Cornell University
ILR School

Cornell University ILR School
DigitalCommons@ILR

Cornell Higher Education Research Institute (CHERI)

Centers, Institutes, Programs

11-1-2002

The Effects of Class Size on Student Achievement in Higher Education: Applying an Earnings Function

Michael Dillon
SUNY Binghamton

E. C. Kokkelenberg
SUNY Binghamton

Sean M. Christy
SUNY Binghamton

Follow this and additional works at: http://digitalcommons.ilr.cornell.edu/cheri

This Article is brought to you for free and open access by the Centers, Institutes, Programs at DigitalCommons@ILR. It has been accepted for inclusion in Cornell Higher Education Research Institute (CHERI) by an authorized administrator of DigitalCommons@ILR. For more information, please contact jdd10@cornell.edu.

The Effects of Class Size on Student Achievement in Higher Education: Applying an Earnings Function

Abstract
This paper uses an earnings function to model how class size affects the grade students earn. We test the model using an ordinal logit with and without fixed effects on 363,023 undergraduate observations. We find that class size negatively affects grades. Average grade point declines as class size increases, precipitously up to class sizes of ten, and more gradually but monotonically through class sizes of 400 plus. The probability of getting a B plus or better declines from 0.9 for class sizes 20 to about 0.5 for class sizes of 120 and almost 0.4 for class sizes of 400.

Keywords
class size, undergraduate students, grades, higher education

Comments

Suggested Citation
Dillon, M., Kokkelenberg, E. C., and Christy, S. M. (2002) *The effects of class size on student achievement in higher education: applying an earnings function* (CHERI Working Paper #28). Retrieved [insert date], from Cornell University, ILR School site: http://digitalcommons.ilr.cornell.edu/workingpapers/15/

Required Publisher Statement
Published by the Cornell Higher Education Research Institute, Cornell University.

This article is available at DigitalCommons@ILR: http://digitalcommons.ilr.cornell.edu/cheri/15

DO NOT QUOTE

The Effects of Class Size on Student Achievement in Higher Education: Applying an Earnings Function[a]

Michael Dillon*
E. C. Kokkelenberg**
Sean M. Christy***

P.O. Box 6000
Binghamton University
State University of New York

November 2002

* Senior Analyst
Office of Institutional Research

** Professor of Economics
Department of Economics
607-777-2550 (phone)
kokkele@binghamton.edu
607-777-2681 (fax)

***Department of Economics

[a] This study is an extension of an analysis started by Jack Keil and Peter Partell of Binghamton University, Office of Institutional Research in the late 1990s. We wish to thank them for their help and insights. We would also like to thank Jessica Richards who has helped enormously in editing, critiquing and creating the tables and charts used in the paper. Yu Zhu and Hester Han have also brought their considerable analytic expertise to the project. All errors and omissions are the responsibility of the authors.

ABSTRACT

This paper uses an earnings function to model how class size affects the grade students earn. We test the model using an ordinal logit with and without fixed effects on 363,023 undergraduate observations. We find that class size negatively affects grades. Average grade point declines as class size increases, precipitously up to class sizes of ten, and more gradually but monotonically through class sizes of 400 plus. The probability of getting a B plus or better declines from 0.9 for class sizes 20 to about 0.5 for class sizes of 120 and almost 0.4 for class sizes of 400.

INTRODUCTION

We estimate the influence of class size on student achievement in higher education. The K-12 literature on class size suggests class size negatively influences student outcomes (at least under certain circumstances). In the case of higher education, the evidence is more mixed.

We present a parsimonious model of grades employing wage theory as a framework and test this model using data from a medium-sized public research university. Applying a logistic regression with and without a fixed effects model we find that class size is a very important variable in predicting grades and that the functional form of the relationship is consistent with the theoretical model developed by Glass et al. (1982) to explain the negative effect of class grades on K-12 student performance. We also explore additional models, various proxies for a key variable (student ability) and how the effect of class size on grades differs for advance placement, at-risk, underrepresented and female undergraduates.

BACKGROUND

K-12 studies.

By the 1970's there was near consensus in the educational research community that class size had little effect on student achievement.[1] However, Glass and Smith, in a series of articles beginning in the late 1970s (Glass and Smith, 1979; Smith and Glass, 1980; Glass, McGraw and Smith, 1981) presented a theoretical model suggesting that the functional form of the relationship between class size and student achievement should be negatively sloped and concave.[2] This model has become a basis for further normative discussion on whether, or how, class sizes should vary[3]. Glass and Smith also presented the results of their own meta-analysis

of studies looking at the effect of class size sustaining the negative logarithmic relationship between class size and student performance.[4] Given this apparently beneficial evidence of smaller class sizes, several states designed experiments to replicate Glass's et al. findings.[5]

Even though there is now clear evidence that smaller class sizes improve student performance, at least in some circumstances, the debate continues over what to do with that evidence. In particular, economists point out the need to weigh the costs of achieving smaller classes versus the costs of improving student achievement by other means. (Nelson and Hevert, 1992; Maxwell and Lopus, 1995).[6]

Class Size at the College Level

Though there is debate about the extent of benefits small classes bring, or how much it costs to achieve, there is at least some agreement in the K-12 literature that class size matters in certain circumstances. No such agreement exists in the literature concerning the effect of class size in higher education. Indeed, in two well respected reviews of the literature (William et al., 1985; Pascarella and Terenzini, 1991), the authors conclude that the overall evidence suggests that class size plays no or little influence on student achievement. This however has not quelled the debate. McKeachie (1980) and McKeachie et. al, (1990) have presented arguments that class size is the primary environmental variable college faculty must contend with when developing effective teaching strategies. They argue that while class size may not be significant in courses best suited for lecture style learning, courses geared toward promoting critical thinking and advanced problem solving are best taught in a smaller classroom environment

McKeachie's view is consistent with findings that suggest that students' (and professors') motivation and attitude toward learning tends to be more negatively affected by larger classes. (Feldman, 1984; Bolander, 1973; McConnell and Sosin, 1984; Spahn, 1999) Though they may have learned the material, students do not feel as satisfied with the classroom experience as they would have in smaller classes, suggesting that some learning opportunities may have been lost. Also, there is some evidence that class size may matter in some courses but not in others. Raimondo et. al (1990) found that students in smaller sized introductory macroeconomic courses did better in subsequent intermediate macroeconomic courses even though the same was not true when conducting the analysis for microeconomic courses. They suggest, consistent with McKeachie argument, that smaller classroom environments enhance the more wide-ranging, non-formula based knowledge necessary for understanding macroeconomic principles.

