Revision for resubmission
January 16, 2006 11:00 AM

# The Effects of Class Size on Student Grades at a Public University

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

Cornell Higher Education Research Institute (CHERI)
School of Industrial and Labor Relations
Cornell University
Ithaca, NY

January, 2006

* Corresponding author
Edward C. Kokkelenberg
Visiting Fellow, Cornell Higher Education Research Institute (CHERI),
School of Industrial and Labor Relations, Cornell University, Ithaca, NY 14853-3901
607-255-9263 (phone)
eck4@cornell.edu
607-255-4496 (fax)
and Professor of Economics, Department of Economics, Binghamton University, Binghamton, NY 13902

**Assistant Director, Institutional Research and Planning, Binghamton University, Binghamton, NY 13902.

***Department of Economics, Binghamton University, Binghamton, NY 13902.

ABSTRACT

We model how class size affects the grade that higher education undergraduate students earn and test the model using an ordinal logit with and without fixed effects on over 670,000 observations from a public university. We find that class size negatively affects grades for several specifications and subsets of the data, as well as for the whole data set. The specifications tested hold constant for academic department, peer effects, student ability, level of student, level of course, gender, minority status, etc. Average grade point declines as class size increases, precipitously up to class sizes of twenty, and gradually but monotonically through larger class sizes. The associated elasticity of Grade-Class Size is estimated to be -0.066 and this is the largest absolute value for variables controllable by the university. We conclude that there are dis-economies of scale associated with a deterioration of student outcomes as class sizes grow larger. The cost of this deterioration is not easily quantifiable as much of the costs are non-market costs and unobservable. Future studies of economies of scale in higher education need to address the traditional assumption of constant product quality.

JEL Classification; I21

Key Words; Economies Of Scale, Educational Economics, Student Performance, Logit Analysis, Fixed Effects Models

# INTRODUCTION

This present paper makes a contribution to understanding a major problem of resource allocation in the faculty staffing of classes in higher education. It has been observed that if faculty can teach larger class sizes with no adverse outcomes, then economies of scale may not always be utilized. If larger classes adversely affect student outcomes, then perhaps institutions are incurring diseconomies of scale (see Hancock, 1996.)[1]

If they exist, economies of scale are a particularly attractive way to reduce costs at schools experiencing increasing demands for education and where the quality of the incoming students appears to be rising or steady[2]. Schools often look to spreading the costs of a faculty member over more students by increasing class sizes or by increasing workload (number of courses taught per term). Faculty senates, faculty unions, and often trustees, often take the easier option of marginally increasing class size as a way to realize economies of scale. The pressure to achieve class size economies of scale is discussed in (Nelson and Hevert, 1992; Toth and Montagne, 2002; and Moore, 2003).

But the question arises; is the education received in a large class the same as that in a small class? To bring further light on this question, we estimate the influence of class size on student achievement in higher education. We model grades as an output and test this model using a very large dataset from a medium-sized public research university.

Applying a logistic regression with and without a fixed effects model we find that class size is an important negative 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 size on K-12 student performance. We explore several specifications, 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. We also test the results by academic department. In all cases we find class size negatively affects student grades. The estimated Grade-Class Size elasticity is negative and large relative to the other factors controllable by the school. We conclude that any considerations of economies of scale must consider the scale effects on the quality of output. Schools that seek to reduce costs by increasing class sizes may need to take steps to train faculty or otherwise rectify poorer student outcomes and other diseconomies of scale.

## 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[3]. However, Glass and Smith, in a series of articles beginning in the late 1970’s (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[4]. This model has become a basis for further normative discussion on whether, or how, class sizes should vary[5]. 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[6]. Given this apparently beneficial evidence of smaller class sizes, several states designed experiments to replicate Glass's *et.al.* findings[7]. In 2003, a number of articles appeared in a special edition of The Economic Journal (V113, February) concentrating on U.S. and U.K. experiences and summarizing a vast amount of literature. The papers therein concentrate on data from K-12 to examine this question (see Dustmann, 2003).

Even though there is now strong evidence that smaller class sizes improve student performance, at least in some circumstances, and using common methodologies to test the data, the debate continues. 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; and Hanushek, 2003)[8]. Further methodological challenges have weakened these claims (Maasoum, Millmet, and Rangaprasad, 2003; Kruger, 2003).

### Higher Education

Though there is debate about the extent of benefits small classes bring, or how much it costs to achieve these benefits, there is at least some agreement in the K-12 literature that, using certain tests, class size matters in some 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 (Williams, Cook, Quinn and Jensen, 1985; Pascarella and Terenzini, 1991), the authors conclude that the overall evidence suggests that class size plays little or no influence on student achievement. This however has not quelled the debate. McKeachie (1980) and McKeachie, Iran-Nejad, and Berliner (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.

A summary of more recent research is given in Toth and Montagne (2002) and Kwantlen University (2004). Toth and Montagne summarize eight studies from 1990 to 1999 and find mixed results for three studies, positive increases in outcomes where class sizes are reduced for two studies, outcomes to be better in large classes for one study, and no significant results for two studies. Kwantlen summarizes other recent research showing variously no relationship between class size and achievement; negative relationship (larger classes yield less student achievement); larger classes enhance student outcome; large classes are as effective as smaller classes; and that student characteristics and instructional design are important factors. Kwantlen quotes an Ohio State website “Research Results: Mixed” and concludes that for courses that emphasize recall of facts, large classes are as equally effective as small classes; for courses emphasizing “problem-solving, critical thinking, long-term retention, and attitude toward the discipline… small classes are more successful.” (Kwantlen, 2004, pg 3).

