Effort and Reward: The Assumption that College Grades Are Affected by Quantity of Study*

HOWARD SCHUMAN, *University of Michigan*
EDWARD WALSH, *Pennsylvania State University*
CAMILLE OLSON, *San Francisco, California*
BARBARA ETHERIDGE, *Ada, Michigan*

Abstract

The relation between college grades and self-reported amount of effort was examined in four major and several minor investigations of undergraduates in a large state university. Grades were operationalized mainly by using grade point average (GPA), though in one investigation grades in a particular course were the focus. Effort was measured in several different ways, ranging from student estimates of typical study over the term to reports of study on specific days. Despite evidence that these self-reports provide meaningful estimates of actual studying, there is at best only a very small relation between amount of studying and grades, as compared to the considerably stronger and more monotonic relations between grades and both aptitude measures and self-reported class attendance. The plausible assumption that college grades reflect student effort to an important extent does not receive much support from these investigations. This raises a larger question about the extent to which rewards are linked to effort in other areas of life—a connection often assumed but seldom investigated.

That hard work along normatively prescribed lines produces rewards is a belief fundamental to virtually every social system (Heider). The belief is certainly widespread in the United States (Rytina et al.) and is probably shared to at least some degree even by those who emphasize the impor-

*Several of the senior author's Research Methods classes participated actively in both the design and the execution of the main research described here. We are also grateful to Diana Pearce, Jean Converse, and Daniel DuRoss, each of whom collaborated at some point in our investigations; to Marc Schuman and Jacob Ludwig who provided consultation on computing throughout; and to James House and F. Thomas Juster who offered suggestions on an earlier draft of this paper. Address correspondence to Howard Schuman, Department of Sociology, University of Michigan, Ann Arbor, MI 48109-1382.

tance of other factors such as native ability, social attributes, or sheer luck. Yet the measurement of effort is very difficult, and we do not know of any sociological investigation that attempts to attack the problem directly, though there are a variety of studies of indirect relevance in that they show the influence on income of presumably non-effort factors such as race (e.g., Siegel) or sex (e.g., Featherman and Hauser). Whether the homilies of our culture—for example, "Genius is one percent inspiration and ninety-nine percent perspiration"—have in fact much truth remains not only unsubstantiated but largely uninvestigated.

We do not propose here to remedy this defect for the marketplace, nor for other areas of life from athletics to Zen where success is thought to go to those who strive most mightily for it. But we do present surprising evidence for one area where effort is ordinarily demanded, where irrelevant ascriptive factors would seem to be minimized, and where rewards might be thought to depend at least in part on the quantity of effort put forth: grades in college.

Investigating the effort–grade relationship in schools has several attractive features. First, grades provide a clear analogue to a merit system of salary or wages. Those who do the work, as assessed by exams, papers, or whatever other measures teachers may use, are rewarded according to official theory by higher grades.

Second, there is some evidence that those who take part in or observe the school system do indeed believe that effort and grades are strongly related. College students in one of the studies to be described were asked whether they thought studying very hard was a "useful," "somewhat useful," or "not useful" way to get good grades, and the choices registered were 69, 25, and 6 percent, respectively. (Much less support was registered for several other strategies, such as "cramming" [38 percent, useful], "training in study skills" [12 percent, useful], or "cheating" [15 percent, useful].) At least some sociologists share this general belief in the importance of study, as witness Collins' claim for public schools that "the sheer amount of school work done seems to be the best predictor of high grades" (21). (Collins cites as his only evidence a book by Sexton, which does make such a claim, but Sexton also provides no direct evidence.) On the basis of a similar assumption, presumably, Jencks and Crouse argue that colleges should substitute achievement tests for Scholastic Aptitude Test (SAT) scores in the admission process, since this "would help reinforce the traditional American notion that effort will be rewarded."

Finally, all students at the start of a course are equal in a sense in which incumbents of different occupations are not: we cannot realistically expect the industrious janitor to earn the same amount of money as the similarly industrious doctor, but their equally hard-working sons or daughters might be expected to do equally well in terms of grades, other

things being equal of course; at least they are not paid at totally different rates for their work as are their parents. For all these reasons, the connection of effort and reward can usefully be investigated in terms of a school setting with perhaps as much expectation of it being positive here as in any other sphere of life.

Grades, of course, are generally believed to be a function of ability as well as of effort. But the ability factor has long been the subject of attention, as represented by the development of SAT and similar aptitude measures, along with innumerable investigations of the extent to which such tests can predict grades (e.g., Lavin). The effort factor has been much less thoroughly investigated, probably in part because it is much harder to operationalize in test terms, but perhaps also in part because it has seemed almost self-evident that effort plays some role in academic performance and therefore in grades.[1] It is "effort" that is our focus here.

In this paper we operationalize effort as the *quantity* or amount of studying or other course preparatory work that college students do, as distinct from the *quality* of the work (for example, ease of comprehension) as it might be assessed through aptitude measures. The research was undertaken when an incidental finding in a 1972 survey of University of Michigan undergraduates showed an essentially zero association between self-reported grade point average (GPA) and a single self-report question on amount of study. Our initial interpretation of the null finding was that one or both of the key variables was inadequately measured, or that there was some interfering variable. Subsequent research has been devoted to attempting to overcome these problems and produce a positive relation between amount of study and GPA. This paper reports four different major investigations and several minor ones over a decade, none of which was very successful in yielding the hypothesized substantial association despite varying attacks on the difficult problem of assessing quantity of study, the somewhat easier problem of assessing grades, and further problems of taking into account other possible variables, such as academic aptitude, that might prevent discovery of a true relation. Our results are limited to a single major university, are subject to various qualifications discussed below, and (it must be admitted) leave a certain amount of disbelief among the investigators themselves. But we think the evidence is strong enough to deserve reporting, if only to encourage improved research leading to its refutation. Specific hypotheses are included with each study.

First Investigation: General Study Time and GPA

The purpose of our first serious investigation in 1973 was to devise an adequate summary index of the number of hours each student typically

studied during a term; to provide some validation of the index; to relate the index to an objective measure of GPA for that term; and finally to incorporate in the analysis other variables relevant to GPA and possibly conditioning or otherwise affecting the relation of hours studied to GPA.

