

The effects of class size on student grades at a public university

Edward C. Kokkelenberg^{a,b,*}, Michael Dillon^{a,b}, Sean M. Christy^{a,b}

^a*Department of Economics, Binghamton University, Binghamton, NY 13902, USA*

^b*School of Industrial and Labor Relations, Cornell University, Ithaca, NY 14853, USA*

Received 10 March 2005; accepted 1 September 2006

Abstract

We model how class size affects the grade higher education students earn and we test the model using an ordinal logit with and without fixed effects on over 760,000 undergraduate observations from a northeastern public university. We find that class size negatively affects grades for a variety of specifications and subsets of the data, as well as for the whole data set from this school. The specifications tested hold constant for academic department, peer effects (relative ability in class), student ability, level of student, level of course, gender, minority status, and other factors. Average grade point declines as class size increases, precipitously up to class sizes of twenty, and more gradually but monotonically through larger class sizes. The evidence is that this is not exclusively a small class effect. 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. Future studies of economies of scale in higher education need to address the traditional assumption of constant product quality.

© 2007 Elsevier Ltd. All rights reserved.

JEL classification: I21

Keywords: Economies of scale; Educational economics; Student performance; Logit analysis; Fixed effects models

1. 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 student outcomes are adversely affected by larger

classes, 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

*Corresponding author. Tel.: +1 607 254 4744; +1 607 777 2550; fax: +1 607 255 4496.

E-mail address: eck4@cornell.edu (E.C. Kokkelenberg).

¹Hancock 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.

the quality of the incoming students appears to be rising or steady.² Schools often look to spreading the costs of a faculty over more students by increasing class sizes or by increasing workload (number of courses taught per term). This latter method is resisted by faculty senates, unions, and often trustees, leaving 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 Montagna (2002) and Moore (2003).

But the questions 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, Cahen, Smith, and Filby (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. 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.

2. Background

2.1. K-12 studies

By the 1970s 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 1970s (Glass, McGaw, & Smith, 1981; Glass & Smith, 1979; Smith & Glass, 1980) 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 & 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 et al. (1982) findings.⁷ In 2003, a number of

³Student/pupil ratios in K-12 schools had been dropping since the 1950s without any marked increase 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 effect 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.

⁴The 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 number 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.

⁵Lipman (1990), Kennedy and Siegfried (1996, 1997).

⁶Heavily weighting studies that they considered more experimental in design, and discounted 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.

⁷The most extensive experiment was Tennessee's STAR project (Ritter & Boruch, 1999; Word et al., 1990). 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 & 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, & Konstantopoulos, 1999) and that small classes seem to be most beneficial to those coming from disadvantaged backgrounds (Krueger & Whitmore, 2001; Slavin, 1990). Subsequently, the findings from the STAR program and more modest experiments elsewhere (Tillitski,

²Thus, 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) and increasing tuition fees and corporate sponsorship (Rizzo, 2004).

articles appeared in a special editions 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 [Dustman, 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 & Hevert, 1992](#); [Maxwell & Lopus, 1995](#); [Hanushek, 2003](#)).⁸ Further methodological challenges have weakened these claims ([Maasoumi, Millimet, & Rangaprasad, 2003](#); [Krueger, 2003](#)).

2.2. Higher education

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 using certain tests, class size matters in some circumstances. No such agreement

(footnote continued)

1990; [Molnar et al., 1999](#); [Weiss, 1990](#)) heavily influenced California's decision to spend 6 billion dollars on class size reduction ([Santa Barbara, 2001](#)).

⁸The 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, & 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 in 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 \(1998\)](#), and [Mosteller \(1999\)](#).

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, & Jensen, 1985](#); [Pascarella & 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, 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 ([Bolander, 1973](#); [Feldman, 1984](#); [McConnell & 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 Montagna \(2002\)](#) and [Kwantlen University \(2004\)](#). Toth and Montagna summarize eight studies from 1990 to 1999 which 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 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, p. 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). The dependent variables are many, 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 Gershengberg (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 postgraduation. 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 24 semesters and 45 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.

3. 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 i th student in the j th class during period t ,

$$W_{ijt} = b_0 + \phi(E_{it})'\beta + \theta(A_{it})'\Gamma + Z'_{it}\gamma + V'_{it}\kappa. \quad (1)$$

Here, W represents the grade, E the i th student's experience (e.g. level in college), A represents ability,

⁹Using 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.