These studies generally focus on one discipline (e.g. Communications), and often one level of course work (e.g. Introductory Economics). Sample sizes are often small (comparing students from two classes, 25 and 50 students), arbitrarily grouping students into three categories (small classes of 30 or less, medium of 31 to 90, and large of over 90 students). These studies use several different dependent variables including tests for persistence, scores on a post-class test specific to a discipline, grades, or grade distribution of the class.

Also, there is some further evidence that class size may matter in some courses or disciplines, but not in others. Raimondo, Esposito and Gershenberg (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's argument, that smaller classroom environments enhance the more wide-ranging, non-formula based knowledge necessary for understanding macroeconomic principles. There is also an argument for small classes in the performing arts where skills and techniques are individually taught[9].

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 whatever the variation in teaching method or classroom that is under study, is attributed to the changed element. Investigators have both a control group, and a tested, agreed-upon metric. We lack control groups and an agreed-upon metric in most 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 conclusion. Finally, much of the K-12 testing is done for specific academic subjects, such as chemistry or reading comprehension. We are not aware of a comparable single set of before-and-after test scores that is applicable across academic subjects in higher education.

We address many of the above deficiencies in this paper and present findings, based on a very large dataset from a single institution covering twenty-four semesters and forty-five disciplines, of how class size (measured continuously) affects student outcomes, as measured by grades, after controlling for other relevant student and course characteristics. We motivate the discussion using the economic 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 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. At the end of some period of study they are rewarded with some sort of certification, which in turn may result in earning higher lifetime incomes and increased non-monetary utility. During this time they are paid by a form of scrip, 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 an additional 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 W as the wage, and 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 ith student in the jth class during period t:

$$W_{ijt} = b_0 + \phi(E_{it})'\beta + \theta(A_{it})'\Gamma + Z_{it}'\lambda + V_{jt}'\kappa \tag{1}$$

Here, W represents the grade, E the ith student's experience (e.g. level in college), A represents ability, Z a vector of student-related variables, and V is a vector of environmental, faculty, and subject matter factors including class size (CS). N(E) and 2(A) are allowed to be polynomials in E and A, and ∃, ϶, 8, and 6 are vectors of parameters to be estimated while $b_0$ denotes a vector of constants, also to be estimated.

The null hypothesis is that class size does not affect student learning or performance and this would be reflected in the stability of grade distributions over various class sizes for various subjects, while holding the other independent variables constant.

## 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 totaling 998,898 observations. The data consists of all undergraduate students for the period fall 1992 through spring 2004. Students take courses 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 analysis reducing the number of observations. This results in a censored sample and any assumption that those dropped have the same distributions of characteristics as those retained cannot be made. The resulting bias in our results however would be to support the null hypothesis as students that withdraw most often do so as they expect low grades. Further reductions incurred when certain variables were censored. The resulting basic overall dataset contains over 764,000 observations. The variables and data are discussed further in Appendix A.

## RESULTS[10]

We begin by presenting the results of a model of grades (W) as explained by relative ability ($A_r$), the class mean grade that takes peer effects in account ($\overline{W}_C$), the departmental mean grade ($\overline{W}_D$),class size (CS), initial objective ability (SATM, SATV), the presence of advanced placement courses in high school (AP), experience on campus as a student (entered as a freshman (F), and student level (L)), gender (G), minority (M), and time (Y). The model (*sans* subscripts) is given as:

(2)

$$W = b_0 + \beta_1 A_r + \beta_2 A_r^2 + \beta_3 \overline{W_C} + \beta_4 \overline{W_D} + \beta_5 CS + \beta_6 CS^2 + \beta_7 SATM + \beta_8 F + \beta_9 SATV +$$

$$\beta_{10} G + \beta_{11} M + \beta_{12} L + \beta_{13} Y + \beta_{14} AP + \beta_{15} CS * G$$

The results of estimating an ordered logistic of Equation (2) are shown in Table 1. The first numeric column is for the full dataset whereas the next two columns show the results for two sub samples of the data. The first of these, labeled IQ1, is for the 342,289 observations lying within the interquartile range of class sizes; the second, IQ2, uses the 271,941 observations lying within the interquartile range of grades[11]. This model, using these three subsets of data explains the observed data well. SAS reports a series of measures of association between probabilities predicted by the model and those observed in the data (analogous to an adjusted R-squared). Among these are the "G," the "tau-a" and the "c" statistics. The "G" statistic, a ratio of the likelihoods calculated from the model with only intercepts and that calculated from the model with the independent variables, is distributed chi square with 15 degrees of freedom. The critical value at $P_R = 0.005$ is 32.801, and our "G" values exceed this (see Table 1). The "c" statistic's theoretical range is from 0.0 to 1.0 (0.5 or lower indicates that the model's predictions are no better than chance). Our regression results are 0.765, 0.773, and 0.731 (see Table 1) indicating a high discriminatory power of the model. The "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. The calculated values are 0.450, 0.465, and 0.389. In summary, the model explains the observed data very well indeed.

Turning next to the individual parameters from the logistics regression we find that all independent variables (with one exception for the time variable), including class size, have a statistically significant influence on grades[12], as all the p values are less than 0.0001. Note that Table 1 reports standard errors. Experience and ability are positively related to grade. The coefficient for Minority students is negative suggesting they do less well than non-minorities, females and those with high SAT scores do better (positive coefficients) but females do worse in larger classes (CS*G is negative). The departmental mean grade has the largest single impact on grades. This indicates that further work in this area should account for departmental grading culture, traditions, and the material presented in class, as McKeachie suggests.

