a small simulation study. We assume that data are generated by the model

$$O_i = 2.5 + 0.20\,\mathrm{STUDY}_i + \varepsilon_i \quad (17),$$

$$s_{i,t} = 3.5 + \mu_i + v_{i,t} \quad (18),$$

where $v_{i,t} \sim N(0, 1.70^2)$, $\varepsilon_i \sim N(0, 0.60^2)$, and $\mu_i \sim N(0, 1.30^2)$.[32] We assume that T is large so that $\mathrm{STUDY}_i = 3.5 + \mu_i$. The survey administrator can potentially choose the number of people n in the sample and the number of time-use surveys N that are

[29] Another alternative would be for the researcher to take a guess of the variance of the transitory component.

[30] $\mathrm{Var}(\widehat{\mathrm{STUDY}}_i)$ when $N = 1$ is given by the square of the standard deviations in the first column of the first six rows of Table 1(A). Assume e.g., that the one survey that was taken was survey 3. The standard deviation from survey 3 is 2.18. At end of Section 4.1 we found that $\mathrm{Var}(v_{i,t}) = 2.76$. Thus, Eq. (8) shows that OLS estimator will be too small by a factor of $2.18^2/(2.18^2–2.76) = 2.38$.

[31] This number was estimated using the 139 people who answered all of our time-use surveys. Recall from the end of Section 4.1 that, for this group, $\hat{\alpha}_{S,\text{measurement}} = 0.160$ and $\hat{\alpha}_{S,\text{MLE}} = 0.171$.

In addition, 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.

[32] These numbers were chosen so that they were close to the MLE estimates obtained from our data using the similar model at the end of Section 4.1.

Table 6
Sampling distribution of $\hat{\alpha}_S$ for different sample sizes and different numbers of time-use surveys

	$N=2$ mean (std deviation) of $\hat{\alpha}_S$	$N=4$ mean (std deviation) of $\hat{\alpha}_S$	$N=6$ mean (std deviation) of $\hat{\alpha}_S$	$N=20$ mean (std deviation) of $\hat{\alpha}_S$	$N=30$ mean (std deviation) of $\hat{\alpha}_S$
$n=300$	0.2030 (0.0268)	0.2006 (0.0228)	0.1993 (0.0212)	0.1989 (0.0195)	0.1994 (0.0190)
$n=900$	0.2012 (0.0167)	0.2007 (0.0113)	0.2009 (0.0100)	0.1996 (0.0082)	0.1994 (0.0076)

Each entry shows the simulated mean and variance of $\hat{\alpha}_S$ for a given sample size (n) and number of time-use surveys (N).

taken for each person. For a particular choice of n and N, a dataset can be simulated by drawing n realizations of $\{\mu_i, \varepsilon_i, v_{i,1}, \ldots, v_{i,N}\}$ and using these in Eqs. (17) and (18) to generate the n data realizations of $\{O_i, s_{i,1}, \ldots, s_{i,N}\}$. The parameters of the model described above can then be estimated from the simulated data by MLE.

The sampling distribution of the estimator can then be constructed by repeatedly simulating datasets and estimating the model parameters from the simulated data. In particular, for each choice of n and N we generate 150 sets of MLE estimates and compute the mean and standard deviation of these estimates. To simplify our discussion, we concentrate on the parameter of primary interest, the coefficient on STUDY_i in Eq. (17). Results are shown in Table 6. The first row shows estimates with $n = 300$, which is approximately the sample size used in this paper. Because the MLE estimator is unbiased, the average estimate α_S (the coefficient on STUDY_i) changes only trivially with N. However, as expected, the standard deviation is decreasing with N since more realizations of $s_{i,t}$ allows the model to be more sure about the value of μ_i for a particular person. In essence, as N increases, the distribution of $\text{STUDY}_i = 3.5 + \mu_i$ given the observation $s_{i,1}, \ldots, s_{i,N}$ has smaller variance. We note that increasing N from two to four leads to a non-trivial decrease in the standard deviation of the estimator. Beyond that, however, increasing N has a relatively small effect.

The sampling variation that remains when $N = 20$ or 30 is almost exclusively due to sampling variation in the ε_i's. The second row in Table 6 shows estimates with the sample size increased to $n=900$. As expected, each entry in row 2 is smaller than the corresponding entry in row 1. The effect of increasing N is similar in nature to the first row of Table 6. Note that the combination of $n=300$ and $N=6$ would entail the same total number of time-use surveys as the combination $n = 900$, $N = 2$. However, the latter combination produces an estimator with a somewhat smaller standard deviation (0.021 vs. 0.017). Thus, in this application, at a fairly small value of N it becomes worthwhile for a survey administrator to increase the number of individuals in the sample at the cost of reducing the number of time-use surveys per person.

An important point is that what is required for consistency is for the number of individuals in the sample to become large. As evidence of this, the estimator associated with $n = 30000$ and $N = 2$ has a mean of 0.2007 and a variance of 0.0011.

6. 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 individuals' reported beliefs regarding the causal relationship between studying and outcomes and are consistent in spirit with 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 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 the survey design experiment that we conducted suggests that non-trivial reporting error may exist even in narrowly defined questions of the form "How much did you study in the last twenty-four hours". Our findings suggest that the reporting error is most strongly related to the transitory component of study-time, but is also related to the permanent component of study-time. Modifications to our MLE that take this into account suggest that ignoring this non-classical reporting error leads our estimator of the relationship between study-time and grades to have a downward bias of approximately 20%.