Z a vector of student related variables, and V is a vector of environmental, faculty, and subject matter factors including class size (CS). $\phi(E)$ and $\theta(A)$ are allowed to be polynomials in E and A , and β, Γ, λ , and κ 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.

4. 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.

5. Results¹⁰

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, (\bar{W}_C), the departmental mean grade (\bar{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

$$\begin{aligned} W = & b_0 + \beta_1 A_r + \beta_2 A_r^2 + \beta_3 \bar{W}_C + \beta_4 \bar{W}_D + \beta_5 CS \\ & + \beta_6 CS^2 + \beta_7 SATM + \beta_8 F + \beta_9 SATV \\ & + \beta_{10} G + \beta_{11} M + \beta_{12} L + \beta_{13} Y \\ & + \beta_{14} AP + \beta_{15} CS * G. \end{aligned} \quad (2)$$

The results of estimating an ordered logistic of Eq. (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

(footnote continued)

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 Eq. (1) were estimated using the full dataset of 672,489 observations, and various sub-datasets as explained.

¹¹Inter 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 observation can be run. The use of inter quartile ranges can serve as a test for very large data sets.

¹⁰The model represented by Eq. (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 Eq. (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,

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 (\bar{W}_c)	1.376	0.005	2.762	0.009	1.448	0.008
Dept. mean grade (\bar{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	1.6×10^{-7}	0.00006	3.4×10^{-6}	0.00002	2.9×10^{-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	
$P_R > \text{Chi squared}$	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.

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, 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. Minority students do less well than non-minorities, females and those with high SAT scores do better 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, tradi-

tions, and the material presented in class, as [McKeachie \(1999\)](#) 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, CS², are consistent with the concave model suggested by [Glass et al. \(1982\)](#). 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, faculty and for other classroom environmental variables. Further, the standard errors on the class size terms are small (0.00022, 0.000102, and 0.000038, respectively). The coefficients at plus or minus two standard errors for class size thus range from −0.00201 to −0.00193, −0.00496 to −0.00456, and −0.00231 to −0.00215; all negative and narrow ranges.

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

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 (100%)	−0.00700 (0.00007)	0.000015 (1.64×10^{-7})
Inter			
90%	611,330 (90.9%)	−0.00851 (0.00012)	0.000024 (3.85×10^{-7})
80%	543,965 (80.9%)	−0.00955 (0.00017)	0.000032 (6.64×10^{-7})
Quartile	342,289 (50.9%)	−0.0123 (0.00054)	0.00006 (3.42×10^{-6})
For all classes greater than			
5 students	657,253 (97.7%)	−0.00566 (0.00007)	0.000012 (1.65×10^{-7})
10 students	642,250 (95.5%)	−0.00507 (0.00007)	0.000011 (1.67×10^{-7})
15 students	617,071 (91.8%)	−0.00455 (0.00007)	9.57×10^{-6} (1.69×10^{-7})
20 students	583,815 (86.8%)	−0.00412 (0.00007)	8.69×10^{-6} (1.73×10^{-7})
For all classes less than			
500 students	672,010 (99.9%)	−0.00717 (0.00007)	0.000015 (1.68×10^{-7})
400 students	655,164 (97.4%)	−0.00902 (0.00008)	0.000022 (2.43×10^{-7})
300 students	622,365 (92.5%)	−0.0145 (0.00013)	0.000045 (4.72×10^{-7})
200 students	530,056 (78.8%)	−0.0238 (0.00021)	0.000101 (1.04×10^{-6})

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, *ceteris paribus*.¹²

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% and the second consisting of 276,327 observations where students' GPA was in the lower 40% of the GPA distribution. The model was that of Eq. (2). The coefficients on class size and the square of class size from the ensuing logistics regressions were comparable to those in column one 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 *ceteris paribus*, and students appear not to select into courses based upon their GPA and the anonymity available in large classes.

The analysis, using an abbreviated model, was next extended to 10 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 Eq. (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. These results are available from the authors upon request.

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. Figs. 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, *ceteris 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.

¹²The 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.

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.