There is also a debate about how to measure student outcomes at the university level. In the K-12 studies, pre and post testing is ubiquitous: the change in student performance relative to the improvement found in students not subjected to the whatever the variation in teaching method or classroom that is under study, is attributed to the changed element. In essence, investigators have both a control group, and a tested, agreed upon metric. We lack control groups and an agreed upon metric in most relevant studies focusing on higher education. Hence, the increased student performance in higher education can be measured by a variety of metrics: grade in the class under study or a subsequent course, performance on a graduate admissions exam, graduation or retention rates, percentage going on to graduate or professional work, self reported "satisfaction" with a course, or even salary or wealth at some time post graduation. There are

numerous problems associated with measurement of many of these and as one moves further away through time from the course under study, many extraneous factors cloud the data. Finally, much of the K-12 testing is done for specific academic subjects, such as chemistry or reading comprehension. There is no comparable single set of before and after test scores that is applicable across academic subjects in higher education.

Given that there is a lack of consensus about how to measure student achievement in higher education it is not surprising that there is no definitive answer to the question of how class size relates to it. Nor do we attempt to solve the debate in this paper. We do present findings, based on data from a single institution, of how class size effects student outcomes, as measured by grades, after controlling for other relevant student and course characteristics. In doing so we rely on the theory of wages as a way to think about the nature of grades from a student's perspective.

THE MODEL

Labor theory (Mincer, 1974) suggests that earnings or wages depend upon ability, education, and experience. Applying this to the market for higher education, we postulate the following story: Students attend institutions of higher education to gain experience and education. They pay for this education through tuition, fees, living expenses, living conditions, and foregone wages. They are rewarded with some sort of certification at the end of some period of study. During this time, they are paid by a form of script, that is, credit hours and individual grades, which when amassed, indicate the extent and quality of their performance in school. When accumulated sufficiently, the script can be used to “buy” a certificate or degree. The quality of the script, and indeed its acceptability in buying a degree, is represented by the course grade. Since there often

are grade point standards, course grades have a further screening importance.

We can consider a course grade then, as a form of reward or payment denoting the quality of the script for the performance the student achieved in a specific course. We define the relationship between script (proxied by course credit hours, *H*), and its quality (proxied by course grade, *G*) as *W*, the wage as

(1) $W = f(G, H)$.

We assume the specific functional form is:

(1a) $W = \alpha G + (1 - \alpha)H, \quad 0 < \alpha < 1$

Tests of our data suggest that *H* is virtually constant as the vast preponderance of courses have the same number of credit hours. Hence we set *W* = *G* in what follows. Using labor wage as its analogy, we hypothesize that a student's wage (grade) can be explained by her ability and experience, controlling for individual-specific and environmental characteristics. We thus write for the *i th* students in the *j th* class during period *t*:

(2) $W_{ijt} = b_0 + \phi(E_{it})'\beta + \theta(A_{it})'\Gamma + Z'_{it}\lambda + V'_{jt}\kappa$

Here, *W* represents the wage, or, in this case, the grade, *E* the student's experience (level in college, 1st semester freshmen through 2nd semester senior), *A* represents ability, *Z* a vector of dichotomous student related variables, and *V* is a vector of environmental factors including class size (CS). N(E) and 2(A) are polynomials in E and A, and ∃, ɜ, 8, and 6 are vectors of parameters to be estimated and b_0 a scalar, also to be estimated.

The null hypothesis is that class size does not affect student learning or performance at the

university level and this would be reflected in the stability of grade distributions over various class sizes for various subjects.

DATA

This study was conducted using data from a highly selective research institution (new Carnegie classification) located in a small city in the Northeast. There is one observation per student per course for each semester analyzed. The principle population sampled is all undergraduate students for the period Fall 1996 through Fall 2001. The undergraduate population is in five schools; Arts and Sciences, Education and Human Development, Engineering, Nursing, and Management. The dependent variable is the grade a student receives in a course. Only grades that count toward a student's GPA are considered; thus incompletes and withdrawals are dropped from the data. The variables and data are discussed further in an Appendix.

MODEL ESTIMATION

The model represented by Equation (2) was estimated via the logistic procedure in SAS, version 8.0. Initially, the model was developed using one fifth of the data. Box Plots and statistical analysis alerted us to several data problems which were addressed. A full specification of Equation (2) including a large number of proxies for the environment variables and polynomials in experience, ability and class size, was then estimated. We also included a number of demographic variables such as race, talent level, degree seeking, and county of residence in New York State. The environmental variables also included faculty rank, a variable for majors(s) and various other academic variables.

The model was then simplified using both the forward and backward routines in SAS and the

probability of the chi-square statistic for individual parameters being equal to zero (in a manner analogous to the use of the t statistic in OLS (see Kmenta, 1986)) to eliminate and strive for one proxy for ability, one for experience, several for the environment and several for the demographics. A simplified model with a limited number of observations - limited by deleting the top and bottom class sizes, was next tested on a second subset of the data. After this, three variants of the model given by Equation (2) were estimated using 363,023 observations, and a slightly smaller dataset as explained below.

RESULTS

The Base Model

We begin by presenting the results of a fairly parsimonious model of grades, including the four variables discussed above (experience, education, ability and class size) and four additional variables, the mean grade given out by the department over the course of nine semesters, gender, students whose ethnicity is underrepresented in higher education, and students from the Educational Opportunity Program (EOP). Initially we suspected the performance of underrepresented, EOP and female students may be more sensitive to class size.

Models 1 and 3 are given as:

$$(3) \qquad W = \exists_0 + \exists_1 E + \exists_2 E^2 + \exists_3 E^3 + \exists_4 A + \exists_5 A^2 + \exists_6 A^3 + \exists_7 CS + \exists_8 CS^2 + \exists_9 CS^3 + \exists_{10} AP + \exists_{11} D + \exists_{12} G + \exists_{13} M + \exists_{14} EOP$$

Here, E denotes experience, A ability, CS class size, AP the existence of advanced placement credits, D a historical departmental mean grade, G denotes gender, M minority, and EOP denotes an educational opportunity student. Models 1 and 3 differ in that model 1 estimates equation (3)

for all observations and model 3 for class sizes greater than five students.

Model 2 is essentially equation (3) with four cross terms, AP* CS, G* CS, M* CS, and EOP* CS which provide estimates how these four factors vary over the range of class size.

Models 1, 2, and 3 explain the data well. The "G" statistic, a ratio of the likelihoods, is distributed chi square with 14, 18, and 11 degrees of freedom for models 1, 2, and 3 respectively. The critical values at $P_R = 0.005$ are 31.319 and 38.582, and our G values greatly exceed this (see Table 1). Note that we also estimated these four models stepwise and all the reported variables entered into the model contribute significantly to the likelihood value.[7] The c statistic's range is from 0.5 to 1.0 (0.5 or lower indicates that the model's predictions are no better than chance). Our regression results are 0.758, 0.759, and 0.754 (see Table 1) indicating a high discriminatory power of the model. Tau-a is a test of the null hypothesis that we have an improperly specified model. Calculated Tau-a values of under 0.05 indicate failure to reject the null hypothesis. In summary, the models explain the data very well indeed

Turning next to the individual parameters from the logistics regression we find that all eight independent variables, including the log of class size, have a statistically significant influence on grades (all the p values are less than .0001). Therefore, the null hypothesis that class size does not matter can be rejected. Further, experience and ability are positively related to grade, but at a decreasing rate (the square term is negative). Minorities and EOP students do less well, females and those with AP credit do better. The departmental mean grade has the most impact

on grades (having the largest numerical value) and any further work in this area should account for departmental grading culture and traditions. The chief result is that class size enters all estimations with a negative and large value (-0.912, -0.893, -3.180 for each of the three models (see Table 1)). This result is robust as to various other proxies for experience, ability, and department and for other classroom environment variables.

