The Impact of Class Size on Student Success: The Importance of Controlling for Instructor and Course Characteristics

Paper initially prepared for presentation at the 41st Annual Conference of the Research and Planning Group, Track 1 – Student Learning Outcomes and Success, Santa Barbara, CA, April 30-May 2, 2003

ABSTRACT

Recent policy debates have focused on class size as an explanatory variable that may, or may not, influence student achievement. This research paper reviews the literature on class size, and examines the impact of class size on successful course completion when controlling for instructor and class specific variables, the gender and ethnicity of a student, the reading abilities of students, the time a course is offered, and the length of time an instructor has been teaching. Results indicate that across the sections taught by the same instructor, class size does not have a statistically significant impact on successful course completion. Students are as likely to earn a successful grade in a 20-seat section as an 80-seat section. The most important explanatory variables for student success are whether the student appeared to be "transfer directed" (a full-time student), and whether the student had sufficient reading skills. White students performed slightly better in the courses, raising questions about the possibility of latent ethnic biases on the part of the instructor, but further research is needed. The findings suggest that for a narrow subset of classes in one social science discipline (taught by one instructor), class size has little meaning for explanations of why some students are more successful than others.

Matt Wetstein, Interim Dean
Planning, Research, and Grants Development
San Joaquin Delta College
5151 Pacific Avenue
Stockton, CA 95207
209-954-5472
mwetstein@deltacollege.edu

Francisco Mora
Research Analyst
San Joaquin Delta College
5151 Pacific Avenue
Stockton, CA 95207
209-954-5472
fmora@deltacollege.edu

INTRODUCTION

Public policy debates over class size reduction in the K-12 educational system have stimulated a growing body of research on the connection between class size and student achievement. That scholarship presents a mixed bag of findings. In this study, we seek to examine the inconclusive findings in the literature, and we analyze the impact of class size on student achievement in an isolated community college context. Our research controls for instructor specific effects and finds almost no evidence of class size effects on a simple measure of student achievement. In the pages that follow, we summarize the literature, set out our research model, and present new results that stress the importance of controlling for instructor specific variables when discussing class size effects.

LITERATURE REVIEW

While much has been written about class size and student achievement, there is no scholarly consensus on the issue. However, more scholars than not suggest a direct connection between the two variables. In a meta-analysis of research up to the early 1980s, Glass et al. (1982) maintained that the evidence suggested small classes were associated with higher levels of achievement across all grades at the primary and secondary level. The connection between improved learning and class size became particularly important when class sizes were reduced below 20 students (see Glass et al. 1982; Pritchard 1999).

Some of the most compelling evidence on the connection between class size and student achievement has come from Tennessee's experiment with class size reduction, and the systematic tracking of student performance after the initiation of that program in 1985 (Finn and Achilles 1999; Pritchard 1999). The Tennessee study involved 79 schools, more than 7,000 students, and a random assignment process to control for school level and curricular effects. The

study found that primary school students in smaller classes consistently outperformed their counterparts in large classes on standardized exams; that the impact was even larger for minority students in early stages of the experiment; and that the impact of small classes in early primary grades had a lasting impact that persisted beyond five years (Pritchard 1999, 4; Nye, Hedges, and Konstantopolos 1999, 127). Krueger and Whitmore (2001) discovered that the impact of the Tennessee program could even be traced to ACT and SAT college entrance test taking patterns and student scores, with more noticeable increases for black students.

California's experiment with reduced class sizes began in 1997, with an investment of $1 billion targeted at reducing class sizes to 20 or fewer at the first grade level, and in subsequent grades in later years (CSR 2002). The California initiative has generated some data suggesting a modest positive relationship between smaller primary classes and standardized test scores, however implementation issues like the inability to hire enough qualified teachers may threaten greater success (CSR 2000, 2-3; CSR 2002). Recent budget problems in the state make it unlikely that the California experiment with class size reduction will continue at the same level.

While studies have documented success stories, other research has concluded that smaller classes have only a small effect on student learning, or no proven influence at all (Pritchard 1999, 2). For example, Eric Hanushek (1998) has argued that the evidence drawing connections between class size and student achievement is "meager and unconvincing" and that analysis of 25 years of aggregate data on pupil-teacher ratios suggests that while student teacher ratios declined dramatically between 1970 and the 1990s, there were no corollary increases in achievement scores on the National Assessment of Education Progress (NAEP, see Hanushek 1998, 1-9). Hanushek (1998, 1) believes that policymakers have rushed to embrace supportive findings in this field because class size reduction is a simple fix for public school problems. He maintains that micro-level variables like good teachers in specific class settings with specific cohorts of students are more likely to generate direct connections between class size and

improved learning. In his view, when teacher quality is introduced as a control variable, very little variance in student achievement is explained by class size (Hanushek 1998, 35).