PROCEDURE

We drew a systematic random sample of students from the Literature, Science, and Arts (LS&A) College (*N* about 12,000) of the University of Michigan, using the University's registration records for that term as the sampling frame. Of the 522 students drawn, 424 (81.2 percent) were successfully interviewed; the refusal rate was only 1 percent, the balance of nonrespondents being students not located due to inadequate or erroneous addresses in the registration office. Interviews were carried out by students in our Social Research Methods course over a 17-day period in the middle of the Fall term, 1973, the interviewers themselves having participated in the design of the study. Post-interview validity checks assured us that the interviewing had been done with care.

Although our questionnaire obtained self-reports of both GPA and SAT scores, we later were able to obtain this information from university records, and use the latter more objective measures in analyses, except where otherwise indicated.[2] The key variable based on student reports, therefore, is Hours Studied, and since our main conclusions hinge on the validity of this measure, we invested considerable effort in obtaining and checking it. The series of questions used is given in the Appendix.

The main Hours Studied index employed here was based on a four to three weighted average of the "typical" number of hours a student reported studying on weekdays "during the past week or so" and the "typical" amount reported for most weekends this term (see Appendix). We kept the weekday total to the recent past for concreteness, but it seemed better to avoid relying on a single weekend because of the variable dates of our interviewing and of student activities. Weekday and weekend studying reports are moderately correlated ($r = .47$), and use of either alone did not appear to change our conclusions.[3] Using this index, the median number of hours of studying reported per day is 3.8, with half the sample falling within the range 2.7 to 4.9. These figures may seem high, but for our purposes the important point is whether students are ordered correctly along the study dimension, not its absolute level.

Validation of the Hours Studied Index

Perhaps the best way to validate a self-report measure of studying would be continuous observation of students over the term, though pilot work in the university library revealed that even direct observation is often not

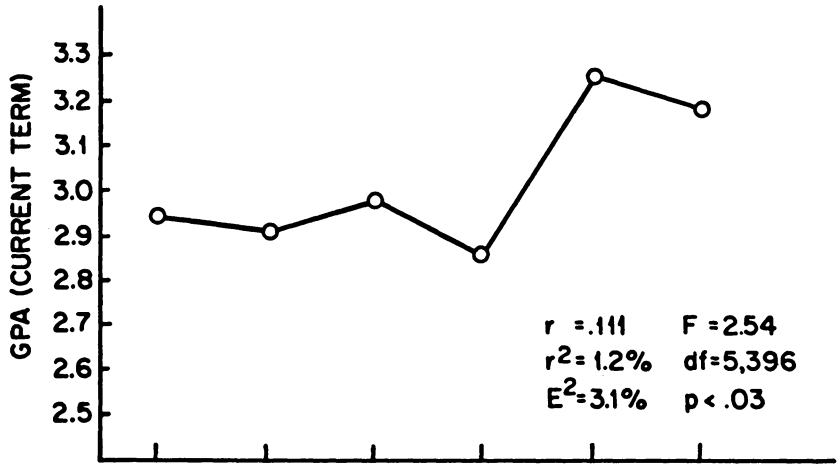
Table 1. HOURS STUDIED BY FIELD OF CONCENTRATION

Major	N	Mean	S.D.
Preprofessional natural sciences	43	3.9	1.4
Natural sciences	108	3.6	1.5
Social sciences	180	3.2	1.5
Humanities	57	3.0	1.7

able to distinguish studying from doodling, reading noncourse books, or especially daydreaming. In the absence of a large-scale observational project—which might also create serious problems of reactivity—we have several indirect indications of the validity of the Hours Studied index. First, the index is significantly related ($r = .19$, $p < .001$) to the number of credit hours students reported carrying for the same term. Second, Hours Studied is related to major field of concentration in a way that fits our and most campus beliefs about course demands, as shown in Table 1. Highest study is reported by Natural Science students, especially premedical students, and appreciably lower study by Social Sciences and Humanities majors ($F = 4.39$, $df = 3, 384$, $p < .01$; $E^2 = 3.3\%$). Third, several weeks after our main interviewing was completed we randomly selected nine "high studyers" (defined as persons studying 6 hours or more a day) and nine "low studyers" (defined as 1 hour or less a day), called them by telephone, and used open-ended indirect questions to obtain a detailed account of their total expenditure of time the previous day. Despite likely regression effects, the median hours studied of the two groups differed by 4 in the direction expected ($p < .01$ using the Mann–Whitney U test because of the small sample size). This finding is reassuring on two points: that our standard questions on study had not been too rigid in formulation and that responses to it were not limited to the particular point in time when they were obtained.

HOURS STUDIED AND OTHER PREDICTORS OF GPA

The relation of GPA to Hours Studied is presented in Figure 1, using both linear and nonlinear measures of correlation. A weak but marginally significant association is shown, but with the expected positive direction appearing only at the high end of the study variable. The relation is not strengthened by standardizing for credit hour load during the term, nor by taking account of responses to a further question on whether recent amount of study time had changed from that earlier in the term, nor is



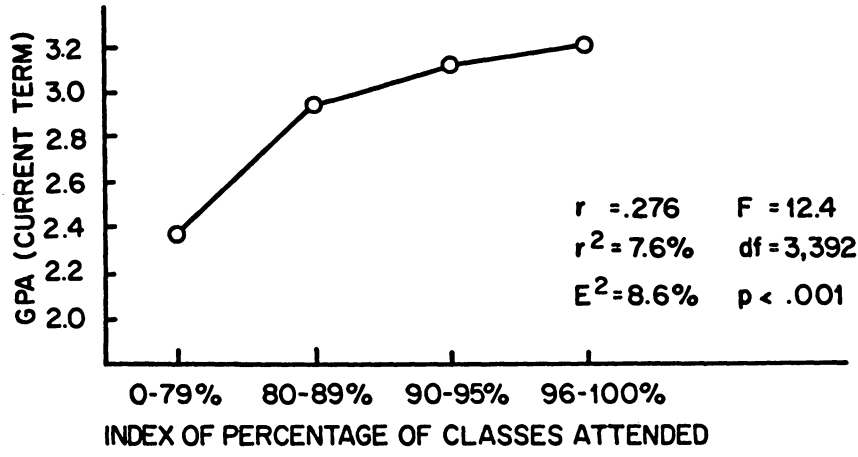
	N	(46)	(69)	(100)	(89)	(53)	(45)
MEAN	GPA	2.94	2.91	2.97	2.86	3.25	3.18
	S.D.	.81	.81	.77	.82	.67	.59

Figure 1. GRADE POINT AVERAGE BY HOURS STUDIED (1973)

there any other variable in our survey that shows associations with both Hours Studied and GPA such that its inclusion in the prediction equation noticeably improves the relation presented in Figure 1.