One could argue that the results are determined by the differing social structures in small versus large classes. Such an argument would claim that faculty are reluctant to give poor grades in small classes and only give low grades to more anonymous students in large classes. To test if this is what drives our results, we estimated our model for successively larger cut-offs at the low end of enrollment. We now turn to these results. Logistic regression results using a slightly larger data set (395,408 observations) for the period fall 1994 through spring 2000 show that the estimates of the ability, experience, course load, demographic and faculty parameters are quite stable over a wide range of enrolment cut-off levels (see Table 2). For example, the estimated value for ability changes from 0.786 to 0.902 for all class sizes, from class sized greater than one student to greater than 100 students. Similar to even more stable estimates are shown for the other parameters except for class size. Because the probabilities for high grades are very sensitive to class size, and the relationship is non linear, the estimates for the log enrolment are increasingly negative.

Adding interactive terms

Next we interact the log of enrollment with four dichotomous variables, AP credits, Female, Underrepresented Minorities and EOP. The expectation are that students entering with AP

credit, since they come in with higher educational achievement, may be less negatively affected by large classes. Females might be expected to be sensitive differentially from males. Underrepresented minorities and Educational Opportunity Program participants may be expected to do worse in large classes, since, as cited earlier, the literature suggests that small classes should be most beneficial to at-risk students.

The results of this estimation (Model 2, Table 1) sustain the view that the effect of class size on grades is negative. Students with AP credits do better than students without AP credit as class size increases. For females the effect is the opposite: In smaller classes they do worse than males (coefficient of –0.114). Underrepresented and at-risk students also perform poorly in larger classes (see Table 1, Model 2).

Next we show our chief results graphically. Figures 1 and 2 show average GPA by class size for total enrolment. The first deals with all classes, the second with classes sized greater than five. Again, the message is that large classes have a high probability of lower grades than smaller classes. Note that the probabilities fall precipitously for classes up to about 20 to 40 students and much more gradually thereafter. Thus, if grades are important, there is less of a decline in the probability of high grades when moving from classes of size 60 to 440 than for increasing class sizes from ten to twenty. The results of estimating Model 2 with different class sizes and interactive terms is shown in Figures 3 through 6. In all sized classes, students with AP credits do better than students without AP credit (Figure 3). For females there is a differential effect. In smaller classes they do better than males, but that difference disappears and indeed, reverses

(Figure 4) as class size increases. Underrepresented and at-risk students, like students without AP credit, do marginally worse than their counterparts in small classes, but they do significantly worse as class size increases (Figures 5 and 6).

The Fixed Effects Model

If one treats the data as a panel data set, where the individual student is the unit of observation (we have data on about 36,000 students) then a fixed effects model can be given as:

(4) $W_{itj} = \exists_0 + \exists_i + \exists_t + \exists_1 E_{it} + \exists_2 A_{it} + \exists_3 CS_j + \exists_4 D_E$

Here $\exists_i$ is the student fixed effect and $\exists_t$ the semester fixed effect. These two variables allow us to control for individual attributes not explicitly contained in the experience and ability variables (which probably evolve over time), and time fixed effects which control for grade inflation, if present. Initially, we estimate the model using the proportional odds assumption for ordinal logistic regression. That is, the marginal effects between an A minus and a B plus is the same as the marginal effects between any other pair, say B minus and C plus.

We estimate a polynomial variant of Equation (4) in both fixed effects and no fixed effects sub-variations. These are Models 1 and 2 of Table 3. In the first Model, the data was for 10,000 students covering 167,928 student grades. The data was differenced by subtracting the average grade the student received from the individual grade; hence, a fixed effects model. Model 2 in Table 3 is for the individual student-grades for the 10,000 students and is shown for comparison. The chief result is that class size again is strongly negative with coefficient values that are one order of magnitude larger than ability or experience. A test of the proportional odds assumption however fails with a p-value of less .0001.

Next, we relaxed the assumption of proportional odds and we estimated a binary fix effects model of equation (4) for a different random sample of 10,000 students chosen from the 167,928 observations. The results are reported in Table 4. Again, the model includes an experience variable, an ability variable to allow for time varying student ability, a departmental variable to accommodate cultural issues, and a class size variable. All fixed student characteristics are differenced out against the individual students' mean value. The binary logit estimates the probability at each grade level. For example, the probability of getting a B plus or better versus the probability of getting a B or lower. Note that the three runs bifurcating the probabilities at F versus D or better, D or lower versus C minus or better, and C minus or lower versus C or better did not converge and are thus not reported. We believe that this has to do with the smaller number of observations at those grade levels. Note that again, the log of class size has a negative coefficient, that the departmental mean grade has the largest impact on grades, and that better students improve with experience. Both of these results of fixed effects models (Tables 3 and 4) are consistent with and confirm the results from the ordinal logit estimation reported above.

DISCUSSION

Applying an earning function to the study of grades in higher education allows us to produce a parsimonious model relating environment, ability, and experience to undergraduate course grades. We use this model to show that class size has a negative relationship to grades and that the effect on class size on grades differs across different categories of student.

Though we have found a link between grades and class size, we cannot conclude that students

learn more in smaller classes; a better test would be to compare the same course with different sized sections. If we can determine that the sections were taught and evaluated in the same manner, we could judge, after controlling for student characteristics, whether students in smaller sections performed better. Alternatively, we could compare different sized sections in terms of how well their students performed in subsequent, more advanced, coursework in the same discipline. As long as students from differently sized courses take the same subsequent course, their grade in the subsequent course can be used to judge the effect of class size in preparing the students for future coursework.

The fact that students receive lower grades in larger classes is not itself a problem. Indeed, some faculty and administrators might suggest the results indicate the need for more large classes to offset grade inflation. Nor have we addressed the higher marginal costs of moving to smaller classes. If however large classes negatively affect persistence as well as grades, this would suggest a non-market cost could exist for relying on large classes both in terms of lost revenue due to the decrease student retention and the loss of reputation caused by lower graduation rates. Indeed, if we could quantify the indirect costs associated with loss of reputation, and the direct costs of losing tuition and other revenue because of lower retention rates, along with the cost saving of using larger classes to teach courses, we could estimate an optimal class size for the institution. Of course, it may be found that larger classes have no effect on retention. The evidence presented in this paper suggests class size mostly influences the likelihood of getting an A; the increase in the likelihood of failing rising only modestly as class size increases. So it is likely that if class size does greatly influence persistence, it will do so by promoting voluntary rather than non-voluntary dropping out. Consequently, future studies should look at the effect

class size has on both kinds of attrition.