The chief result of interest in this paper is that class size enters all estimations with a negative value (-0.007, -

0. 012, -0.008 for each of the three datasets (see Table 1)). Note also the positive estimates of the squared term, $CS^2$, are consistent with the concave model suggested by Glass *et.al*. Therefore, the null hypothesis that class size does not matter can be rejected. We also found this result to be robust as to variations in other proxies for experience, ability, department, and faculty and for other classroom environmental variables. Further, the standard errors on the class size terms are small (.00022, .000102, and .000038 respectively). The coefficients at plus or minus two standard errors for class size thus range from –0.002014 to –0.001926, -0.004964 to –0.004556, and –0.002306 to –0.002154; all negative and narrow ranges.

We also calculated the elasticities of several key variables that the university can control and these are: Grade-Class Size = -0.066; Grade-AP Credits= +0.017; Grade-Relative Ability = +0.044. Class Size is almost exclusively the purview of the school, though given a choice, students can opt for the smaller class on offer. The other two variables are only partially controllable via admissions and the enforcement of prerequisites. This limited control holds even more so for the other independent variables in Table 1, e.g. gender. While modest in size, the Grade-Class Size elasticity is the largest value reported, negative, and based on a statistically significant result. This supports our overall conclusion noted elsewhere; class size is important and large classes adversely affect grades.

One could argue that the results are determined by the differing social structures in small versus large classes and that faculty are reluctant to give poor grades in small classes but more willing to award low grades to more anonymous students in large classes. To test if this is what drives our results, we re-estimated our model for other subsets of the data (the mid 90% and 80% ranges of class size, as well as for successively larger minimum class size cut-offs, and again for successively smaller maximum class size cut-offs). These results (Table 2) show the parameters on CS are consistently statistically significantly negative, *ceterus paribus*[13].

Another possible explanation of our results is prompted by the work of Bressoux, Kramarz and Prost (2005), who found that poor students benefited from small classes whereas good students did not. A set of regressions was run to test this. The data was partitioned by cumulative grade point average at the start of the relevant term into two groups: the first consisted of 146,150 observation where students' GPA was in the upper 40 percent and the second consisting of 276,327 observations where students' GPA was in the lower 40 percent of the GPA distribution. The model was that of Equation (2). The coefficients on class size and the square of class size from the ensuing logistics regressions were comparable to those in column 1 of Table 1. The class size coefficients were –0.0068 and –0.0074 for upper GPA and lower GPA students respectively. Hence, while poorer students seem more adversely affected by large classes, both groups suffer from increased class sizes *ceterus paribus*, and students appear not to elect into courses based upon their GPA and the anonymity available in large classes.

The analysis, using an abbreviated model, was next extended to ten diverse departments, Economics, Psychology, Political Science, Chemistry, Computer Science, English, History, Management, Mathematics and Music. These results also indicate class size has a negative impact in seven and a non-significant effect in the other three departments (see Table 3). Note further, the significant coefficients on class size are approximately one magnitude apart at most across departments, ranging from -0.0003 to -0.007 (Mathematics to Music). Note also, while women do better overall (see Tables 1 and 3), they do worse in Economics and Chemistry.

We also estimated Equation (2) for several introductory courses in Mathematics and Economics that had sections of widely varying sizes over the time period observed. The class size coefficient estimated is always statistically significantly negative (See Appendix B).

In summary, all of these results sustain the view that the effect of class size on grades is negative over a wide range of class sizes, courses, and departments, holding other demographic and student variables constant.

Next we show the results of analyzing a subset of data graphically. Figures 1 and 2 show cumulative probabilities for specific grades by class size. The first deals with all classes, the second with classes sized six students or greater. Again, the message is that students in large classes have a higher probability of lower grades than those in smaller classes *ceterus paribus*. Note that the probabilities fall rapidly 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 80 than for increasing class sizes from 10 to 20.

## The Fixed Effects Model

If one treats the data as a panel data set, where the individual student is the unit of observation then a fixed effects model can be given as:

$$W_{itj} = \beta_0 + \beta_i + \beta_t + \beta_1 E_{it} + \beta_2 A_{it} + \beta_3 CS_j + \beta_4 \overline{W}_{Dj}$$

(3)

Here $\beta_i$ is the student-fixed effect and $\beta_t$ the semester-fixed effect. These two variables allow us to control for individual attributes not explicitly contained in the experience (level) and relative ability variables (which may 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 are assumed to be the same as the marginal effects between any other grade pair, say B minus and C plus.

We estimate a polynomial variant of Equation (3) in both fixed effects and no fixed effects sub-variations. These are Models 1 and 2 of Table 4. In the first Model, the data was for 167,928 students and in the full data set. 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 4 is for the same data but estimated without fixed effects 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 (3) for a random sample of 10,000 students chosen from the 167,928 observations[14]. The results are reported in Table 5. Again, the model includes an experience variable, an ability variable to allow for time varying student ability, a departmental variable, and a class size variable. All fixed student characteristics are differenced out against the individual student's 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 in this subset 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 are consistent with and confirm the results from the ordinal logit estimation reported above in Tables 1, 2, and 3.

## DISCUSSION

This study of grades in higher education, using various models relating environment, ability, and experience to undergraduate course grades, shows that class size has a negative relationship to grades and that while the value of the class size coefficient differs across different departments and subsets of data, it is negative in all cases. Further, the estimated grade–class size elasticity is modest but negative. This is an important result bringing further information to an ongoing debate as to how to achieve efficiencies in higher education.

Though we have found a negative relationship between grades and class size, we cannot conclude, to the extent that grades are but a proxy for knowledge, that students learn more in smaller classes, [10] nor do we offer a reason for our result. As Glass *et. al*. (1982) argued, attitudinal changes among faculty and students might account for the observed results. Recall that McKeachie (1999) suggests that optimal teaching methods and class sizes vary by subject matter and level. He also reminds us that students may self-select class sizes whenever possible. Alternatively, as the K-12 literature suggests, the attention faculty can give to individual students and the intensity of engagement in learning that occurs in small classes could account for the results. We do observe however, that the negative relationship persists even when we account for variations in data subsets, models, included variables, and statistical methodology: a robust result.