6. 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

$$W_{ijt} = \beta_0 + \beta_i + \beta_t + \beta_1 E_{it} + \beta_2 A_{it} + \beta_3 CS_j + \beta_4 \bar{W}_{Dj}. \quad (3)$$

Here β_i is the student fixed effect and β_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 is 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 Eq. (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 0.0001.

Next, we relaxed the assumption of proportional odds and we estimated a binary fix effects model of Eq. (3) for a random sample of 10,000 students chosen from the 167,928 observations.¹³ 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,

¹³Computer memory limitations imposed this constraint so a random sample of students was chosen.

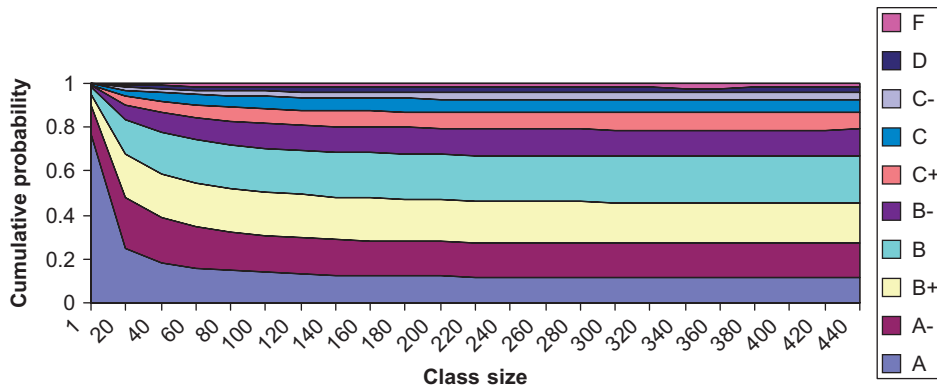


Fig. 1. Cumulative probability of grades received vs. class size.

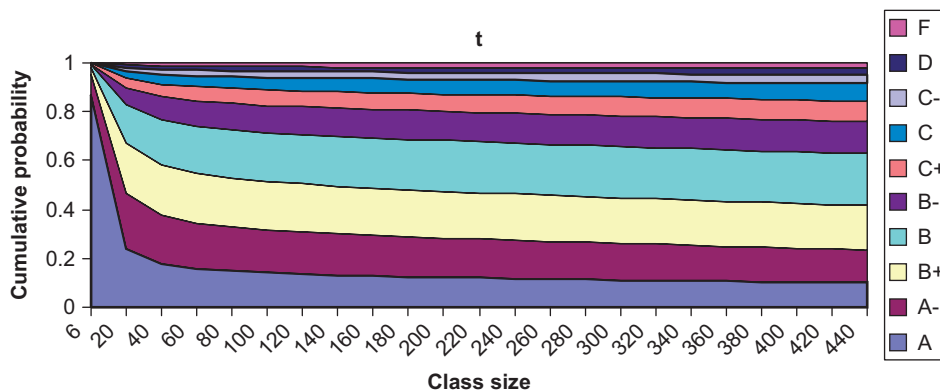


Fig. 2. Cumulative probability of grades received vs. class size for classes six or greater.

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 1–3.

7. 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.

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

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	625,261	625,261
Intercept only		
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.

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.

Acknowledgments

We wish to thank Ronald Ehrenberg, Jack Keil, Peter Partell and Corine Prost for helpful comments. We would also like to thank Jessica Richards, Yu Zhu and Hester Han who have brought considerable analytic expertise to the project. 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 and this list is available from the authors upon request. The overall grade distribution is listed in Table A1. Further statistics are also available upon request.

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	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 β 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.

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 an 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 freshmen, 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: Individual's 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 A1
Letter grade distribution

Letter grade	Frequency	Percent
A ^a	200,705	22.88
A ^{−a}	131,627	15.00
AU	792	0.09
B ^a	109,172	12.44
B ⁺ ^a	114,740	13.08
B ^{−a}	63,703	7.26
C ^a	41,648	4.74
C ⁺ ^a	43,146	4.92
C ^{−a}	23,121	2.64
D ^a	19,851	2.26
F ^a	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	8160	0.93
WF ^a	63	0.01
WP	346	0.04
X	23	0.00
Total	877,294	100.00
Missing	111,604	

^aUsed 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% of the unusable grades.