TABLE 1

Estimated Coefficients via Maximum Likelihood-Logistics Procedure
(Values in Parenthesis are Standard Errors)

Variable/Statistic	Model 1	Model 2	Model 3
Experience	0.171 (0.021)	0.174 (0.021)	0.162 (0.021)
Experience Squared	-0.029 (0.005)	-0.029 (0.005)	-0.028 (0.005)
Experience Cubed	0.002 (0.0004)	0.002 (0.0004)	0.002 (0.0004)
Ability	0.785 (0.006)	0.784 (0.006)	0.787 (0.006)
Ability Squared	-0.079 (0.003)	-0.079 (0.003)	-0.078 (0.003)
Ability Cubed	0.005 (0.0002)	0.005 (0.0002)	0.005 (0.0002)
Class Size	-0.912 (0.044)	-0.893 (0.044)	-3.180 (0.135)
Class Size Squared	0.031 (0.013)*	0.041 (0.013)	0.603 (0.034)
Class Size Cubed	0.005 (0.001)	0.004 (0.001)	-0.041 (0.003)
AP Credit	0.300 (0.006)	0.029 (0.026)**	0.300 (0.006)
Department	2.210 (0.011)	2.197 (0.011)	2.230 (0.011)
Gender	0.136 (0.006)	0.610 (0.025)	0.134 (0.006)
Minority	-0.218 (0.011)	0.150 (0.043)	-0.217 (0.011)
EOP	-0.277 (0.014)	0.205 (0.057)	-0.276 (0.014)
AP x CS		0.065 (0.006)	
Gender x CS		-0.114 (0.006)	
Minority x CS		-0.092 (0.010)	
EOP x CS		-0.127 (0.014)	
N	363,023	363,023	354,454
Tau-a	0.439	0.440	0.433
C	0.7758	0.759	0.754
-2 Log Likelihood			
Intercept only	1,496,936	1,496,936	1,470,368
Full Model	1,309,876	1,309,020	1,293,052
Difference (G)	187,060	187,916	177,316

All chi square statistics <0.005 except as noted below.
* Chi Square = 0.0165; marginally significant
** Chi Square = 0.2502; not significant

TABLE 2

Estimated Coefficients via Maximum Likelihood-Logistics Procedures
Class Sizes (Number of Students)

	>1	>3	>5	>10	>25	>50	>100
Experience	0.194	0.193	0.193	0.190	0.200	0.168	0.095
Experience ^{2}	-0.035	-0.036	-0.035	-0.035	-0.036	-0.025	NS
Experience ^{3}	0.003	0.003	0.003	0.003	0.003	0.002	NS
AP Credits	0.304	0.304	0.304	0.304	0.320	0.361	0.391
Ability	0.786	0.787	0.788	0.797	0.842	0.894	0.902
Ability ^{2}	-0.080	-0.079	-0.079	-0.081	-0.089	-0.098	-0.094
Ability ^{3}	0.005	0.005	0.005	0.005	0.005	0.006	0.005
Class Size	-1.729	-2.681	-2.677	-2.319	-8.351	-7.853	NS
Class Size ^{2}	0.231	0.466	0.465	0.380	1.685	1.606	NS
Class Size ^{3}	-0.010	-0.029	-0.029	-0.022	-0.115	-0.011	NS
Department	2.222	2.232	2.242	2.275	2.437	2.444	2.521
Gender	0.140	0.140	0.139	0.135	0.106	0.025	NS
Minority	-0.218	-0.218	-0.217	-0.219	-0.231	-0.306	-0.375
EOP	-0.275	-0.274	-0.275	-0.276	-0.321	-0.433	-0.512
N	395,408	391,432	388,222	378,897	312,798	198,533	131,543
Tau-a	0.436	0.433	0.431	0.429	0.432	0.436	0.448
C	0.757	0.755	0.753	0.751	0.750	0.750	0.756

NS = not statistically significantly different from zero.

Note: a modified dataset with more observations was used for this test.

TABLE 3

Ordinal Logit Estimation of Data by Students: Dependent Variable is Grade
Estimated Coefficients via Maximum Likelihood-Logistics Procedure
(Values in Parenthesis are Standard Errors)

Variable/Statistic	Model 1: Fixed Effects		Model 2: No Fixed Effects	
Experience	0.184	(0.035)	0.329	(0.031)
Experience Squared	-0.024	(0.008)	-0.067	(0.008)
Experience Cubed	0.002*	(0.001)	0.005	(0.001)
Ability	0.183	(0.011)	0.871	(0.009)
Ability Squared	-0.018	(0.005)	-0.102	(0.004)
Ability Cubed	0.002	(0.001)	0.005	(0.0004)
Class Size	-2.195	(0.209)	-2.341	(0.190)
Class Size Squared	0.324	(0.052)	0.361	(0.048)
Class Size Cubed	-0.017	(0.004)	-0.019	(0.004)
Department	2.577	(0.021)	2.205	(0.017)
Proportion of fixed Effects significant at <.005 or better	0.673			
N	167,928		167,928	
-2 Log Likelihood				
Intercept only	625,261		625,261	
Full Model	512,487		547,065	

All Wald Chi square statistics <0.005 except as noted below.
* Chi Square = 0.0130; marginally significant

TABLE 4

Estimated Coefficients via Fixed Effects Binary Logit Model
(t - statistics in parentheses)

Variable/statistic	C or lower Versus C+ or better	C+ or lower Versus B- or better	B- or lower Versus B or better	B or lower Versus B+ or better	B+ or lower Versus A- or better	A- or lower versus A	Ordinal Non-fixed effects
Ability	0.176 (6.50)	0.184 (7.14)	0.149 (6.13)	0.169 (6.89)	0.153 (5.91)	0.142 (4.68)	0.620 (219.75)
Experience	NS	0.027 (1.29)	0.57 (3.09)	0.069 (4.08)	0.076 (4.48)	0.063 (3.33)	0.067 (32.25)
Log class size	-0.445 (-9.38)	-0.450 (-10.62)	-0.401 (-10.79)	-0.426 (-12.38)	-0.441 (-12.85)	-0.504 (-13.11)	-0.338 (-68.6)
Department	3.060 (17.59)	3.257 (20.22)	2.843 (20.09)	2.762 (21.08)	2.370 (18.63)	1.958 (14.12)	2.174 (127.84)
N	10,000	10,000	10,000	10,000	10,000	10,000	167,928
Percent $\exists_l$ significant at							
0.01	42	71	80	88	89	81	
0.05	48	71	80	88	89	81	
0.10	51	71	80	88	89	81	

NS = not statistically significantly different from zero.