We conclude that there are diseconomies of scale associated with a deterioration of student outcomes as class sizes grow larger. The cost of this deterioration is not quantifiable with our data, as much of the costs are non-market costs and unobservable. For example, these costs may include lost revenue due to a decrease in student persistence and a resulting lower student retention rate as well as the loss of reputation caused by lower graduation rates.  Nor have we quantified the added cost of training and properly staffing large classes to negate adverse grade effects. We do conclude that any institutional benefits from larger classes must be set against the short and long-term costs associated with the resulting poorer student performance. The evidence presented in this paper suggests class size influences the likelihood of getting good grades and that future studies of economies of scale in higher education need to consider the effect class size has on student outcomes.

## Acknowledgements

We wish to thank Ronald Ehrenberg for helpful comments. We would also like to thank Jack Keil, Peter Partell and Corine Prost for comments on early drafts and Jessica Richards, Yu Zhu and Hester Han for analytic expertise. Finally, we thank Mary Elizabeth Eddlestone for computational guidance. All errors and omissions are the responsibility of the authors.

# APPENDIX A: DATA

The data comes from a wide variety of courses taught by over 40 academic departments listed in Table A-1. The overall grade distribution is listed in Table A-2.

Relative Ability: Normalized grade point average of all other courses student is taking in a given semester relative to that of other students in the course in question. 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 invariant. We allow for temporal variation that can be thought of as a combination of specific ability, motivation and learning by doing.

GPA: Grade Point Average on 4-point scale.

Departmental Mean Grade: Average grade awarded by relevant department over entire time period covered by this study on a 4-point scale.

Class Size: Class size after add deadline or the third week of class.

Student Level: Student level based upon earned credit hours on scale of 1 to 8 where 1 and 2 are freshman, etc.

Female: Dichotomous variable, one if female, zero otherwise.

Minority: Dichotomous variable, one if under-represented minority (Black, Hispanic, Alaskan Native/American Indian), zero otherwise.

Grade: Numeric value of course grade student received in credit bearing section; F = 0, D = 1, C minus = 2, C = 3, C plus = 4, B minus = 5, B = 6, B plus = 7, A minus = 8, A = 9.

Cumulative GPA: Individuals cumulative GPA at the start of the relevant term; an alternative measure of ability, motivation and circumstances or prior success in college.

AP Credit: Dichotomous variable; one if student entered with Advanced Placement credit, zero otherwise.

Year: Scaled log of time variable.

Entered as freshman: Dichotomous variable, one if so entered, zero otherwise.

Class Mean: Grade point average of peers enrolled in specific course of interest.

SAT Scores: normalized SAT scores (0,1).

TABLE A-1

(Origin of Department Course)

| Department | Frequency | Percent |
|---|---|---|
| Africana Studies | 214 | 0.44 |
| Anthropology | 1626 | 3.32 |
| Art History | 443 | 0.90 |
| Art Studio | 517 | 1.06 |
| Biological Sciences | 2932 | 5.98 |
| Chemistry | 1729 | 3.53 |
| Cinema | 361 | 0.74 |
| Classics & Near Eastern Studies | 118 | 0.24 |
| Comparative Literature | 676 | 1.38 |
| Computer Sciences | 1691 | 3.45 |
| Economics | 2159 | 4.41 |
| Electrical Engineering | 515 | 1.05 |
| English, Gen. Lit. & Rhet | 3951 | 8.06 |
| Engineering Design | 327 | 0.67 |
| Geological Sciences | 1204 | 2.46 |
| Geography | 786 | 1.60 |
| German, Russian & East Asian Languages | 601 | 1.23 |
| Harpur – Dean's Office | 174 | 0.36 |
| History | 2385 | 4.87 |
| Human Development | 1265 | 2.58 |
| Judaic Studies | 407 | 0.83 |
| Latin American Studies | 127 | 0.26 |
| Linguistics | 108 | 0.22 |
| Management | 3621 | 7.39 |
| Mathematical Sciences | 2488 | 5.08 |
| Mechanical Engineering | 643 | 1.31 |
| Medieval Studies | 41 | 0.08 |
| Music | 2198 | 4.49 |
| Nursing | 940 | 1.92 |
| Off Campus College | 234 | 0.48 |
| Philosophy | 1751 | 3.57 |
| Physics, Applied Physics and Astronomy | 1013 | 2.07 |
| Physical Education | 3043 | 6.21 |
| Political Science | 1393 | 2.84 |
| Psychology | 3471 | 7.08 |
| Romance Languages | 1142 | 2.33 |
| Sociology | 1066 | 2.18 |
| Systems Science/Industrial Engineering | 94 | 0.19 |

| | | |
|---|---|---|
| Theatre | 1207 | 2.46 |
| Women’s Studies | 225 | 0.46 |
| Other* | 106 | 0.22 |

* Other includes Asian Studies, Bioengineering, Education, Latin American Studies, Public Administration and certain courses assigned to administration totaling less than ¼ of 1%.

TABLE A-2
Letter Grade Distribution

| s7grad | Frequency | Percent |
|---|---|---|
| A* | 200,705 | 22.88 |
| A-* | 131,627 | 15.00 |
| AU | 792 | 0.09 |
| B* | 109,172 | 12.44 |
| B+* | 114,740 | 13.08 |
| B-* | 63,703 | 7.26 |
| C* | 41,648 | 4.74 |
| C+* | 43,146 | 4.92 |
| C-* | 23,121 | 2.64 |
| D* | 19,851 | 2.26 |
| F* | 24,479 | 2.79 |
| I | 1855 | 0.21 |
| MG | 507 | 0.06 |
| P | 92,720 | 10.57 |
| R | 145 | 0.02 |
| S | 513 | 0.06 |
| U | 8 | 0.00 |
| W | 8,160 | 0.93 |
| WF* | 63 | 0.01 |
| WP | 346 | 0.04 |
| X | 23 | 0.00 |
| Total | 877,294 | 100.00 |
| Missing | 111,604 | |

*Used in statistical analysis totaling 772,225. The pass grade (P) is assigned to any student earning a pass/fail option who earns a grade of D or better, and accounts for over 88 percent of the unusable grades.