Despite the growing body of scholarship suggesting a link between K-12 class sizes and student performance, the literature on universities and community colleges is quite sparse. A search of the Educational Resources Information Center (ERIC) database for articles and reports listing "class size" in the title and "college" as a key word turned up only 15 documents between 1995 and the present. Within that group, one study notes that recent findings point to factors other than class size as being important to student learning at the college level (Gilbert 1995). Factors like teaching effectiveness, instructor practices, and course organization tend to be more important than class size for higher rates of learning. A similar conclusion emerges from a study focused on community college students and their opinions about instruction (Lesser and Ferrand 2000). These researchers discovered that class size, grades given, and college and field rankings had no significant correlations with student opinions about the quality of instruction. In an unpublished study at Fullerton College (2001), Institutional Researcher Ken Meehan found no correlation between class size and rates of student retention and success.

Two recent studies have examined the influence of class size on student achievement in college level introductory courses (see Kennedy and Siegfried 1997; Borden and Burton 1999). Kennedy and Siegfried (1997) examined whether class size had an impact on achievement levels in introductory economics courses. Using a standardized national test from 1988-89, they found that class size had no significant association with achievement scores in their economics courses, even when controlling for student characteristics like SAT scores. In the Borden and Burton study (1999), five years of data from introductory Math and Sociology courses were used to examine whether course completion and course grades were influenced by class size. They found only a "small, overall negative impact of increasing class size on student grades and course completion rates" (Borden and Burton 1999, 1).

Our study attempts to contribute to the literature on class size and student performance by providing a micro-level analysis that controls for some course specific factors in a community college setting. Most notably, the study controls for possible teacher impacts by isolating the analysis on one specific instructor teaching a limited set of courses over a five-year period. By using one instructor's courses, we can control for variables that Hanushek (1998) and others have identified as significant: variation in teaching quality, presuming that all things being equal, an individual brings relatively constant levels of "quality" to the various section she/he teaches. Thus, an analysis of one instructor can hold constant this potential variable. Additionally, the study aims to examine whether certain course characteristics like class size, time of the course offering, and a time trend variable might have an impact on student achievement.

Data and Variables

The research proposes a multivariate logistic regression equation that explores the impact of various independent variables on successful course completion in introductory Political Science courses at a California community college. While class size is the variable of most interest to the study, other factors control for the demographic features of students (to explore potential ethnic and gender differences), the nature of a course (in terms of U.S. Government versus other Political Science topics), time of day the course was offered, and a time trend variable that attempts to assess whether the instructor has become "tougher" with age. Two control variables are also built into the study for student skill levels and transfer intention.

The study focuses on one professor's courses over a five-year period. In this case one of the authors of this study served as a "guinea pig" for analysis. The period under review begins in 1996 with the start of the author's teaching tenure at San Joaquin Delta College to the spring of 2001 (1996-2001). Complete data are available on 1,578 students who enrolled in the classes during that timeframe, and they represent the population studied.

In order to assess grading patterns over time, and to test the hypothesis that a particular instructor might become a "tougher grader" over time, a count variable was included in the study for each successive semester. Thus, courses in the first semester were scored "1," with each successive semester increasing in value by one. This particular variable captures the consistency (or inconsistency) of the instructor's grading method. If the coefficient is positive, it means that with time the instructor is becoming an easier grader. Conversely, if the coefficient is negative, it implies that the instructor is becoming a tougher grader over time.

Ethnicity is a variable intended to capture which ethnic groups do better in the author's classes. It could also imply whether or not the instructor favors certain ethnic groups by giving students from that group higher grades. Ethnicity is measured by a variable identifying whether the student is white (a score of 1), or non-white/unknown race (0). If the coefficient comes out positive it means that the likelihood of successful completion is positively associated with being white. Conversely, if the coefficient is negative the likelihood of success decreases as a result of being non-white. Moreover, gender is used in the analysis to see if higher percentages of men or women are successful in completing a Political Science course, and to explore the possibility that the instructor may have certain sexist biases in one direction.