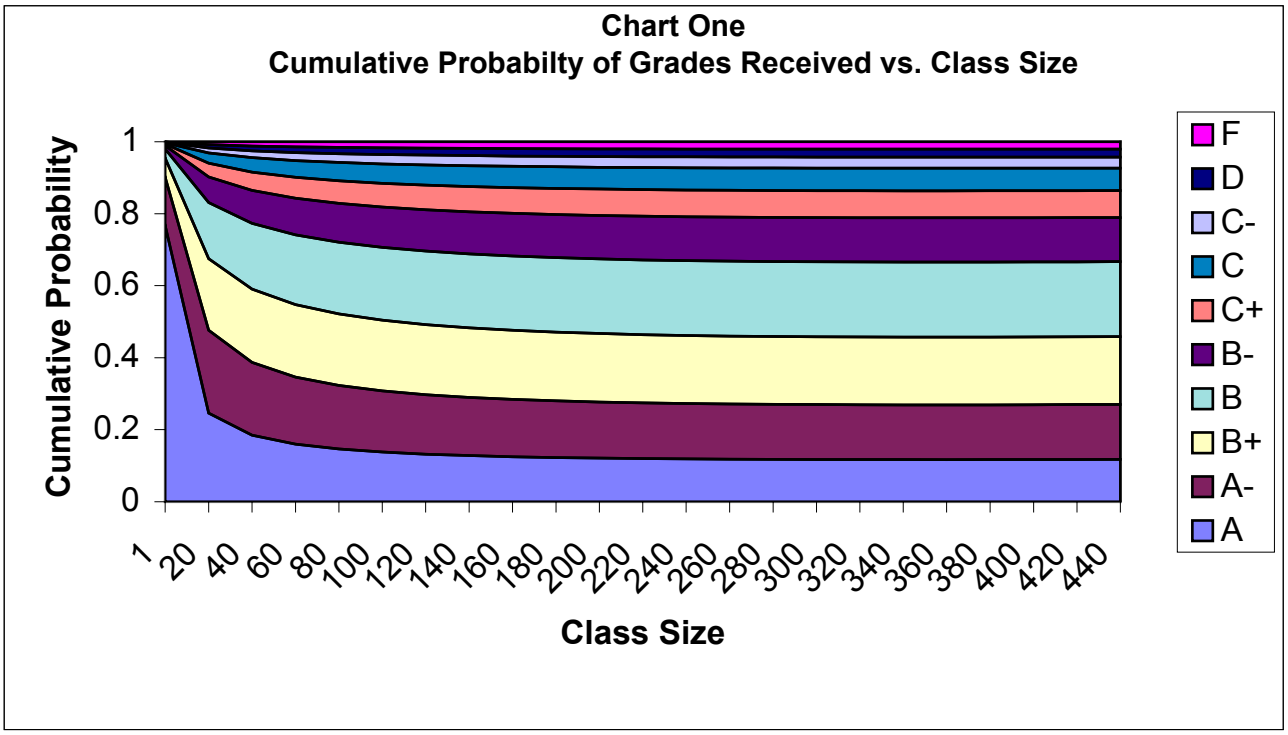
Chart One
Cumulative Probabilty of Grades Received vs. Class Size
Cumulative Probability
1
0.8
0.6
0.4
0.2
0
1 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380 400 420 440
Class Size
F
D
C-
C
C+
B-
B
B+
A-
A

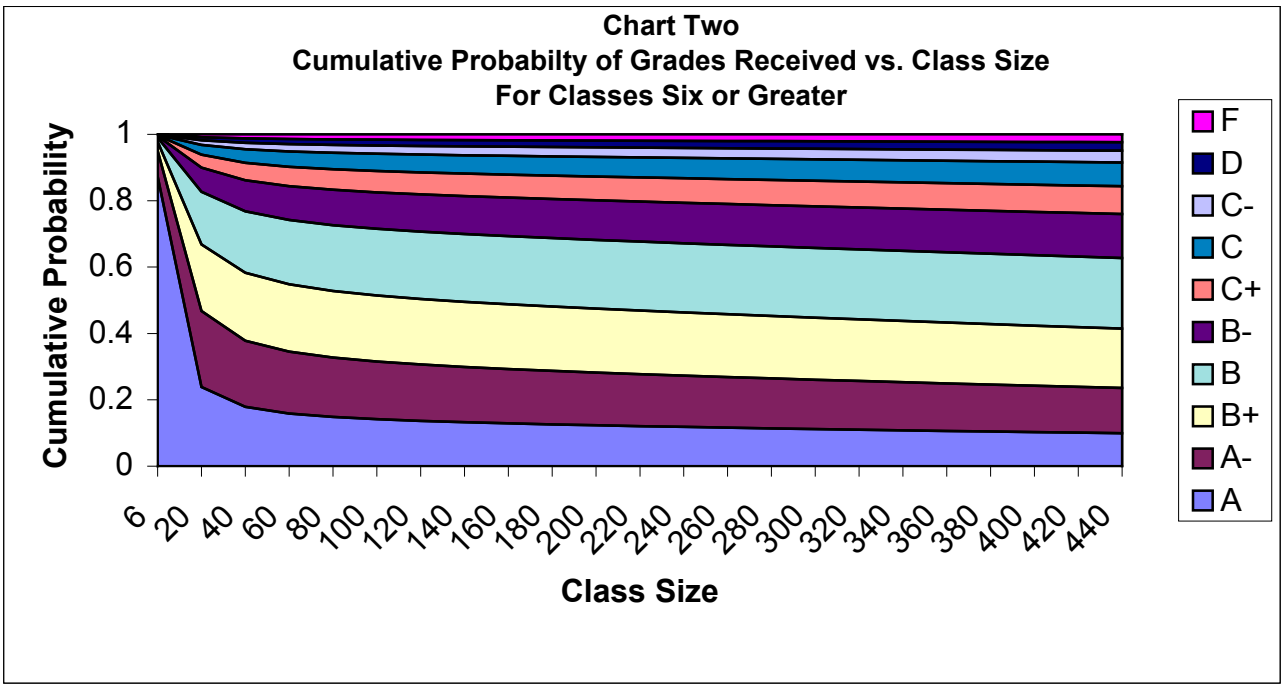
Chart Two
Cumulative Probabilty of Grades Received vs. Class Size
For Classes Six or Greater
Cumulative Probability
1
0.8
0.6
0.4
0.2
0
6 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380 400 420 440
Class Size
F
D
C-
C
C+
B-
B
B+
A-
A