## TABLE A-3
Descriptive Statistics

| | Sample Size | Min | Interquartile range | Max | Mean | 5% Trimmed Mean | Standard Deviation | Median | Mode |
|---|---|---|---|---|---|---|---|---|---|
| Grade | 772,225 | 0.0 | 4.000 | 9.000 | 6.474 | 6.673 | 2.450 | 7.000 | 9.000 |
| Student Level | 988,898 | 1.0 | 4.000 | 8.000 | 4.967 | 5.019 | 2.237 | 5.000 | 8.000 |
| AP Credit | 988,137 | 0.0 | 1.000 | 1.0 | 0.419 | 0.410 | 0.493 | 0.0 | 0.0 |
| Relative Ability | 772,225 | -16.679 | 2.203 | 22.695 | 0.605 | 0.605 | 2.158 | 0.661 | 0.0 |
| Class Size | 988,898 | 1.0 | 117 | 547 | 97.419 | 97.419 | 105.161 | 48 | 25 |
| Dept. Mean Grade | 970,439 | 1.783 | 0.375 | 3.881 | 3.120 | 3.120 | 0.272 | 3.166 | 3.167 |
| Female | 988,137 | 0.0 | 1.000 | 1.00 | 0.538 | 0.538 | 0.499 | 1.000 | 1.000 |
| Minority | 988,137 | 0.0 | 0.0 | 1.00 | 0.109 | 0.109 | 0.312 | 0.0 | 0.0 |
| Year | 988,898 | 6.931 | 10.560 | 47.449 | 37.824 | 37.824 | 9.942 | 41.271 | 47.185 |
| Class mean | 772,225 | 0.0 | 0.609 | 3.949 | 2.953 | 2.953 | 0.531 | 3.000 | NA |
| Entered as freshman | 988,137 | 0.0 | 0.0 | 1.0 | 0.762 | 0.762 | 0.426 | 1.00 | 1.00 |
| Verbal SAT | 871,803 | -5.594 | 1.022 | 3.163 | -0.003 | -0.003 | 1.092 | 0.0 | 0.0 |
| Math SAT | 871,803 | -6.268 | 1.070 | 2.904 | 0.001 | 0.001 | 1.083 | 0.0 | 0.0 |
| GPA | 764,432 | 0.0 | 0.900 | 4.000 | 3.106 | 3.106 | 0.755 | 3.292 | 4.000 |
| Cumulative GPA | 836,536 | 1.000 | 0.693 | 4.000 | 3.102 | 3.102 | 0.484 | 3.135 | 3.000 |

# APPENDIX B

## Class Size Effects In Introductory Mathematics And Economics Courses

| Department<br>Course | Number of Observations | Class size Q1 | Class Size Q2 (Median) | Class Size Q3 | Class Size Coefficient (Std. Error) |
|---|---|---|---|---|---|
| Mathematics | | | | | |
| Introductory Statistics | 5882 | 225 | 309 | 374 | -0.00401<br>(0.00094) |
| Calculus I | 6054 | 31 | 56 | 304 | -0.00065<br>(0.00016) |
| Economics | | | | | |
| Poverty and Discrimination | 2976 | 61 | 98 | 121 | -0.00790<br>(0.00426) |
| Introductory Micro | 9205 | 142 | 206 | 224 | -0.00786<br>(0.00298) |
| Introductory Macro | 7391 | 99 | 143 | 217 | -0.00866<br>(0.00161) |

# References

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.

Bressoux, Pascal, Francis Kramarz, and Corinne Prost, 2005, "Teachers' Training, Class Size and Students" Outcomes" mimeo, February.

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.

Dustman, Christian 2003. "The Class Size Debate and Educational Mechanisms: Editorial." The Economic Journal 113 (February), F1-F2.

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

Ehrenberg, Ronald G. 2004. "Key Issues Currently Facing American Higher Education," Mimeo, July.

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.

Hancock, T. M. 1996."Effects of Class Size on College Student Achievement." College Student Journal 30(4): 479-81.

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. 2003. "The Failure of Input-Based Schooling Policies." Economic Journal 113 (February) F64-F98.

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.

Krueger, Alan B. 2003. "Economic Consideration and Class Size." The EconomicJourna 113 (February) F34H.

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.

Kwantlen University Institutional Analysis and Planning, 2004, Summary of Research Findings on Impact of Class Size on Student Learning and Satisfaction, from a/sites/instanaplan/nsf/files/classes%20size%20research%20for%20web.doc downloaded September 16, 2004.

Lesser, Diane, and Judy Ferrand 1998. "Effects of Class Sizes, Grades Given, and Academic Field on Student Opinion of Instruction," Mimeo, War-Wic Community College.

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

Maasoumi, Esfandiar, Daniel L. Millimet and Vasudha Rangaprasad 2003. Robust Inference Concerning the Impact of Class Size on Student Achievement, mimeo, (July).

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.

McKeachie,Wilbert J. 1999. Teaching Tips, Houghton Mifflin; Boston.

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.

Moore, Jenny 2003. "Data and Issues Regarding Increasing Class Size and Quality of Education," mimeo.

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.