If we assume for the moment that sheer amount of study does not account for much of the variation in GPA, what other relevant factors are there? Perhaps the most obvious is SAT scores, although such scores have already had part of their effect absorbed in the initial university admission process and cannot be expected to account for a great deal of variation within an already highly selected college population. In fact, the combined Verbal and Math SAT scores condensed to an eight level total (TSAT) show a correlation (r) of .25 ($p < .001$) with GPA, and the relation is generally positive and monotonic throughout.[4] TSAT scores and Hours Studied are essentially uncorrelated, the product-moment r being $-.05$ ($t = .94$, $n = 379$, $p > .30$). This argues against the otherwise reasonable possibility that high aptitude students reduce their studying because they perceive less need for it.

One other measure included in our questionnaire turned out to be an important predictor of GPA. We asked respondents to report the "per-



N	(39)	(64)	(98)	(195)
MEAN GPA	2.38	2.85	3.03	3.14
S.D. of GPA	1.01	.76	.72	.67

Figure 2. GPA BY CLASS ATTENDANCE INDEX[a]

[a]As the standard deviations suggest, the variances of GPA for the four Class Attendance categories differ significantly among themselves, making the F-ratio for the difference among means of uncertain interpretation. Therefore, the nonparametric Kruskal–Wallis one-way analysis of variance was also used: $h = 24.5$, $p < .001$.

centage of the regularly scheduled classes [that] you get to" for each enrolled course. We then weighted these percentages by the credit hours for each course to form a Class Attendance index. The index is quite skewed and does not reveal much absolute variation, yet as Figure 2 shows, it is a much better predictor of GPA than Hours Studied and even slightly better than SAT scores. (It should be noted that we were as dependent on student reports for Class Attendance as for Hours Studied.) The effect of Class Attendance is greatest at the lowest level, unlike Hours Studied where the effect occurs only at the highest level.

We are not able to pin down exactly what it is about Class Attendance that creates this relatively strong relationship. Neither self-reported participation in class discussions nor amount of note-taking in classes is associated with GPA or affects the relation of GPA to the Class Attendance index. Moreover, we cannot even be certain of the causal direction of this GPA–Attendance relationship, for it is possible that students who are already doing poorly in their academic work then cease to attend classes regularly. Yet the results do fit a common student belief that exams and

other grading are based largely on what is said in lectures, rather than on assigned reading or other extra-class activity.[5]

Class Attendance, Hours Studied, and Total SAT scores are the three significant predictors of GPA in our study that can readily be interpreted in terms of direct effects on grades. If all three are included in a multiple regression equation, together they yield an R^2 (adjusted for degrees of freedom) of .15. The intercorrelations among the three predictors are small (for Class Attendance x Hours Studied, $r = .23$; for each of the two with TSAT, r is .08 and $-.05$, respectively) and the net contribution of each to the total explained variance is very little different from the E^2's already reported. Standardized regression coefficients for TSAT and Class Attendance are each highly significant ($B_2 = .28$, $p < .001$ and $B_3 = .28$, $p < .001$, respectively), but the coefficient for Hours Studied barely reaches borderline significance ($B_1 = .08$, $p = .10$).[6]

Specifications

Since there appears to be *some* relation between Hours Studied and GPA, it becomes important to discover whether the relation is heightened within certain identifiable categories of students. It is easy to hypothesize a number of such possible interactions, but here we simply present results for four academic variables of general importance.

As an ability measure, we divided students into approximate tertiles by TSAT scores, and used this three-category variable as the control in an analysis of covariance, with GPA as the dependent variable and Hours Studied as the interval level independent variable. The hypothesis of equal slopes cannot be rejected ($p > .40$), though there is a slight trend for the relation in Figure 1 to be located mainly among the high SAT students. If the trend should prove replicable, it indicates that whatever relation between studying and grades exists reflects primarily a lowering of grades among high SAT students who do not study, rather than a raising of grades by low or medium SAT students who study a great deal.

The Hours Studied–GPA relation was also examined within tertiles of the Class Attendance index and within the four college years. For the former we speculated that high studying might either intensify the effect of Class Attendance or substitute for lack of attendance. For the latter, no particular hypotheses were advanced, but some sign of a systematically increasing or decreasing Study–GPA relation by college year would provide clues to the nature of the relationship. However, neither examination of graphs nor formal analysis of covariance offers any hint of reliable interactions involving either Class Attendance or college year.

A fourth possible interaction examined, using major field of concentration, has special importance because it provides some reduction in the ambiguity of the basic GPA measure. While GPA is an important summary index of academic success, it is possible that the relation of studying

to grades occurs within courses, but then is swamped by other types of between course variations in grading. Specifically, Table 1 showed that Hours Studied varied by student majors, yet other analysis indicates no difference in mean GPA by major. It seems likely that grading is done relatively within each course, even though average studying varies greatly among courses. Thus studying would have both a between and a within course component, while grade variation would have only within course variation. This hypothesis can be adequately tested only on a course by course basis (see the second investigation below), but an approach can be made by considering major areas of concentration separately. Using the four broad concentration areas shown in Table 1, we found no consistent increase in the Study–GPA relation within any concentration.[7]

Second Investigation: Study and Grades for a Specific Course

Our main analysis to this point has employed overall GPA to represent college grades. This is justified by the social significance of GPA as a measure of student achievement, whatever GPA may actually represent. Nevertheless, it is important to discover whether the tiny degree of relationship found in the previous investigation holds within separate courses as well. Furthermore, it also seemed possible that student reports of study for a single course might be more accurate than for all course work treated as a whole.