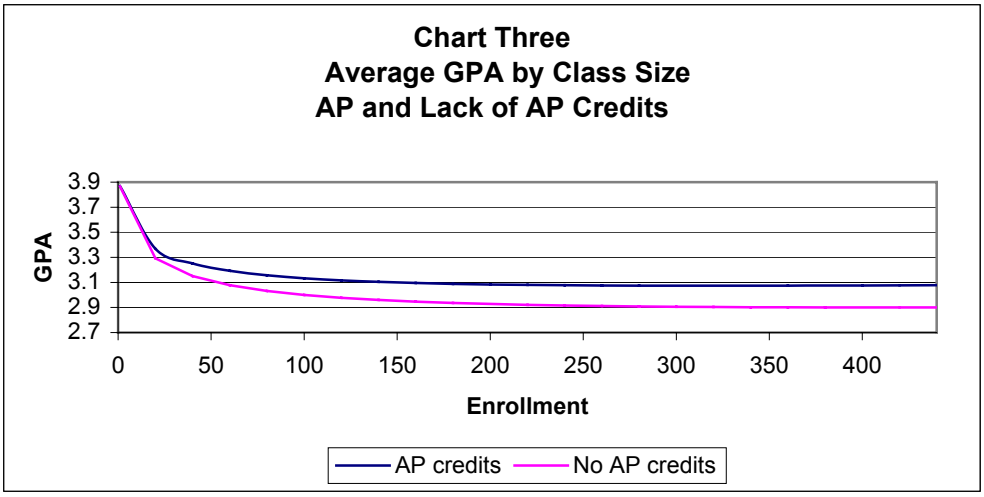
Chart Three
Average GPA by Class Size
AP and Lack of AP Credits
GPA
3.9
3.7
3.5
3.3
3.1
2.9
2.7
0
50
100
150
200
250
300
350
400
Enrollment
AP credits
No AP credits

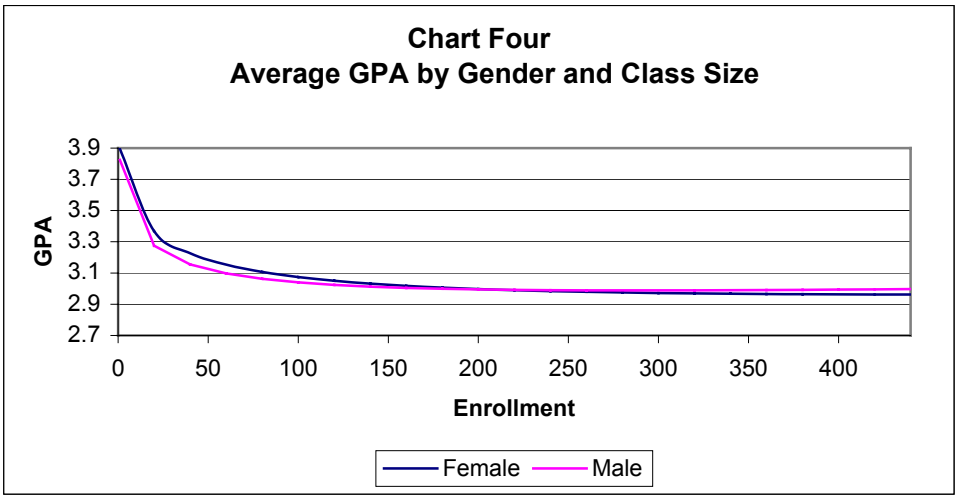
Chart Four
Average GPA by Gender and Class Size
GPA
3.9
3.7
3.5
3.3
3.1
2.9
2.7
0
50
100
150
200
250
300
350
400
Enrollment
Female
Male

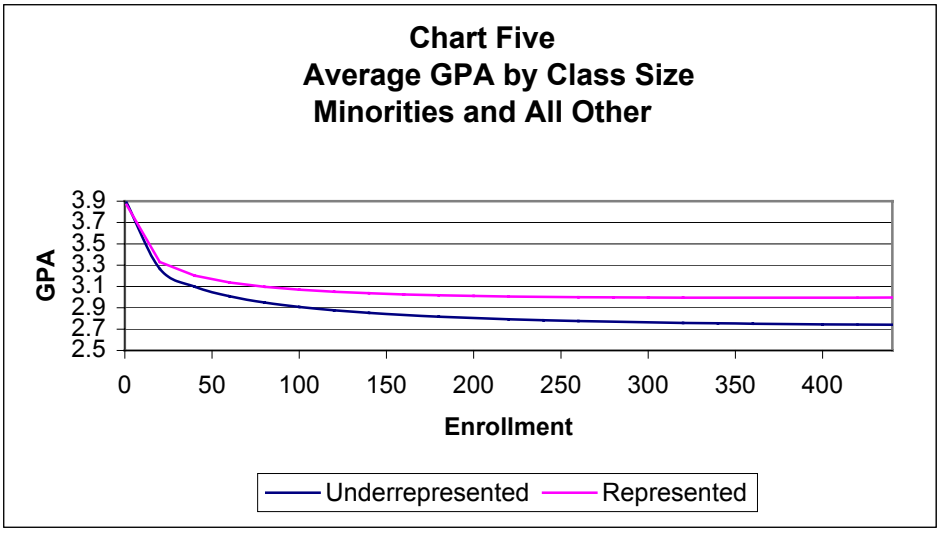
Chart Five
Average GPA by Class Size
Minorities and All Other
GPA
3.9
3.7
3.5
3.3
3.1
2.9
2.7
2.5
0
50
100
150
200
250
300
350
400
Enrollment
Underrepresented
Represented

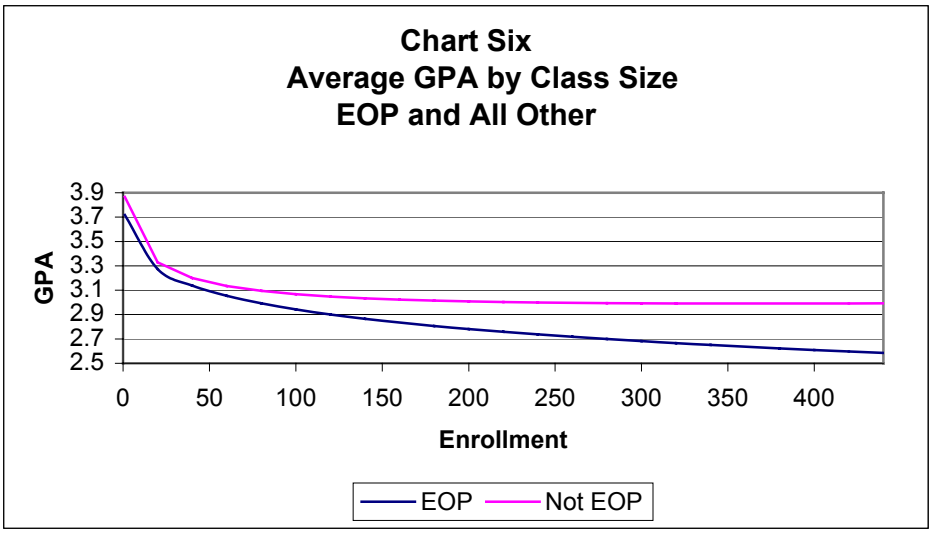
Chart Six
Average GPA by Class Size
EOP and All Other
GPA
3.9
3.7
3.5
3.3
3.1
2.9
2.7
2.5
0
50
100
150
200
250
300
350
400
Enrollment
EOP
Not EOP

Bibliography

Akerhielm, Karen 1995. "Does Class Size Matter?" *Economics of Education Review* 14 (3) (9): 229-241.