Acknowledgements

Generous support for this project was provided by The Mellon Foundation, The Social Science Humanities Research Council (SSHRC), Berea College, and the Canadian Imperial Bank of Commerce. This project would not have been completed without the invaluable research assistance of Anne Kee, Lori Scafidi, Dianne Stinebrickner, and Pam Thomas. We are also grateful for the useful comments that we received from two referees. Numerous useful discussions took place at the Osceola Research Institute and an earlier version of this paper was circulated as part of the ORI working paper series.

Appendix A.

When T is not large, the estimator must take into account that $\mathrm{Var}((1/T)\sum_{t=1}^{T} v_{i,t})$ may not be close to zero. This influences the calculation of the STUDY_i variable slightly. Consider a person who has completed all six time-use surveys. For this person,

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

Eq. (A.1) indicates that STUDY_i is the average of $s_{i,1},\ldots,s_{i,T}$. The values $s_{i,1},\ldots,s_{i,6}$ are observed and the observed values are used in computing STUDY_i. $s_{i,7},\ldots,s_{i,T}$ are not known but have a stochastic process which is given by

$$s_{i,t} = C_t + \beta X_i + \mu_i + v_{i,t}, \qquad t = 7,\ldots,T. \tag{A.2}$$

We do not have information about the values of $C_7,\ldots,C_T$ since time-use surveys are not collected at these times. Instead, we assume that $(T-6)/6$ of the constants $C_7,\ldots,C_T$ are the same as C_1, $(T-6)/6$, of the constants $C_7,\ldots,C_T$ are the same as $C_2,\ldots,$ and $(T-6)/6$, of the constants $C_7,\ldots,C_T$ are the same as C_6.[33]

The likelihood contribution for person i becomes

$$L_i = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\sigma_\varepsilon^2 + \alpha_S^2\sigma_{\bar v}^2}}\,\varphi\left(\frac{O_i - \alpha_S\,\mathrm{STUDY}_i^*(\mu_i, C, \beta) - \alpha_X X_i}{\sqrt{\sigma_\varepsilon^2 + \alpha_S^2\sigma_{\bar v}^2}}\right)$$
$$\times \prod_{t=1}^{N_i} \frac{1}{\sigma_v}\,\varphi\left(\frac{s_{i,t} - C_t - \beta X_i - \mu_i}{\sigma_v}\right) h(\mu_i)\,\mathrm{d}\mu_i, \tag{A.3}$$

where $\bar v = (1/T)\sum_{t=7}^{T} v_{i,t}, \sigma_{\bar v}^2 = \mathrm{var}(\bar v)$, and $\mathrm{STUDY}_i^* = \mathrm{STUDY}_i - \bar v$, and φ is the standard normal probability density function. Eq. (A.2) takes into account variance of $\bar v = (1/T)\sum_{t=7}^{T} v_{i,t}$. When T is large, this variance approaches zero and Eq. (A.2) is identical to Eq. (10).[34]

References

Betts, J., 1997. The role of homework in improving school quality. Unpublished manuscript.

Bound, J., 1991. Self-reported objective measures of health in retirement models. The Journal of Human Resources 26 (1), 106–138.

Bound, J., Brown, C., Mathiowetz, N., 2001. Measurement error in survey data. In: Leamer, E.E., Heckman, J.J. (Eds.), Handbook of Econometrics, Vol. 5, North-Holland, Amsterdam, pp. 3705–3843.

Brownstone, D., Valletta, R., 1996. Modeling earnings measurement error: a multiple imputation approach. Review of Economics and Statistics 78, 705–717.

[33] Information is available about $C_1,\ldots,C_6$ since $t = 1,\ldots,6$ are the weeks in which time-use is observed. This assumption just says that the average study time in each of the observed days (weeks) is like that in $T-6/6$ unobserved days (weeks) during the year.

[34] We assume that there are 24 weeks and 120 weekdays during the academic year The 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.

Card, D., 1999. The causal effect of education on earnings. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, Vol. 3A, Elsevier, New York.

Ehrenberg, R., Sherman, D., 1987. Employment while in college, academic achievement and postcollege outcomes. Journal of Human Resources 22 (1), 1–24.

Heckman, J., Singer, B., 1984. A method for minimizing the impact of distributional assumptions in econometric models for duration data. Econometrica 52, 271–320.

Juster, T., Stafford, F., 1985. Time, goods, and well-being. Survey Research Center, Institute for Social Research, The University of Michigan.

Juster, F.T., Stafford, F., 1991. The allocation of time: empirical findings, behavioral models, and problems of measurement. Journal of Economic Literature 26, 471–522.

Lavy, V., Palumbo, M., Stern, S., 1998. Simulation of multinomial probit probabilities and imputation. In: Fomby, T., Hill, R.C. (Eds.), Advances in Econometrics. JAI Press, Greenwich, CT.

Lee, L.F., Sepanski, J., 1995. Estimation of linear and nonlinear errors-in-variables models using validation data. Journal of the American Statistical Association 90, 130–140.

Little, R.J.A., Rubin, D.B., 1987. Statistical Analysis with Missing Data. Wiley, New York.

Pishke, S., 1995. Measurement error and earnings dynamics: some estimates from the PSID validation study. Journal of Business and Economic Statistics 13, 305–314.

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

Stern, D., Nakata, Y.-F., 1991. Paid employment among U.S. college students: trends, effects, and possible causes. Journal of Higher Education 62 (1), 25–43.

Stinebrickner, T., 1999. Estimation of a duration model in the presence of missing data. The Review of Economics and Statistics 81 (August), 529–542.

Stinebrickner, R., Stinebrickner, T., 2003a. Working during school and academic performance. Journal of Labor Economics 21 (2), 473–491.

Stinebrickner, R., Stinebrickner, T., 2003b. Understanding educational outcomes of students from low income families: evidence from a liberal arts college with a full tuition subsidy program. Journal of Human Resources, Summer 2003, 591–617.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1–25.