A variable measuring the time of day a course was taught is included in the model. This variable is designed to capture any impacts that might come from taking a class in the afternoon, as opposed to a morning section. The hypothesis behind this measure is that students taking a course later in the day might have a drop off in their scores compared against morning students, essentially because of "study fatigue" and a loss of alertness that comes from the passing of the day. In short, the expectation was a negative relationship between successful course completion and a variable measuring time of day. Scores on this indicator range between 7.50 (for a 7:30 am class) and 13.50 for a 1:30 pm class).

Two key variables to include in any study are indicators of student skill level and what we call “transfer directed behavior.” We use reading assessment test scores as a crude proxy for academic skill levels, assuming that stronger readers have an obvious advantage of over weaker readers. Thus, we expect a strong positive relationship between reading scores and successful completion of a course. Similarly, we examine course-taking patterns of students during the semester of their enrollment in the Political Science courses. By tracking the number of units enrolled, we are able to determine whether a student is “transfer directed” and taking a full time load of 15 units or more, or whether the student is taking only a part time load. We hypothesize that transfer directed students would have greater odds of completing the Political Science courses, simply because their course taking patterns suggest a seriousness of purpose about completing the college’s general education requirements, and necessarily, a greater probability of transfer in subsequent years.

Of course the variable of most interest is the measure of class size and how it might affect successful course completion. Given the findings in some of the literature, the expectation was that larger class sizes would produce lower probabilities of student success. Thus, we expected to find a statistically significant relationship between class size and course completion.

Model

Using successful course completion as a dichotomous dependent variable, the logistic regression model is defined as:

Success = Intercept +
b1 (Semester Count Variable) +
b2 (U.S. Government Course) +
b3 (Gender) +
b4 (White/Non-White) +
b5 (Class size) +
b6 (Units Attempted) +
b7 (Start Time of Class) +
b8 (Reading Assessment Score) +
Error

The variables in the model are listed in Table 1, with their mean, minimum, and maximum values. The mean value for successful completion (found at the bottom of the table) was .51, suggesting that 51 percent of the students were able to complete the courses with a grade of A, B, C, or CR.

The statistical model used in this study is derived from a logistic regression analysis in the SPSS software package. Logistic regression was used because of the dichotomous nature of the dependent variable (see Aldrich and Nelson 1984).

Table 1

Mean, Minimum, and Maximum Values for Variables in the Study

Variable	Mean	Min	Max
Semester Count Variable	5.26	1.0	10.0
U.S. Govt. Class	.90	0.0 (no)	1.0 (yes)
Gender	.49	0.0 (female)	1.0 (male)
White/Non-White	.49	0.0 (non-white/other)	1.0 (white)
Class Size	60.62	19.0	87.0
Units Attempted	13.31	3.0	35.0
Start Time of Class	9.85	7.5 (7:30 am)	13.50 (1:30 pm)
Reading Assessment Score	42.46	26.0	53.0
Successful Completion	.51	0.0	1.0

RESULTS

The logistic regression model resulted in some rather robust findings and model fit statistics. One useful indicator of a model's utility is to examine its ability to correctly classify the outcomes in a dichotomous data set. A model that can correctly classify cases at a much higher rate than a modal guessing strategy is better than a model that fails to improve over the same guessing strategy. In this case, employing a modal guessing strategy would result in a correct classification 51 percent of the time (by simply guessing every time that a student would be successful, one would be correct 51 percent of the time). The model employed in this study improves beyond this point to a correct classification 67 percent of the time (see the bottom of Table 2). This represents a 33% reduction in the amount of error.

TABLE 2

Logistic Regression Coefficients for Successful Course Completion in Political Science Courses Taught by the Instructor, 1996-2001

Variable	B	signif.	P if x is low	P if x is high	Change in P
Semester Count Variable (x = 1, x = 10)	.0133	.5244	.49	.52	.03
U.S. Govt. Class (x = 0, x = 1)	-.0587	.7973	.52	.50	-.02
Gender (x = 0, x = 1)	-.0479	.6666	.51	.50	-.01
White/Non-White (x = 0, x = 1)	.3441	.0023 **	.46	.54	.08
Class Size (x = 20, x = 80)	-.0003	.9388	.51	.50	-.01
Units Attempted (x = 6, x = 15)	.1033	.0000 ***	.32	.55	.23
Start Time of Class (x = 7:30 am, x = 1:30 pm)	.0427	.2565	.48	.54	.06
Reading Assessment Score (x = 26, x = 53)	.1670	.0000 ***	.06	.86	.80
Constant	-9.0068				