Appendix B

Class size effects in introductory mathematics and economics courses:

Department course	Number of observations	Class size Q1	Class size Q2 (median)	Class size Q3	Class size coefficient (Std. er)
Mathematics					
Introductory statistics	5882	225	309	374	−0.00401 (0.00094)
Calculus I	6054	31	56	304	−0.00065 (0.00016)
Economics					
Poverty & discrm'n.	2976	61	98	121	−0.00790 (0.00426)
Introductory micro	9205	142	206	224	−0.00786 (0.00298)
Introductory macro	7391	99	143	217	−0.00866 (0.00161)

References

- Akerhielm, K. (1995). Does class size matter? *Economics of Education Review*, 14(3,9), 229–241.
- Bolander, S. F. (1973). Class size and levels of student motivation. *Journal of Experimental Design*, 42(2), 12–18.
- Borden, V. M. H., & Burton, K. L. (1999). The impact of class size on student performance in introductory courses. In Paper presented at the 39th annual conference of the AIR in Seattle, WA.
- Bressoux, P., Kramarz, F., & Prost, C. (2005). *Teachers' training, class size and students' outcomes*. Mimeo, February.
- Brewer, D., Krop, C., Gill, B. P., & Reichardt, R. (1999). Estimating the cost of national class size reductions under different policy alternatives. *Educational Evaluation and Policy Analysis*, 21(2), 179–192.
- Correa, H. (1993). An economic analysis of class size and achievement in education. *Education Economics*, 1(2), 129–135.
- Dustman, C. (2003). The class size debate and educational mechanisms, editorial. *The Economic Journal*, 113, F1–F2.
- Ehrenberg, R. G., Brewer, D. J., Gamoran, A., Willms, J. D., & Zorpette, G. (2001). Does class size matter? *Scientific American*, 285(5), 78–86.
- Ehrenberg, R. G. (2004). *Key issues currently facing American higher education*. Mimeo, July.
- Feldman, K. A. (1984). Class size and college students' evaluations of teachers and courses, a closer look. *Research in Higher Education*, 21(1), 45–91.
- Finn, J. D., & Achilles, C. M. (1990). Answers and questions about class size, a statewide experiment. *American Educational Research Journal*, 27(3), 557–577.
- Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study, findings, implications, misconceptions. *Educational Evaluation and Policy Analysis*, 21(2), 97–110.
- Glass, G. V., Cahen, L. S., Smith, M. L., & Filby, N. N. (1982). *School class size, research and policy*. Beverly Hills: Sage Publications.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills: Sage Publications.
- Glass, G. V., & Smith, M. L. (1979). Meta-analysis of research on the relationship of class-size and achievement. *Evaluation and Policy Analysis*, 1, 2–16.
- Gursky, D. (1998). Class size does matter. *The Education Digest*, 64, 15–18.
- Hancock, T. M. (1996). Effects of class size on college student achievement. *College Student Journal*, 30(4), 479–481.
- Hanushek, E. A. (1995). Moving beyond spending fetishes. *Educational Leadership*, 53(3), 60–64.
- Hanushek, E. A. (1996). Outcomes, costs, and incentives in schools. *Improving America's schools: The role of incentives*. Washington: National Academy Press.
- Hanushek, E. 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, E. A. (1999b). The evidence on class size. In S. E. Mayer, & P. E. Peterson (Eds.), *Earning and learning, how schools matter* (pp. 131–168). Washington: Brookings Institution Press.
- Hanushek, E. A. (2003). The failure of input-based schooling policies. *Economic Journal*, 113, F64–F98.
- Hanushek, E. A., & Taylor, L. L. (1989). Alternative assessments of the performance of schools, measurement of state variations in achievement. *The Journal of Human Resources*, 25(2), 179–200.
- Harder, H. (1990). A critical look at reduced class size. *Contemporary Education*, 62(1), 28–30.
- Hoff, D. J. (1998). Federal class-size reports do an about-face. *Education Week*, 17(9 June), 10–22.
- Hoxby, C. M. (2000). The effects of class size on student achievement, new evidence from population variation. *The Quarterly Journal of Economics*, 11, 1239–1285.
- Kennedy, P. E., & Siegfried, J. J. (1996). On the optimality of unequal class sizes. *Economics Letters*, 50, 299–304.
- Kennedy, P. E., & Siegfried, J. J. (1997). Class size and achievement in introductory economics, evidence from the TUCE III data. *Economics of Education Review*, 16, 385–394.
- Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics*, 5, 497–531.
- Krueger, A. B. (2003). Economic consideration and class size. *The Economic Journal*, 113(February), F34H.
- Krueger, A. B., & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: evidence from project star. *Economic Journal*, 111(468), 1–28.
- 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, D., & Ferrand, J. (1998). *Effects of class sizes, grades given, and academic field on student opinion of instruction*. Mimeo, War-Wic Community College.
- Lipman, S. A. (1990). On the optimality of unequal class sizes. *Economics Letters*, 33, 193–196.
- Maasoumi, E., Millimet, D. L., & Rangaprasad, V. (2003). *Robust inference concerning the impact of class size on student achievement*. Mimeo, July.
- Maxwell, N. L., & Lopus, J. S. (1995). A cost effectiveness analysis of large and small classes in the university. *Educational Evaluation and Policy Analysis*, 17(2), 167–178.
- McConnell, C. R., & Sosin, K. (1984). Some determinants of student attitudes toward large classes. *Journal of Economic Education*, 181–190.
- McKeachie, W. J. (1980). Class size, large classes and multiple sections. *Academe*, (2), 24–28.
- McKeachie, W. J. (1999). *Teaching tips*. Boston: Houghton Mifflin.
- McKeachie, W. J., Iran-Nejad, A., & Berliner, D. C. (1990). The multi-source nature of learning, an introduction. *Review of Educational Research*, 60(4), 509–516.
- Mincer, J. (1974). *Schooling, experience and earnings*. New York: Columbia University Press.
- Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A., & Ehrle, K. (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, J. (2003). *Data and issues regarding increasing class size and quality of education*. Mimeo.
- Mosteller, F. (1999). How does class size relate to achievement in schools? In S. E. Mayer, & P. E. Peterson (Eds.), *Earning and learning: How schools matter* (pp. 117–129). Washington DC: Brookings Institutional Press.