We chose a general organic chemistry lecture course as one with a large number of students, a systematic method of grading, and presumably a content for which amount of study should be of some importance. With the instructor's agreement we administered a very brief questionnaire at the beginning of a class hour late in the Fall term, 1973, emphasizing to students that the instructor would not see their reports of studying or class attendance.[8] We also asked their permission to obtain their final grades at the end of the term, and our analysis covers only the 114 students who consented to this request by appending their names to the end of the questionnaire. Since there were 206 students officially registered at the beginning of the course, we missed some 70 students who had dropped out, did not attend class that day, or returned no questionnaire, plus 22 who returned questionnaires but did not give their names. Despite the rather large nonresponse rate, the results are valuable, though they cannot be unequivocally generalized back to the full class, nor of course to other classes or subjects.[9]

Table 2 presents the intercorrelations among the eight variables obtained on these students. Three are measures of achievement in the course: the final course grade and two of the five test scores on which that final grade was based. One of the specific tests was actually given in a

Table 2. INTERCORRELATIONS AMONG GRADE AND PREDICTOR VARIABLES FOR A SINGLE CHEMISTRY CLASS[a]

	Study					Grades		
	Yesterday	Past Week	Weekends	Term	Attendance	Test 3	Test 4	Final
Yesterday	--							
Past 2 weeks	.27**	--						
Weekends	.09	.21*	--					
Term	.25**	.26**	.28**	--				
Class attendance	-.13	.02	.09	.05	--			
Test 3	.01	.08	.11	.09	.35**	--		
Test 4	.05	.13	.15	.09	.21*	.63**	--	
Final grade	.06	.07	.08	.12	.26**	.81**	.85**	
Mean	5.1 hrs.	2.0 hrs.	5.5 hrs.	1.5 hrs.	94.0%	64	77	326
S.D.	2.3	1.2	3.8	1.0	9.9	23	13	74

[a]Correlations involving Study Weekends are based on N=111; correlations involving Study Term on N=113; all other correlations on N=114.

*.05 level of significance; **.01 level of significance.

special session later the same day on which our questionnaire was administered, making the student's report of his or her studying the day before of direct relevance. This probably accounts for the high mean in Table 2 for our first variable, Study Yesterday, as indeed was pointed out on a large number of the questionnaires themselves. Study on weekdays over the past week is less than half as large, and over the entire Term study averages to a still lower level. (Weekend study figures are higher because they represent the entire weekend, not the average per day.) Some idea of the variation in these figures can be gained by noting that 28 percent of the sample reported averaging one hour or less of study on weekdays over the previous week, while 22 percent reported averaging three or more hours a day during the same period. Practically nobody claims to have not studied at all, but above that zero level we seem to have registered a fair degree of variation. There is less variation on the Class Attendance variable, with 57 percent of the class reporting 99 or 100 percent attendance over the term, 30 percent reporting 90 to 98 percent attendance, and only 13 percent reporting less than 90 percent attendance.

As Table 2 shows, the two tests and the final grade are all substantially intercorrelated, although the relations involving the final grade are artifactually inflated to a small extent due to the inclusion of the test scores in the calculation of the final grade. Likewise, the three measures of studying are all significantly though weakly correlated, with one exception (Yesterday and Weekends) for which we have no explanation. In addition,

for this course the associations between Class Attendance and the several measures of studying are weak, and in fact the highest one is negative.

Turning to the causal relations that are our primary interest, the pattern of associations for this single class is quite similar to that found in the previous investigation. Reports of Class Attendance are modestly but significantly related to each of the three grade indicators, despite the fact that variation in such attendance is small and we almost certainly underrepresent those with low attendance records. If attendance is condensed to allow its use as a categorical variable in a simple analysis of variance, the final numerical grade for those attending less than 90 percent of the time averages 279; for those attending 90–98 percent of the time, 312; and for those attending 99 or 100 percent, 345.

On the other hand, none of the four measures of studying in Table 2 shows a significant association with any of the three types of grades, although all the correlations are in a positive direction. Several of the relationships were also examined using simple analysis of variance to allow for nonlinearity, but this did not produce signs of meaningful relationships different from those in Table 2. In addition, regressing Final Grade on Studying Weekdays, Weekends, and over the Term produces a multiple correlation of only .10, which does not approach significance ($F = .38$, $df = 3, 107$). Finally, we may note that Test 4, which was given on the day of our questionnaire administration, is predicted a bit more successfully by two of the study measures than are other grades, but the difference is trivial and nonsignificant; moreover, variation in intensive study the day before the test seems to have no special effect at all.

In sum, our data for this single class provide even less evidence than the larger survey that quantity of studying has any appreciable effect on achievement as measured by grades. We would be inclined to regard lack of variance in our measures of studying, or lack of truthfulness in student reports, as possible explanations, except that both explanations would also seem to apply to self-reported class attendance. Yet the latter is significantly and meaningfully related to grades in this chemistry class, just as it was in the general survey carried out in our first investigation. It is possible of course that there is something peculiar about organic chemistry that eliminates the value of study, and until similar research is carried out in other types of courses, these results should not be overgeneralized. It is indeed hard to imagine how amount of study could fail to be involved in grade achievement in language or history classes, though chemistry was in fact chosen for investigation because a substantial amount of new information needs to be learned.