- 2 Log Likelihood Ratio	1577.228
Model Chi Square	291.687
Percent Successfully Classified	67.17%
Pseudo R Square Measure	.156
Proportional Reduction in Error	33.0%

* p < .05, ** p <.01, *** p< .001

The coefficients in Table 2 reveal that three of the variables in the study had statistically significant impacts on successful course completion. Those variables include reading level (b = .167, significant at .000), units attempted (b = .1033, significant at .000), and the indicator of white or non-white students (b = .344, significant at .01). The impact of the first two variables meshes with the expected hypotheses. That is, students scoring higher on the college's ASSET reading assessment test were significantly more likely to pass Political Science courses than those with lower scores. Similarly, transfer-directed students who were enrolled in a full complement of classes were more likely to successfully complete the courses than their part-time counterparts. This meets with the expectation that full-time students might bring more serious attention to their studies, perhaps because they are less prone to work part-time or have other life commitments that impinge on study time.

One of the surprising findings (for the author) was the finding of a statistically significant impact for the white versus non-white variable. The positive coefficient suggests that white students earn higher grades than non-white students, even when controlling for other variables in the equation like reading levels and unit loads. This suggests one of two possibilities: either the white/non-white variable is correlated with some unmeasured variables that are omitted from the equation that would truly explain this gap (class attendance, study time, time reading, study group behavior, outside work commitments), or it may be possible that the coefficient unmasks some latent ethnic biases that the instructor applies to student work unconsciously. Further additions to the theoretical model might examine this problem to see if there are other variables that drive this difference rather than ethnic differences.

Perhaps the most significant non-finding in the equation is the paltry effect that the class size variable had on successful course completion. The negative coefficient suggests a relationship in the expected direction (b = -.0003), but the magnitude of the impact is quite small.

The relationship suggests that as class size increases, probabilities of successful completion are only marginally smaller. Indeed, in relation to the other variables in the equation, class size has the weakest impact of all measures. This does not provide overwhelming evidence of a connection between class size and successful course completion in Political Science courses.

Logistic regression coefficients are difficult to interpret because of the curvilinear relationship that is implied in the model (as opposed to linear relationships assumed in most social science models). The equation is "fitted" to take on the shape of an "S" curve in a logistic regression equation, implying that the slope of the relationship between the independent variables and the dependent variable changes in magnitude along the possible range of the dependent variable (see Aldrich and Nelson 1984). To make the coefficients more understandable, a logistic calculation can be used to transform the effects into an odds ratio, based on a specified level of an independent variable. These transformations can be noted in the last three columns of Table 2, which feature the "probability" of an occurrence when a variable (x) is set at a low value, the probability when x is set high, and the change in probability when moving from a low to a high level of a variable. These impacts are calculated while controlling for all other variables in the equation. In essence, each variable is set at a mean value, while one independent variable is set at a low and high value to determine the change in odds of an outcome. In this case, the probability scores tell one the odds of a successful course completion in Political Science.

The probability calculations allow one to inspect not only the change in probability, but also the relative magnitude of each variable in the equation. Note for example the scores for the first variable listed in Table 2 (Semester Count). The logistic transformations suggest that a student enrolled in a Political Science course in fall 1996 (semester 1 in the study) had a 49 percent chance of successful completion. A student in semester 10 – controlling for all other variables – had a 52 percent chance of success. The difference between the two time frames is

quite low, and clearly not large enough to be statistically significant. The data also suggest no real "softening" or "hardening" in grading standards by the instructor over a five year period. Contrast that minimal change in probability with the impact measured for the last variable in Table 2 (Reading Assessment Score). A student scoring the lowest on the ASSET reading test had only a six percent chance of success in the Political Science courses taught by the instructor. Meanwhile, students earning the highest reading score had an 86 percent chance of successful completion. This 80-point gap represents the largest swing in the odds for the variables in the equation, and far outweighs the next strongest variable (Units Attempted).

The insignificance of class size can be seen in the probability calculations. When the class size value is set at 20 students (the smallest a Political Science course would likely ever be at the college), the probability of successful completion was 51 percent. When the class size is set at 80, the probability of successful completion only fell to 50 percent. This is a clear confirmation that when considering the impact of more relevant variables, class size had little meaning for explaining course completion patterns in one instructor's Political Science offerings.