NYS 2005-06 Executive Budget: Higher Education, http://publications.budget.STATE.NY.US/fy05-06 littlebook/higherEd.html, downloaded 02/03/05.

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.

Rizzo, Michael J. 2004. "A (less than) Zero Sum Game? State Funding for Public Education: How Public Higher Education Institutions Have Lost," dissertation, Cornell University, August.

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.

Toth, Linda S. and Linda G. Montagna 2002. "Class Size and Achievement in Higher Education: A Summary of Current Research," College Student Journal 36(2) pp 253-61.

U.S. Census Bureau 2005. "Statistical Abstract of the United States 2004-5," Washington, D.C.

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.

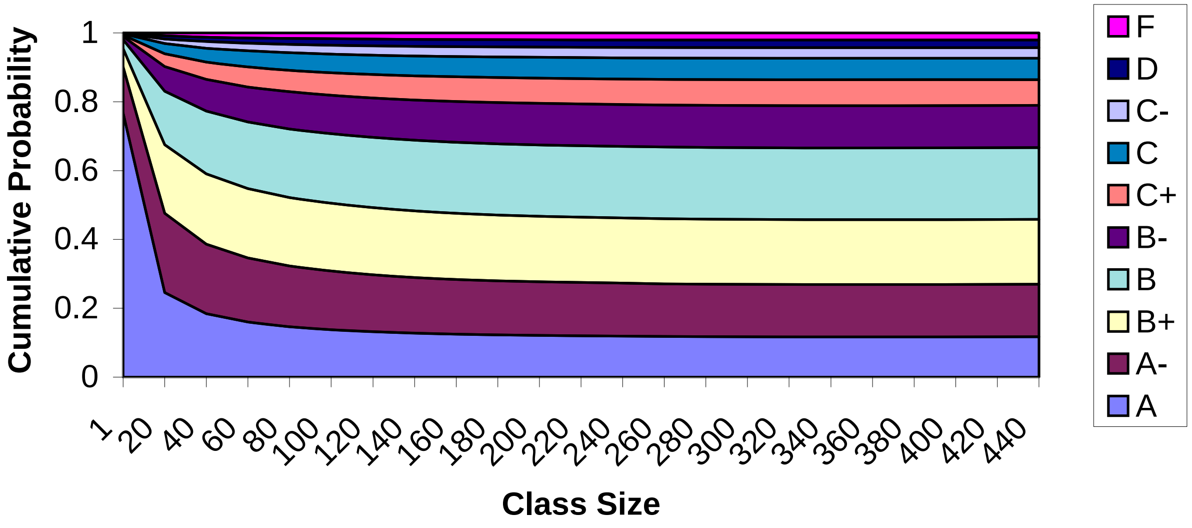

Figure One
Cumulative Probabilty of Grades Received vs. Class Size

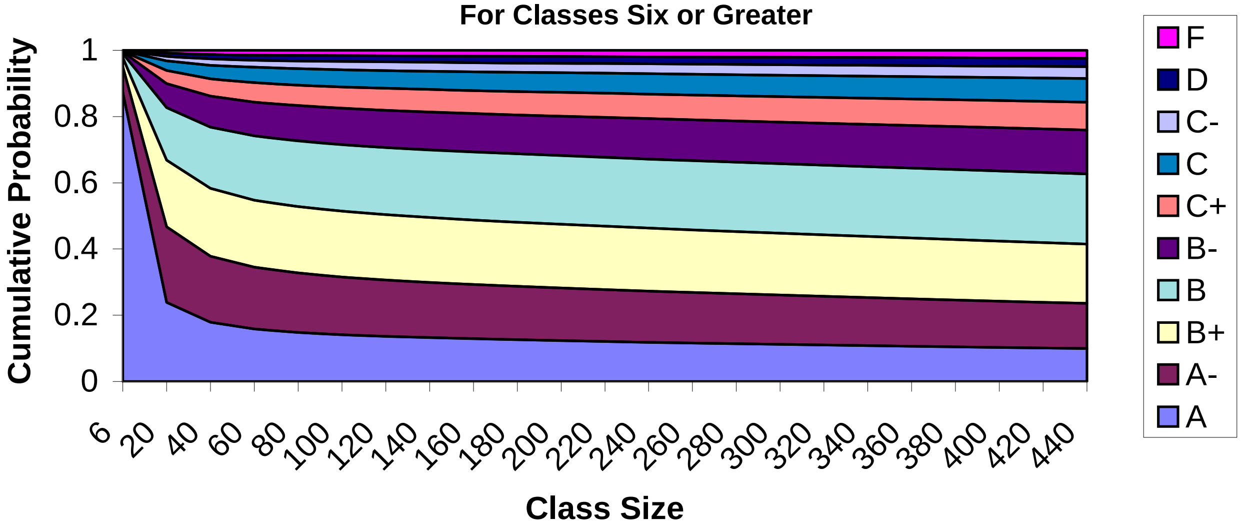

Figure Two
Cumulative Probabilty of Grades Received vs. Class Size
For Classes Six or Greater