Third Investigation: Measuring Effort Unobtrusively

The purpose of this investigation was to attempt a different and perhaps better approach to measuring Hours Studied. It seemed possible that students could not report accurately enough on study because they were not used to thinking in such terms. The task involved in classifying and summarizing study periods over an extended time might simply be too difficult. Our approach in Investigation III therefore was to say nothing to students about our interest in studying, but rather to ask each respondent to provide a detailed "time chart" of *all* their activities for the day previous to the interview. (Interviewing was arranged so that the "previous day" always referred to a weekday in order to provide as standard a frame of reference as possible, a seemingly sensible way of proceeding at that early point.) We then constructed a measure of the time each student had apparently studied by adding all such references. The measure has the advantage of focusing respondent attention on a concrete and full account of his or her activities for a 24-hour period. Both self-consciousness about studying and problems of memory should also be minimized with this approach. The students for Investigation III came from a random sample of the LS&A College, final $N = 273$, with interviews carried out in the middle of the Winter term, 1975, following procedures described for Investigation I.[10]

Although student accounts of a single day may be exceedingly accurate, a sample of only one day in a term might not provide a reliable enough indicator of total studying. Obviously factors can occur on a given day that greatly raise or lower a student's normal pattern of studying, though we expected that such exceptions would tend to cancel out and that there would be enough typicality captured to yield the hypothesized relationship. Evidence that the time-use approach does produce a meaningful Hours Studied measure comes from the relation of the new index to major field. Preprofessional and regular Natural Science majors are almost tied for highest hours studied, with a considerable gap between both groups and Social Science majors, followed then by a smaller gap to the lowest scores for Humanities majors ($F = 2.41$, $df = 3, 262$, $p = .07$; $E^2 = 2.7\%$). The result is not quite as strong or reliable as in Investigation I (Table 1), but given the close similarity to the earlier pattern and the smaller sample in Investigation III, the findings add confidence to the time-use based measure. Investigation III also produced, as expected, a somewhat lower Hours Studied median (2.9) than Investigation I (3.8), though still higher than we had thought likely.

We also asked respondents whether "yesterday was a fairly typical weekday as far as the way you spent your time?" Those who said no were asked in what ways it was not typical; 19 percent of the total sample

mentioned having studied less than usual, 25 percent mentioned having studied more than usual, and later analysis was run with and without these self-reported atypical respondents. Finally, after obtaining detailed information on time use, we asked the following question to all those who reported any time in study:

A person rarely spends 100% of study time actually studying. Some time is usually spent getting settled down, being distracted, or just plain daydreaming. Of the ____hour(s) that you were studying yesterday, how much of the time would you say you were actually working?

The Investigation III measure of Hours Studied shows no correlation at all with GPA for the total sample ($r = -.02$), nor for the subsample of those who regarded their studying for the day as typical ($r = -.03$). Nor is there any sign in either scattergram of a meaningful nonlinear association. Use of the information about "actual studying" vs. total study time does not alter these negative findings.

Fourth Investigation: Cumulating Measures of Effort Over Time

In searching for an explanation for the failure to find the expected relation between study and grades, one might note recent research pointing to the unreliability of single-act measures (Epstein, a; Fishbein and Ajzen). It is possible that studying on a single day, no matter how accurately reported, is too unrepresentative to provide a useful predictor of grades. Moreover, student answers to questions about their "typical" studying may either be too highly influenced by the immediate past, or simply wildly inaccurate for other reasons. Although we have presented indirect evidence against either of these possibilities, an attempt was made in 1979 to directly sample student studying on more than one occasion during a term.[11]

At the beginning of the 1979 Fall term a stratified random sample of 64 students was drawn from the same total Michigan undergraduate body studied previously. Stratification was used to allow equal proportions from each year (1st to 4th), since the relatively small size of the sample might have allowed too much inequality in this respect. Forty-nine (77 percent) of these 64 students were successfully interviewed by telephone at three points over the term (early November, the third week in November, and early December), and these students constitute the sample analyzed here. The investigation was presented to the students as a class project, and included among other items a set of questions on time spent the previous day in various activities including (but not limited to) study. "Typical" time spent studying "during the past several weeks" was also obtained at each time point, as was study time lost due to distractions or daydreaming. The questions on study were not obtrusive and students were not told the

main purpose of the investigation until after the final interview. Grade point averages and SAT scores were later obtained from official records. (GPA was also requested at the end of the third interview, and there is little evidence of an upward biasing of respondent reports of this information.)

The main correlational results are presented in Table 3. The primary conclusion from this panel study is that GPA is not significantly predicted from hours studied "yesterday" at any of the three time points, nor from any combination, including the average of all three ($r = -.05$, $p > .50$). Graphs also fail to show meaningful nonlinear trends. When the values for a "typical hours studied" question similar to that in Investigation I are averaged over the three time points, there is a slight positive trend at the highest end of the study dimension resembling that noted earlier for Figure 1, but it does not approach significance and could easily be due to chance. These main conclusions hold both with and without a control for SAT scores, which here as previously are unrelated to any measure of Hours Studied.

Several other findings about studying are of interest. The Hours Studied Yesterday measures from the first and the third of the three time points are moderately correlated ($r = .47$), but the correlation of times one and three with the measure obtained at the second time point are both near-zero. However, reports of typical hours studied from the three time points are all substantially intercorrelated (r's range from .51 to .61), and all three are also significantly correlated with Hours Studied Yesterday at times one and three, but not at time two. From this set of associations we conclude that: (a) at time point two, which fell during the days just preceding the Thanksgiving vacation, actual studying patterns were atypical, but (b) student reports of typical studying reflect well their actual studying over the entire term (since even at time two the typical measure continued to correlate well with all reports from times one and three). In sum, this investigation increased our confidence in the accuracy with which students report their amount of studying, but it did not increase the correlation of study with GPA, nor did it cast doubt on the validity of the measure of "typical study" that we employed in our earlier investigations. The latter, as Epstein (b) suggests, may in some cases be better than records of individual days, since it can take account of fluctuations in study time over the term.

One other finding that supports our earlier results is a relatively strong negative association between GPA and reports at time three of the average number of classes missed per week during the term: $-.46$ ($p < .001$). Again, it is class attendance much more than hours studied that helps explain variation in grade point averages. SAT scores also show a strong relation to GPA: $r = .44$ ($p < .01$).