Bolander, Steven F. 1973. "Class Size and Levels of Student Motivation," *Journal of Experimental Design* 42(2): 12-18.

Borden, Victor M. H. and Kathy L. Burton 1999. "The Impact of Class Size on Student Performance in Introductory Courses," paper presented at the 39th Annual Conference of the AIR in Seattle, WA.

Brewer, Dominic, Cathy Krop, Brian P. Gill, and Robert Reichardt 1999. "Estimating the Cost of National Class Size Reductions Under Different Policy Alternatives," *Educational Evaluation and Policy Analysis*, 21(2): 179-192.

Correa, Hector 1993. "An Economic Analysis of Class Size and Achievement in Education", *Education Economics*, 1(2): 129-135.

Ehrenberg, Ronald G., Dominic J. Brewer, Adam Gamoran, J. Douglas Willms, and Glenn Zorpette 2001. "Does Class Size Matter?" *Scientific American.*

Feldman, Kenneth A. 1984. "Class Size and College Students' Evaluations of Teachers and Courses: A Closer Look." *Research in Higher Education* 21(1):45-91.

Finn, Jeremy D., and Charles M. Achilles 1990. "Answers and Questions About Class Size: A Statewide Experiment." *American Educational Research Journal* 27(3):557-577.

Finn, Jeremy D., and Charles M. Achilles 1999. "Tennessee's Class Size Study: Findings, Implications, Misconceptions." *Educational Evaluation and Policy Analysis* 21(2):97-110.

Glass, G. V., L. S. Cahen, M. L. Smith, and N. N. Filby 1982. *School Class Size: Research and Policy*. Beverly Hills: Sage.

Glass, Gene V., Barry McGaw, and Mary Lee Smith 1981. *Meta-Analysis in Social Research*. Beverly Hills, California: Sage Publications.

Glass, Gene V. and Mary Lee Smith 1979. "Meta-Analysis of Research on the Relationship of Class-Size and Achievement." *Evaluation and Policy Analysis* 1: 2-16.

Gursky, Daniel 1998. "Class Size Does Matter." *The Education Digest* 64(2 October):15-18.

Hanushek, Erik A. 1995. "Moving Beyond Spending Fetishes," *Educational Leadership* 53(3) 60-64.

Hanushek, Eric A. 1996. "Outcomes, Costs, and Incentives in Schools", *Improving America's Schools: The Role of Incentives* Washington, D. C. National Academy Press.

Hanushek, Eric A. 1999a. "Some Findings From an Independent Investigation of the Tennessee STAR Experiment and From Other Investigations of Class Size Effects," *Educational Evaluation and Policy Analysis,* 21(2): 143-164.

Hanushek, Eric A. 1999b. The Evidence on Class Size, in Mayer, Susan E. and Paul E. Peterson (Eds), *Earning*

and Learning: How Schools Matter (Washington D.C.,Brookings Institutional Press)

Hanushek, Eric A., and Lori L. Taylor 1989. "Alternative Assessments of the Performance of Schools: Measurement of State Variations in Achievement," *The Journal of Human Resources*, 25(2): 179-200.

Harder, Heather 1990. "A Critical Look at Reduced Class Size", *Contemporary Education,* 62(1): 28-30.

Hoff, David J. 1998. "Federal Class-Size Reports Do an About-face," *Education Week* 17(9 June 10): 22.

Hoxby, Caroline M. 2000. "The Effects of Class Size on Student Achievement: New Evidence From Population Variation," *The Quarterly Journal of Economics* 11: 1239-1285.

Kennedy, Peter E and John J. Siegfried 1996. "On the Optimality of Unequal Class Sizes," *Economics Letters*, 50: 299-304.

Kennedy, Peter E and John J. Siegfried 1997. "Class Size and Achievement in Introductory Economics: Evidence from the TUCE III Data" *Economics of Education Review* 16(4October): 385-394.

Kmenta, Jan 1986. *The Elements of Econometrics*, 2nd ed., MacMillan: New York

Krueger, Alan B. and Diane M. Whitmore 2000. "The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project Star"

Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions," *The Quarterly Journal of Economics* 5: 497-531.

Lipman, Steven A. 1990. "On the Optimality of Unequal Class Sizes," *Economics Letters*, 33: 193-196.

Maxwell Nan L. and Jane S. Lopus 1995. "A Cost Effectiveness Analysis of Large and Small Classes in the University," *Educational Evaluation and Policy Analysis* 17(2): 167-178.

McConnell, Campbell R. and Kim Sosin 1984. "Some Determinants of Student Attitudes Toward Large Classes" *Journal of Economic Education* Summer 1984 181-190.

McKeachie W. J. 1980. "Class Size, Large Classes and Multiple Sections," *Academe* (2): 24-28.

McKeachie W. J., Asghar Iran-Nejad, and David C. Berliner. 1990. "The Multi-source nature of learning: An Introduction," *Review of Educational Research* 60(4) 509-516.

Mincer, J. (1974). "Schooling, Experience and Earnings" NBER.

Molnar, Alex, Phillip Smith, John Zahorik, Amanda Palmer, Anke Halbach, and Karen Ehrle. 1999. "Evaluating the SAGE Program: A Pilot Program in Targeted Pupil-Teacher Reduction in Wisconsin," *Educational Evaluation and Policy Analysis* 21(2): 165-178.

Mosteller, Fredrick, 1999. How Does Class Size Relate to Achievement in Schools?, in Mayer, Susan E. and Paul E. Peterson (Eds), *Earning and Learning: How Schools Matter* (Washington D.C., Brookings Institutional Press)

Nelson, Randy and Kathleen T. Hevert 1992. "Effect of Class Size on Economies of Scale and Marginal Costs in Higher Education," *Applied Economics*, 24: 473-482.

Nye, Barbara, Larry V. Hedges, and Spyros Konstantopoulos 1999. "The Long-Term Effects of Small Classes: A

Five-Year Follow-Up of the Tennessee Class Size Experiment," *Educational Evaluation and Policy Analysis* 21(2): 127-142.

Pascarella, E.T and P. T. Terenzini 1991. *How College Affects Students*, San Francisco: Jossey-Bass.

Raimondo, H. J., L. Esposito, and I. Gershenberg 1990. "Introductory Class Size and Student Performance in Intermediate Theory Courses," *Journal of Economic Education*, 21: 369-381.

Ritter Gary W. and Robert F. Boruch 1999. "The Political and Institutional Origins of a Randomized Controlled Trial on Elementary School Class Size: Tennessee Project STAR," *Educational Evaluation and Policy Analysis* 21(2): 111-125.

Santa Barbara 2001: www.cc.santa barbara.ca.us/cced/budget/research / proposedbudget /document/secIB.pdf

Slavin, Robert 1990. "Class Size and Student Achievement: Is Smaller Better?" *Contemporary Education* 62(1): 6-12.