# TABLE 1

Estimated Coefficients via Maximum Likelihood-Logistics Procedures
Dependent Variable: Grade (W)

| Variable/Statistic | All Data | | IQ1 | | IQ2 | |
|---|---|---|---|---|---|---|
| | COEF | SE | COEF | SE | COEF | SE |
| Relative Ability ($A_r$) | 0.502 | 0.001 | 0.532 | 0.002 | 0.428 | 0.002 |
| Relative Ability Squared ($A_r^2$) | -0.008 | 0.0003 | -0.007 | 0.0004 | -0.022 | 0.0005 |
| Class Mean $(\overline{W_c})$ | 1.376 | 0.005 | 2.762 | 0.009 | 1.448 | 0.008 |
| Dept. Mean Grade $(\overline{W_D})$ | 1.253 | 0.010 | 0.640 | 0.015 | 1.297 | 0.015 |
| Class Size (CS) | -0.007 | 0.00007 | -0.012 | 0.0005 | -0.008 | 0.0001 |
| Class Size ($CS^2$) | 0.00002 | $1.6x10^{-7}$ | 0.00006 | $3.4x10^{-6}$ | 0.00002 | $2.9x10^{-7}$ |
| Math SAT (SATM) | 0.098 | 0.003 | 0.089 | 0.004 | 0.054 | 0.004 |
| Entered as Freshman (F) | 0.194 | 0.006 | 0.129 | 0.008 | 0.240 | 0.010 |
| Verbal SAT (SATV) | 0.077 | 0.002 | 0.082 | 0.003 | 0.053 | 0.004 |
| Female (G) | 0.278 | 0.007 | 0.343 | 0.014 | 0.198 | 0.010 |
| Minority (M) | -0.228 | 0.008 | -0.200 | 0.011 | -0.228 | 0.012 |
| Level (L) | 0.037 | 0.001 | 0.006 | 0.002 | 0.019 | 0.002 |
| Time (Y) | 0.002 | 0.0003 | 0.001* | 0.0044* | 0.004 | 0.0006 |
| AP | 0.240 | 0.005 | 0.240 | 0.007 | 0.144 | 0.008 |
| CS*G | -0.001 | 0.00004 | -0.003 | 0.0002 | -0.0009 | 0.00007 |
| | | | | | | |
| N | 672,489 | | 342,289 | | 271,941 | |
| Tau-a | 0.450 | | 0.465 | | 0.389 | |
| c | 0.765 | | 0.773 | | 0.731 | |
| Difference (G) | 359,295 | | 197,658 | | 105,486 | |
| $P_R$ > Chi Squared | 0.0001 | | 0.0001 | | 0.0001 | |

IQ 1: Interquartile class size data.
IQ 2: Interquartile grade data.
* Not statistically significant by an $X^2$ test.

## TABLE 2
## Class Size Coefficients for Various Subsets of Data
(Standard Errors in Parenthesis)

| Dataset | No. of observations | CS | $CS^2$ |
|---|---|---|---|
| Total Dataset | 672,489 | -0.00700 | 0.000015 |
| | (100%) | (0.00007) | $(1.64 \times 10^{-7})$ |
| Inter | | | |
| 90% | 611,330 | -0.00851 | 0.000024 |
| | (90.9%) | (0.00012) | $(3.85 \times 10^{-7})$ |
| 80% | 543,965 | -0.00955 | 0.000032 |
| | (80.9%) | (0.00017) | $(6.64 \times 10^{-7})$ |
| Quartile | 342,289 | -0.0123 | 0.00006 |
| | (50.9%) | (0.00054) | $(3.42 \times 10^{-6})$ |
| For all classes greater than | | | |
| 5 students | 657,253 | -0.00566 | 0.000012 |
| | (97.7%) | (0.00007) | $(1.65 \times 10^{-7})$ |
| 10 students | 642,250 | -0.00507 | 0.000011 |
| | (95.5%) | (0.00007) | $(1.67 \times 10^{-7})$ |
| 15 students | 617,071 | -0.00455 | $9.57 \times 10^{-6}$ |
| | (91.8%) | (0.00007) | $(1.69 \times 10^{-7})$ |
| 20 students | 583,815 | -0.00412 | $8.69 \times 10^{-6}$ |
| | (86.8%) | (0.00007) | $(1.73 \times 10^{-7})$ |
| For all classes less than | | | |
| 500 students | 672,010 | -0.00717 | 0.000015 |
| | (99.9%) | (0.00007) | $(1.68 \times 10^{-7})$ |
| 450 students | 664,322 | -0.00811 | 0.000019 |
| | (98.7%) | (0.00008) | $(1.99 \times 10^{-7})$ |
| 400 students | 655,164 | -0.00902 | 0.000022 |
| | (97.4%) | (0.00008) | $(2.43 \times 10^{-7})$ |
| 350 students | 638,186 | -0.0117 | 0.000033 |
| | (94.9%) | (0.00011) | $(3.48 \times 10^{-7})$ |
| 300 students | 622,365 | -0.0145 | 0.000045 |
| | (92.5%) | (0.00013) | $(4.72 \times 10^{-7})$ |
| 250 students | 602,641 | -0.0171 | 0.000058 |
| | (84.6%) | (0.00015) | $(5.84 \times 10^{-7})$ |
| 200 students | 530,056 | -0.0238 | 0.000101 |
| | (78.8%) | (0.00021) | $(1.04 \times 10^{-6})$ |
| 150 students | 487,061 | -0.0360 | 0.00020 |