Table 3. INTERCORRELATIONS (*r*) AMONG HOURS STUDIED AND GPA MEASURES (1979)[a]

	HY1	HY2	HY3	TS1	TS2	TS3	AVG HY	AVG TS
HY1								
HY2	-.03							
HY3	.47**	.16						
TS1	.63**	-.16	.33*					
TS2	.61**	.14	.47**	.54**				
TS3	.72**	.07	.60**	.60**	.60**			
AVG HY	.72**	.48**	.82**	.40**	.58**	.68**		
AVG TS	.74**	.03	.54**	.82**	.84**	.86**	.64**	
GPA	.06	-.20	-.04	.10	-.10	.03	-.05	.02

Abbreviations: HY=Hours Studied Yesterday; TS=Typical Hours Studied Over Term; T1=Time 1; T2=Time 2; T3=Time 3; AVG=Mean.

[a]N=49 for all correlations.
*p<.05 **p<.01

Later Investigations

The 1979 study just reported was carried out on a special sample, but in 1977, 1979, and thereafter our regular surveys of LS&A College students routinely asked one or two summary questions on "typical study" and a question on GPA. The questions on study were never as detailed as in the previous years and the measure of GPA always relied on self-report, but separation of weekend study from weekday study is possible for most of these years and turns out to be useful. (The separation was not due to recognition of the interpretive importance of the distinction, but was made simply because student activities differ on the two types of days and it seemed easier to recognize this difference in interviewing and in standardized comparisons or scale construction.) Table 4 presents correlations for all of the available years from 1977 to 1984; examination of scattergrams shows the relations to be sufficiently linear so that product–moment coefficients are adequate summary measures of association.

Taken one at a time, as they occurred, the results from these more superficial survey inquiries did not seem to change appreciably the conclusions reached on the basis of the earlier, more thorough investigations, especially since our focus continued to be on *total* hours studied. But when the later findings are separated by time of the week and laid out in sequence in Table 4, an unexpected difference in associations emerges. First, we shall partly disregard the relatively large correlation in 1984 for weekday study (r = .18), since it is out of line with all other weekday associations in Table 4, as well as with those from the more thorough

Table 4. PRODUCT–MOMENT CORRELATIONS OF HOURS STUDIED AND GRADE POINT AVERAGE FOR WEEKDAYS AND WEEKENDS: 1977–84[a]

	Weekday Study			Weekend Study		
Year	r	N	S.D.	r	N	S.D.
1977	.09	300	2.14	.24*	56	2.32
1979	-.04	256	2.10	.12	61	2.33
1980	-.01	370	1.72	--	--	--
1981	.07	394	1.93	--	--	--
1982	.06	423	1.95	.09*	425	2.07
1983	.00	345	1.86	.16**	347	1.92
1984	.18**	298	1.64	.16**	297	1.64

[a] In 1977 and 1979, students were first asked about hours studied "yesterday" and then about how typical that time was for other weekdays/weekends earlier in the term; hence each N represents only those students interviewed following a weekday or a weekend. In 1980 and 1981 only typical weekday study was asked about. In 1982, 1983, and 1984, all students were asked about both weekdays and weekends. (No survey was carried out in 1978.) S.D.s are for hours studied on weekdays and weekends, respectively.

investigations reported earlier. (A difference in questioning in 1984 does not explain this outlier, and there is no sign of a trend over time that makes the result meaningful.) Instead, we note simply that five of the seven weekday correlations are positive, the others are essentially zero, and the mean is .06. It seems probable that there is a very tiny positive correlation between weekday hours studied and GPA, but it is so close to zero as to be regarded as trivial. The 1984 result is best seen as a chance deviation from this small mean.

However, the correlations involving weekend study are generally larger, more significant, and more consistent. The average is still small (.15, or .14 if *N*'s are used as weights), but unless there is some kind of correlated error due to self-report, a modest but meaningful relation must exist between weekend study and GPA. On the basis of the pattern in Table 4, we reexamined the 1973 data where GPA was obtained from records and found that the separate correlation (*r*) for weekend study was .12, just a little below the corresponding mean in Table 4; the correlation for GPA and weekday study was .06, exactly the same as the correspond-

ing Table 4 mean. (The 1975 survey did not obtain weekend hours, and the 1979 panel sample does not permit a distinction to be made. Only the 1975 chemistry class analysis fails to produce creditable evidence that weekend study is more highly correlated with grades than is weekday study.)

Some of the difference between weekday and weekend correlations can be explained by the greater variation in reported hours studied for weekends (compare the standard deviations in Table 4). This seems to be only a small part of the story, however, for the differences in variances are too slight to alter the correlations by more than a tiny amount. Apparently there is something about extensive studying over the weekend period—defined by most students as heavily social—that is associated with higher grades in a way that extensive studying during the week is not. In a larger sense, these findings do fit our earlier conclusion that it is only unusual amounts of studying that seem to relate to higher grades.

Although the findings reported in this section add to our confidence that hours of study do make some difference in grades, it is important to bear in mind that the associations reported in this section are very small not only in absolute terms ($r = .15$ being our highest estimate) but in comparison with the correlations reported earlier for class attendance, SAT scores, and even ethnicity. If effort in the sense of study does raise grades, these data indicate that it does so to only a very slight extent.

Discussion

A number of investigations were carried out to assess the relation between study and grades among students in a large liberal arts college of a major university. An association between hours of study and grades can sometimes be uncovered, but it appears to be very small and largely limited to those who study well above average in terms of amount or period of the week. Controls for SAT scores, the most likely confounding variable, do not significantly affect the study–GPA relation, and most other attempts to improve measurement and analysis also bore little fruit.

Results that are negative or much slighter than expected always pose a special problem for investigators, since they seem more likely than positive results to be attributable to inadequacies in measurement or other aspects of the research process. Our major concern throughout these investigations has been with the possible invalidity of student reports of their own studying. Students may not know how much they study, and there may also be some bias in willingness to report honestly. The problem is an inherently difficult one, for studying is not always directly observable (e.g., a student staring at a book may well be daydreaming), and in addition attempts at direct observation or detailed recording (as through diaries) might distort the phenomenon itself. Our own reliance

on various forms of self-report has its difficulties as well, but the total pattern of results from these investigations provides a fair degree of confidence that our measures, especially those in which an index of "typical study" was created, provide a reasonable approximation to a valid ordering of the actual studying that students do.