Students taking the introductory U.S. government course did only marginally worse than students in other courses (International Relations and Introduction to Political Science). The difference in probabilities (.50 and .52) was not large enough to achieve statistical significance. Students taking classes later in the day did marginally better than students who enrolled in earlier sections, but the difference was not statistically significant. Likewise, there was virtually no difference between the odds of success for female and male students.

CONCLUSION

The results from this study explain a little more than fifteen percent of the variance in successful course completion in Political Science courses taught by one instructor. While reading levels and course loads were strong predictors of success, other factors such as the time spent studying and class attendance perhaps would help explain a greater extent of the variance in course completion. Surprisingly, grade differences between white and non-white students appeared, with white students having a higher probability of successful completion in the instructor's courses. Perhaps additional variables in the theoretical model would eliminate this surprising finding.

If class size is an institutional factor that can influence student performance levels, there is little in this study to suggest it has an impact when controlling for other variables. Indeed, class size mattered little in the likelihood of success for students enrolled in the particular courses taught by one instructor. This methodological approach allows one to control for specific teaching techniques and approaches brought to the classroom, since the instructor essentially teaches in the same style regardless of class section.

The results do not fit with the literature on primary school class size reductions. There is no apparent connection between class size and student achievement when controlling for instructor and class specific variables in the community college setting analyzed here. Yet there is the lingering question of whether these findings can be generalized beyond the instructor studied. The next stage of research should attempt to expand the analysis beyond one instructor in one field of study. The results would be more impressive if they were found to hold across a wide range of individual instructors in a wide range of academic and vocational fields at the same college. The importance of the methodology is that it has the potential to control for instructor specific effects while examining class size as a variable that influences student achievement.

REFERENCES

Aldrich, John H., and Forrest D. Nelson. 1984. *Linear Probability, Logit, and Probit Models.* Sage University Paper Series: Quantitative Applications in the Social Sciences, Number 07-045. Beverly Hills: Sage Publications.

Borden, Victor M. H., and Kathy L. Burton. 1999. The Impact of Class Size on Student Performance in Introductory Courses. Paper presented at the annual Forum of the Association for Institutional Research, Seattle, WA, May 30-June 3, 1999.

CSR Research Consortium. 2002. Class Size Reduction in California (Fact Sheet). Palo Alto: CRS, June 2002.

_____. 2000. Class Size Reduction in California: The 1998-99 Evaluation Findings. Palo Alto: CSR, 2000. Available online at: http://www.classize.org/summary/98-99/

Finn, Jeremy D., and Charles M. Achilles. 1999. Tennessee's Class Size Study: Findings, Implications, and Misconceptions. *Educational Evaluation and Policy Analysis* 21 (2): 97-100.

Fullerton College Faculty Senate. 2001. Approved Minutes of the Faculty Senate, October 18, 2001. Fullerton, CA: Fullerton College. Available online at: http://academicsenate.fullcoll.edu/

Gilbert, Sid. 1995. Quality Education: Does Class Size Matter? *CSSHE Professional File* 14 (Winter): 1-7.

Glass, Gene V., Leonard S. Cahen, Mary L. Smith, and Nikola N. Filby. 1982. *School Class Size: Research and Policy*. Beverly Hills, CA: Sage Publications.

Hanushek, Eric A. 1998. The Evidence on Class Size. W. Allen Wallis Institute of Political Economy Occasional Paper Number 98-1, University of Rochester, February 1998.

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 (4: October): 385-394.

Kreuger, Alan B., and Diane M. Whitmore. 2001. Would Smaller Classes Help Close the Black-White Achievement Gap? Princeton University Industrial Relations Section Working Paper #451. Available online at: http://www.irs.princeton.edu/pubs/working_papers.html

Lesser, Diane, and Judy Ferrand. 2000. Effect of Class Size, Grades Given, and Academic Field on Student Opinion of Instruction. *Community College Journal of Research and Practice* 24 (4): 269-277.

Nye, Barbara, Larry V. Hedges, and Spyros Konstantopulos. 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.

Pritchard, Ivor. 1999. *Reducing Class Size: What Do We Know?* Washington DC: US Department of Education, Office of Educational Research and Improvement, March 1999. Available online at: http://www.ed.gov/pubs/ReducingClass/Class_size.html