| | | |
|---|---|---|
| (72.4%) | (0.00031) | (2.04x10$^{-6}$) |

TABLE 3
Estimated Coefficients by Department

| | Economics | Psychology | Political Sc. | Chemistry | Computer Sc. | English | History | Mgmt | Math | Music |
|---|---|---|---|---|---|---|---|---|---|---|
| Relative Ability | 0.618 | 0.540 | 0.556 | 0.585 | 0.531 | 0.417 | 0.501 | 0.642 | 0.562 | 0.211 |
| Class Size | -0.003450 | -0.000289 | 0.0001* | -0.00137 | -0.0002* | -0.0025 | 0.0002* | -0.00097 | -0.00055 | -0.00722 |
| Entered as Freshman | 0.297 | 0.597 | 0.312 | 0.424 | 0.197 | 0.217 | 0.296 | 0.294 | 0.257 | 0.490 |
| Female | -0.073 | 0.162 | 0.029* | -0.259 | 0.511 | 0.200 | 0.183 | -0.001* | 0.189 | 0.151 |
| Minority | -0.657 | -0.625 | -0.588 | -0.556 | -0.574 | -0.517 | -0.401 | -0.556 | -0.239 | -0.513 |
| Level | 0.039 | 0.139 | 0.143 | 0.108 | 0.105 | 0.018 | 0.118 | 0.053 | 0.024 | 0.068 |
| Year | -0.012 | 0.001* | 0.006 | 0.006 | 0.0* | 0.010 | -0.004 | 0.005 | -0.015 | 0.010 |
| N | 38,199 | 60,392 | 23,601 | 31,028 | 26,806 | 67,014 | 36,953 | 70,122 | 38,044 | 28,994 |
| Tau-a | 0.421 | 0.419 | 0.384 | 0.428 | 0.378 | 0.297 | 0.364 | 0.375 | 0.401 | 0.265 |
| C | 0.737 | 0.742 | 0.726 | 0.744 | 0.719 | 0.687 | 0.713 | 0.722 | 0.724 | 0.712 |

* Not statistically significant.

## TABLE 4

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

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

## TABLE 5

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 |
|---|---|---|---|---|---|---|
| Ability | 0.176 | 0.184 | 0.149 | 0.169 | 0.153 | 0.142 |
| | (6.50) | (7.14) | (6.13) | (6.89) | (5.91) | (4.68) |
| Experience | NS | 0.027 | 0.57 | 0.069 | 0.076 | 0.063 |
| | | (1.29) | (3.09) | (4.08) | (4.48) | (3.33) |
| Log class size | -0.445 | -0.450 | -0.401 | -0.426 | -0.441 | -0.504 |
| | (-9.38) | (-10.62) | (-10.79) | (-12.38) | (-12.85) | (-13.11) |
| Department | 3.060 | 3.257 | 2.843 | 2.762 | 2.370 | 1.958 |
| | (17.59) | (20.22) | (20.09) | (21.08) | (18.63) | (14.12) |
| N | 10,000 | 10,000 | 10,000 | 10,000 | 10,000 | 10,000 |
| Percent $\exists_I$ 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.

---

[1] Hancock noted that if the performance outcomes of students in different sized classes was indeed not class size dependent, and if the "… learning experience is not demonstrably harmed by significant increases in enrollment caps, then it is certainly harmed by not increasing them." While Hancock admits that outcomes may be a function of size in some disciplines beyond statistics courses (the data Hancock used), he is properly concerned about expending resources in staffing unnecessary sections throughout higher education.

[2] We find concerns about graduation rates and the average time-to-degree performance of universities (NYS Executive Budget, 2005-06), the increasing use of part-time and non-tenure track faculty (Ehrenberg, 2004),

---

increasing tuition, fees, and corporate sponsorship (Rizzo, 2004).

[3] 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.

[4] 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 are typical, it may also decline per student as the numbers of students per class 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 diminution in outcomes equals the rate of diminution in per student costs. We leave this for future research.

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

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

[7] The most extensive experiment was Tennessee's STAR project (Word, Achilles, Bain, Folger, Johnston and Lintz, 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, Hedges, and Konstantopoulos, 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, Smith, Zahorik, Palmer, Halbach and Ehrle, 1999;Weiss, 1990) heavily influenced California's decision to spend 6 billion dollars on class size reduction (Santa Barbara, 2001).

[8] 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, Krop, Gill and Reichardt, 1999). In view of current total spending on K-12 education nationwide of $655 billion in 1998 and over $790 billion in 2002 (U.S. Census Bureau, 2005), this seems modest. 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, Brewer, Gamoran, Willms and Zorpette (2001), Gursky (1998), Hanushek and Taylor (1989), Hoff (1998), Mosteller (1999).

[9] Using another perspective, Lesser and Ferrand, 1998 reviewed student opinion and eliminated class size as a factor affecting the student's perception of instruction, attributing observed variations to majors, faculty ability and student preparation. McKeachie, 1999 gives further references on class size research differentiating among learning methods, types of material and student motivation

[10] The model represented by Equation (1) was estimated via the logistic procedure in SAS, version 9.0. Initially, the model was developed using one fifth of the data. A full specification of Equation (1) including a large number of proxies for several

---

variables and polynomials in experience, ability and class size, was estimated. We also tested a number of demographic variables such as race, EOP, talent level, registration as a degree seeker, and county of residence. Other variables explored included faculty rank, a variable for majors(s), whether the course was a laboratory course, and whether the course had a discussion section and used teaching assistants. The model was then simplified using statistical tests for the significance of explanatory variables and tests for multicollinearity. 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, variants of the model given by Equation (1) were estimated using the full dataset of 672,489 observations, and various sub-datasets as explained.

11 The inter quartile range for class size lies between 25 and 144 students. The inter quartile range for grades lies between 2.78 (a B minus) to 3.47 (approximately half way between a B plus and an A minus.) These inter quartile ranges eliminate the more extreme class sizes and grades. Observations at the upper and lower ranges of independent variables can have a large impact on non-intercept coefficients serving as leverage points. For modest sized data sets, tests of the influence of an individual observation can be run. The use of inter quartile ranges can serve as a method of moderating the influence of possible leverage observations where testing is impractical (e. g. for very large data sets.)

[12] It might be argued that given the large number of observations most relevant variables will be statistically significant. But we tested models with variables for lab courses, for courses with discussion sections, courses with and without teaching assistants, courses that are and are not team taught, etc. none of which were statistically signifigant.

[13] The model of Table 2 is the same as the model for Table 1 and the overall statistics are consistent with those of Table 1. Detailed results are available from the author upon request.

[14] Computer memory limitations imposed this constraint so a random sample of students was chosen for testing.