A number of listeners to these results have found them so inconsistent with their own intuition or experience that they have simply assumed that some important aspect of studying has been omitted in our research. While it is always possible that this is the case, a variety of additional inquiries not reported in detail have failed to find evidence for other study variables suggested by listeners. For example, whether students report studying in groups or alone bears no consistent relation to grades, although there is a small trend for solitary study to be connected with higher grades. Self-reported cramming, degree of note-taking, review of past exams, and going over readings twice are some of the other variables we have examined, but without finding either a direct relation to grades or an interactive effect with study and grades. We doubt that other similar kinds of variables will have a substantial effect on the basic study–grade relationship examined in this paper.

We are left with three possible lines of explanation for the study–grade findings beyond the possibility that the nonrandom component in grades is largely a reflection of SAT and similar aptitudes. (There is doubtless a component of grades that is best conceptualized as random, but GPA is too highly correlated with other variables to allow this as a sufficient explanation for its low correlation with study.)

1. There may be a substantial relation between study and grades, but study behavior may be so unstable that only a much larger number of individual study day observations will capture it. However, Epstein's (a, b) data suggest that three time points should ordinarily begin to show more promising results than appeared in our 1979 investigation. In that investigation, only the "typical study" measure gave even a sign of predicting grades, not the addition of three time points.

2. Hours Studied may have little effect on grades for the simple reason that most instructors may grade largely on the basis of the material they provide or at least emphasize in lectures. This is the simple and in a sense naive way to interpret our results, and it fits nicely the positive associations discovered between class attendance and grades. More generally, it is worth noting that there was one factor that students in our surveys cited as being as useful as study as a determinant of grades: "figuring our what a professor really grades on." It may be that this kind of practical wisdom about *what* to study is highly important and is both correlated with class attendance and uncorrelated with quantity of study.

3. The separation of "aptitude" and "effort" may be too artificial or too incomplete. Perhaps such factors as ability to concentrate or personal skill in organizing material or sheer memory affect the outcome of study in large and complex ways. Strictly speaking, such personality characteristics do not fall within the category of "effort," yet they are not really measured by standard college aptitude tests either. And of course grading also presumably reflects specialized aptitudes that fit particular subjects, for example, spatial ability for organic chemistry or an ear for language.

Without doubting either the reality or the desirability of such sources of grade ordering, we would still be left with our initial question: Is the formal reward system known as grading influenced to an important degree by effort and industry? And if not, what about other reward systems outside of the university setting, where universalistic standards are much less the ideal?

Notes

1. We do not know of any comprehensive review of research on the relation of amount of effort to either overall GPA or individual course grades. Most of the specific studies we have located were done long ago and yield contradictory conclusions: both May and Ryans report positive relations; Jones and Ruch report a negative relation; each of Crawford, Williamson, and Strang report no relation. More recent research on predicting grades (e.g., Brown and Holzman) tends to combine studying in an index with other nonstudy variables, which makes it difficult to know how much studying as such is an effective predictor. There is also a widespread and indeed unquestioned assumption that the success of high school grades reflects a large component of effort, as for example in a statement by Lerner that "Our closest current approximation to [effort] seems to be grade-point average. . . ."[!] (134).
2. We obtained GPA for the term in which the survey was carried out and also cumulatively through that term. The current term is used in most of our analysis, since Hours Studied and Class Attendance are for that term. In fact, results are similar for the two GPA dependent variables, and the two GPA measures themselves are highly correlated ($r = .85$). We should note that at the time of this first study, GPA at Michigan was based on a five-point scale (A, B, C, D, E) without plus and minus qualifiers, but the 1979 panel study (reported below) was done after the grading system was amended in Fall, 1975, to include plus and minus grades. (For various reasons, GPA records could not be obtained for 22 students for 1973.)
3. Weekend study did somewhat better than weekday study ($r = .12$ and .06, respectively) when the two were used separately, but at the time the importance of the distinction was not recognized. The difference is discussed further in a later section.
4. Verbal SAT and Math SAT scores were simply added together to give the total score. Each test can range from 200 to 800, and TSAT from 400 to 1600. In comparing the effect of SAT scores with other variables in our survey, it is necessary to keep in mind that SAT scores have quite high reliability, while the other measures are of more uncertain but doubtless lower reliability.
5. In our fourth investigation (reported below) we found that students who report sitting in the first few rows in a class have a significantly higher mean GPA than other students, though again causal direction is uncertain.
6. In comparing SAT scores and Hours Studied, we must keep in mind that the former is based on university records and the latter on self-report. But we also obtained student re-

ports of their SAT scores, at a point when neither they nor we knew that access to university records would be possible. A comparison of self-reported and actual SAT scores shows that 148 students (35 percent of the total 424) reported their scores quite accurately (within 50 points of recorded scores on both Verbal and Math SAT's), while the rest of the students were to varying degrees less accurate—the error almost always involving an upward bias in one or both self-reported scores. If the Hours Studied–GPA relation is being lowered because of misreporting of hours studied, it might plausibly be attributed to some of these same inaccurate students—whether such inaccuracy is due to carelessness, desire to impress interviewers, or other individual traits. (When errors on VSAT and MSAT scores are coded in terms of inflation, they are correlated $r = .41$.) If we separate out the 148 respondents who are accurate within the 50-point categories on both VSAT and MSAT scores, we have 148 individuals who might be expected to be somewhat more accurate *also* in reporting Hours Studied. For these individuals, the relation of GPA to Hours Studied was again computed, but its magnitude remains exactly what it was in Figure 2 ($E^2 = 3.1\%$) and its shape is no more clearly monotonic. Thus, if the assumption is correct that we have reduced reporting error in Hours Studied, then one must conclude that this does not help in predicting GPA.