- Nelson, R., & Hevert, K. T. (1992). Effect of class size on economies of scale and marginal costs in higher education. *Applied Economics*, 24, 473–482.
- Nye, B., Hedges, L. V., & Konstanopoulos, S. (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*, from http://publications.budget.STATE.NY.US/fy05-06_littlebook/higherEd.html, downloaded February 3, 2005.
- Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students*. San Francisco: Jossey-Bass.
- Raimondo, H. J., Esposito, L., & Gershenberg, I. (1990). Introductory class size and student performance in intermediate theory courses. *Journal of Economic Education*, 21, 369–381.
- Ritter, G. W., & Boruch, R. F. (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, M. 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). *Proposed budget*. From www.cc.santa-barbara.ca.us/cced/budget/research/proposedbudget/document/secIB.pdf, downloaded March, 2004.
- Slavin, R. (1990). Class size and student achievement, is smaller better? *Contemporary Education*, 62(1), 6–12.
- Smith, M. L., & Glass, G. V. (1980). Meta-analysis of research on class size and its relationship to attitudes and instruction. *American Educational Research Journal*, 17(4), 419–433.
- Spahn, K. (1999). *Class size and faculty effectiveness and quality*. Paper presented at the 39th Annual Institutional Research forum in Seattle, WA.
- Tillitski, C. (1990). The longitudinal effects size in prime time, Indiana's state sponsored reduced class size program. *Contemporary Education*, 62(1), 31–32.
- Toth, L. S., & Montagna, L. G. (2002). Class size and achievement in higher education, a summary of current research. *College Student Journal*, 36(2), 253–261.
- U.S. Census Bureau. (2005). Statistical abstract of the United States 2004–5, Washington, D.C.
- Weiss, T. (1990). Indiana's prime time. *Contemporary Education*, 62(1), 31–32.
- Williams, D. D., Cook, P. F., Quinn, B., & Jensen, R. P. (1985). University class size, is smaller better? *Research in Higher Education*, 23(3), 307–317.
- Word, E., Achilles, C. M., Bain, H., Folger, Johnston, J., & Lintz, N. (1990). STAR final executive summary, kindergarten through third grade results. *Contemporary Education*, 62(1), 13–16.