Smith, Mary Lee and Gene V. Glass 1980. "Meta-analysis of Research on Class Size and Its Relationship to Attitudes and Instruction," *American Educational Research Journal* 17(4): 419-433.

Spahn, Karen 1999. "Class Size and Faculty Effectiveness and Quality", Paper Presented at the 39th Annual Institutional Research Forum in Seattle, WA.

Tillitski, Christopher 1990. "The Longitudinal Effects Size in PRIME TIME, Indiana's State Sponsored Reduced Class Size Program," *Contemporary Education* 62(1): 31-32.

Weiss, Trish 1990. "Indiana's PRIME TIME," *Contemporary Education* 62(1): 31-32.

Williams David D., Paul F. Cook, Bill Quinn and Randall P. Jensen 1985. "University Class Size: Is Smaller Better?" *Research in Higher Education* 23(3): 307-317.

Word, Elizabeth, Charles M. Achilles, Helen Bain, John Folger, John Johnston, and Nan Lintz 1990. "STAR Final Executive Summary: Kindergarten Through Third Grade Results," *Contemporary Education* 62(1): 13-16.

DATA APPENDIX

Grade: Is the grade a student received in a credit bearing section that counts towards their cumulative GPA (Pass/Fail, Satisfactory, Incomplete etc. are not included). Grades are re-numbered 0=F, 1=D, 2=C-,3=C,4=C+,5=B-,6=B,7=B+,8=A-,9=A.

Experience: Is measured by student level at the start of the semester and range from 1 for first semester freshmen to 8 for second semester senior.

Ability: For each other course a student took in a given semester, we measure the difference between the grade they received and the average grade given in that course. The sum of these differences, Other Grades, measures a student's ability in that semester. We also tested scores, high school standing, and cumulative GPA from prior college work. The overall results are essentially the same. Note that in labor theory, ability is generally considered to be temporally in variant. We allow for temporal variation that can be thought of as a combination of course specific ability and motivation.

Class size: The natural log of the number of students registered for the class at the end of the third week of classes.

Department: Acknowledging that different academic departments grade differently, we included a variable (Dept Mean) that is the value of the mean grade given by the academic department over the time period covered in this analysis.

Education: A dummy variable equaling 1 for students admitted with AP Credit**.**

Female: A dummy variable equaling 1 for students that are Female.

Minority: A dummy variable equaling 1 for students who report their ethnic background as Black-Non Hispanic, Hispanic, or American Indian/Alaskan Native.

Educational Opportunity Program: A dummy variable equaling 1 for students admitted into the EOP program.

TABLE A-1
Descriptive Statistics
Sample Size = 363,023

	Min	Max	Mean	SD	Median
Grade	1.0	9.0	6.41	2.49	7.00
Student Level	1.0	8.0	4.87	2.24	5.00
AP Credit	0.0	1.0	0.43	0.50	0.00
Ability	-20.0	8.7	0.09	2.22	0.31
LN Class Size	0.0	6.1	4.05	1.09	3.91
Class Size	1	445	57.4	2.9	49
Dept. Mean Grade	4.6	9.0	6.41	0.82	6.42
Female	0.0	1.0	0.54	0.50	1.00
Minority	0.0	1.0	0.11	0.31	0.00
EOP	0.0	1.0	0.06	0.23	0.00

Endnotes

[1] Student/pupil ratios in K-12 schools had been dropping since the 1950's without any marked increased in standardized test scores or other indicators of overall student performance, and the majority of the studies conducted at the classroom level showed either no or very modest affect of class size on student performance. The U.S. Department of Education reports that K-12 student teacher ratios fell from 26.9 in 1955 to 17.2 in 1998. Yet average class sizes remain at about 24. The increase in special education teachers is believed to be the principle reason for this apparent contradiction.

[2] The negative slope suggests that the ideal class size from the point of view of the student's learning is size one. The concavity suggests an optimal tradeoff might exist between the student and the school (society). If concave, the rate of fall off in student outcome decreases slowly at first, and then more rapidly. If the costs of providing student outcomes is typical, it may also decline per student as the output numbers increase, but rapidly at first as the costs of facilities and faculty are distributed over more students, and less rapidly at larger numbers of students as marginal efficiencies diminish. Hence, there may be a societal optimum, assuming society bears the costs of education and receives its benefits, where the rate of dimunation in outcomes equals the rate of dimunation in per student costs. This would not hold, or at least not hold over the entire range of student numbers if benefits were a convex function.

[3] Lipman, 1990; Kennedy and Siegfried, 1996, 1997.

[4] Heavily weighting studies that they considered more experimental in design, and discounting those they considered non- or quasi-experimental, Glass et al. (1982) argued that the positive effect of smaller class sizes results from attitudinal changes in both teachers and students in that environment.

[5] The most extensive experiment was Tennessee's STAR project. (Word et al., 1990; Ritter and Boruch, 1999) The results of the STAR Project showed that students scored better on 3rd grade standardized tests in math and reading if they had attended smaller sized kindergartens (Finn and Achilles, 1990, 1999; Krueger, 1999). Follow up studies showed that those students who continued in small classes beyond kindergarten did better than those that did not (Nye et al., 1999), and that small classes seem to be most beneficial to those coming from disadvantaged backgrounds (Krueger and Whitmore 2000; Slavin 1990). Subsequently, the findings from the STAR program and more modest experiments elsewhere (Tillitski, 1990; Molnar et al., 1999;Weiss, 1990) heavily influenced California's decision to spend 6 billion dollars on class size reduction (Santa Barbara, 2001).

[6] The evidence suggests that average class sizes must be reduced to 15 to achieve significant improvement in test scores, yet it has been estimated that this would cost up to eleven billion dollars a year if enacted nationwide at the K-12 level (Brewer et al., 1999). While the STAR project does show significant improvement in students attending smaller sized kindergarten, the estimated beneficial effect of continuing in small classes is modest and its significance debatable (Harder, 1990; Slavin, 1990). Further, the implementation of the STAR experiment has been question. The attempts to randomly assign students to different sized classrooms may not have been perfect, given that some parents may have tried to get their child into the treatment group of smaller classes. For similar reasons, the morale of teachers and students in control groups might have been different than those assigned to the treatment groups (Hanushek, 1995, 1996, 1999a, 1999b). Indeed, in a recent sophisticated statistical analysis, Hoxby (2000) critiques numerous class size studies on the basis of how they assigned students to different size classrooms. Using an exogenous assignment model she found only sketchy evidence that class size positively influences performance. See also Akerhielm (1995), Borden and Burton (1999), Correa (1993), Ehrenberg et. al (2001), Gursky (1998), Hanushek and Taylor (1989), Hoff (1998), Mosteller (1999).

[7] The variables entered in to the stepwise in the following order: Ability, Department, Class Size, AP Credit, Experience squared, Class Size cubed, Ability squared, Minority Status, Gender, EOP, Ability cubed, Experience, Experience cubed, Class size squared.