7. There is one other piece of puzzling evidence in our data that should be noted. In our main 1973 survey, we found that Jewish students (identified in terms of parental religious background) had a GPA 0.42 points higher than non-Jewish students ($F = 21.3$, $df = 1, 34$, $p < .001$; $E^2 = 5.8\%$). Such a religious difference has been noted previously (Lavin) and also occurs in our later replications, but what is interesting here is our complete inability to explain the difference in terms of *either* ability or effort variables. Jewish students average slightly higher on both VSAT (11 points) and MSAT (17 points), but the differences are nonsignificant and much too small in terms of explained variation ($E^2 = 1.0\%$ for TSAT) to account for the mean GPA difference by religion. At the same time, there is no difference at all in Hours Studied by religion, and Jewish students actually report less class attendance than non-Jews ($p < .04$). Since religious category by itself cannot be a proximate cause of GPA, some mediating behavioral variable must either be missing or inadequately measured. (A further finding involving religion is that the relation of GPA to Class Attendance tends to be stronger ($p < .10$) among non-Jewish students (.35) than among Jewish students (.13), though the relation of GPA to Hours Studied is equally low and nonmonotonic for both religious categories.)

8. The questions on studying were essentially the same as used in the previous investigation and shown in the Appendix, except that "for this course" was added at an appropriate point in each question (". . . study yesterday for this course?"). Grades were obtained from the instructor at the end of the course. Numerical grading was used by the instructor and in Table 2; it was transformed into letter grades by the instructor at the end of the course.

9. The administration took only ten minutes at the beginning of the hour and we did not obtain a count of the total number of students present. Our impression was that most students there at the beginning of the hour did return questionnaires, and that nonreturnees either arrived late or were not at the class that morning. It seems likely that the latter students had a lower than average class attendance, and this fits both our results with the class attendance variable and the fact that our questionnaire sample had a slightly higher mean final course grade (low B−) than the total class (mean of C).

10. In the 1975 study objective records of GPA were not available and we were forced to depend on student self-reports of GPA. Given the complete absence of any trend in the Hours Studied–GPA relationship, this seems unlikely to be the source of the null association. Moreover, as noted below in Investigation IV, student reports of GPA do not seem to be biased in the same way as reports of SAT scores.

11. The full report of this investigation appears in an undergraduate Honors Thesis by Camille Olson, *GPA and Hours Studied: Is There a Positive Relation?*, University of Michigan, 1980. All the interviewing for Investigation IV was also done by Olson, but without knowledge of grades until all study data were obtained.

Appendix. Measurement of Hours Studied (Investigation I)

Original Questions

Q25. The next few questions are about studying. First, apart from time spent in class, how much time, if any, did you study *yesterday*? By studying we mean reading or any other assignment, writing, or review done outside of class. _____

Q26. Was that amount of time typical of the time you spent studying on *weekdays* during the past week or so?

Q26A (If No) What *was* the typical amount of time you spent studying on *weekdays* over the past week or so? _____

Q27. If you think of weekends as running from Friday afternoon through Sunday evening, how much time, if any, did you spend studying *last* weekend? _____

Q28. Was that typical of most weekends this term?

Q28A (If No) What *was* the typical amount of time you spent studying this term? _____

Hours Studied Index*

N	
(11)	0. Missing data
(6)	1. 0 to 0.9 hours per day
(42)	2. 1.0 to 1.9 hours per day
(72)	3. 2.0 to 2.9 hours per day
(102)	4. 3.0 to 3.9 hours per day
(91)	5. 4.0 to 4.9 hours per day
(55)	6. 5.0 to 5.9 hours per day
(45)	7. 6.0 or more hours per day
424	

*This index was constructed by summing the appropriate combination of Q's 25, 26, and 26A, weighted by 4, and the appropriate combination of Q's 27, 28, and 28A, first divided by three to obtain a per day figure for other analyses, then weighted by 3 for this index. The total sum was then divided by 7 to obtain a per day figure for each respondent.

References

Brown, W. F., and W. H. Holtzman. 1966. *Survey of Study Habits and Attitudes*. Psychological Corporation.

Collins, Randall. 1979. *The Credential Society*. Academic Press.

Crawford, A. B. 1929. *Incentives to Study*. Yale University Press.

Epstein, S. a:1979. "The Stability of Behavior: I. On Predicting Most of the People Much of the Time." *Journal of Personality and Social Psychology* 37:1097–1126.

_____. b:1980. "The Stability of Behavior: II. Implications for Psychological Research." *American Psychologist* 35:790–806.

Featherman, D. L., and R. M. Hauser. 1976. "Sexual Inequalities and Socioeconomic Achievement in the U.S., 1962–73." *American Sociological Review* 41:462–83.

Fishbein, M., and I. Ajzen. 1974. "Attitudes Toward Objects as Predictors of Single and Multiple Behavioral Criteria." *Psychological Review* 81:649–53.
Heider, F. 1958. *The Psychology of Interpersonal Relations*. Wiley.
Jencks, C., and J. Crouse. 1982. "Aptitude vs. Achievement: Should We Replace the SAT?" *The Public Interest* 67:21–35.
Jones, L., and G. M. Ruch. 1928. "Achievement as Affected by Amount of Time Spent in Study." *27th Yearbook, National Society for the Study of Education*, Part 2.
Lavin, David. 1965. *The Prediction of Academic Performance*. Russell Sage.
Lerner, B. 1980. "The War on Testing: David, Goliath, and Gallup." *The Public Interest* 60:119–47.
May, M. A. 1923. "Predicting Academic Success." *Journal of Educational Psychology* 14:429–40.
Ryans, D. G. 1939. "Some Observations Concerning the Relationship of Time Spent at Study to Scholarship and Other Factors." *Journal of Educational Psychology* 30:372–76.
Rytina, J. H., W. H. Form, and J. Pease. 1970. "Income and Stratification Ideology: Beliefs about the American Occupational Structure." *American Journal of Sociology* 75:703–16.
Sexton, P. C. 1961. *Education and Income*. Viking Press.
Siegel, P. M. 1965. "On the Cost of Being a Negro." *Sociological Inquiry* XXXV:41–57.
Strang, R. 1937. *Behavior and Background of Students in College and Secondary School*. Harper & Row.
Williamson, E. G. 1935. "The Relationship of Number of Hours of Study to Scholarship." *Journal of Educational Psychology* 26:682–88.