December 2010

Class size, type of exam and student achievement

Erik Strøjer Madsen

Department of Economics, Aarhus School of Business, Aarhus University

Abstract

Education as a road to growth has been on the political agenda in recent years and promoted not least by the institutions of higher education. At the same time the universities have been squeezed for resources for a long period and the average class size has increased as a result. However, the production technology for higher education is not well known and this study highlights the relation between class size and student achievement using a large dataset of 80.000 gradings from the Aarhus School of Business. The estimations show a large negative effect of larger classes on the grade level of students. The type of exam also has a large and significant effect on student achievements and oral exam, take-home exam and group exam reward the student with a significantly higher grade compared with an on-site written exam.

Papiret er præsenteret ved det 33. Symposium I anvendt statistik den 24.-26. januar 2011, Det Biovidenskabelige Fakultet, Københavns Universitet.

1. Introduction

In recent years the universities in Denmark have been asked to increase productivity in teaching by about 2 percent each year through budget cuts. For teaching this means larger class size through labour saving as it is difficult to make cost savings on teaching infrastructure like class rooms and other teaching facilities. While larger class size saves costs it is unclear how it affects student achievements and whether there would be any increases in productivity at all. This question has not been addressed so far for university teaching.

However, class size effects have been studied extensively in the elementary school especially in the US and for a survey of these studies see Hedges et al (1994) and Hanushek (1998). Hanushek found no consistent relationship between class size and student achievements as 35 studies report a weak positive effect of smaller classes, but 18 studies found the opposite and 32 were inclusive. This seems to be a paradox as the results would make large cost cutting possible without affecting student achievements and that is probably what triggered the large amount of empirical research on this topic in the 70thies and 80thies in the US.

One explanation of this paradox was the econometric method used in these studies and which among other problems assumed class size to be exogenous. However, there are good reasons to expect class size to be endogenous in explaining student achievements. As class size is determined by the resources allocated to schools and a lot of special teaching programs allocate more resources to classes with children who are less able to learn due to their social background or for other reasons. Thereby resource allocations establish a negative relationship between class size and student achievements, which is as likely to show up in a simple regression. This has among other researchers been pointed out in Krueger (1999, 2003), who also refereed the largely cited natural experiment known as Tennessee's Project STAR, where 11,600 students and their teachers were randomly assigned to small and regular sized classes**.** When eliminating the endogeneity of class size in this project, the small class students turn out to perform significantly better than students in regular sized classes. For a recent Danish study which also addresses the problem of endogeneity of classes in the primary school and find class size effects comparable with Kruger (1993) see Heinesen (2010).

The endogeneity of class size has nicely been explained in an equilibrium model by Lazear (2001) who views education as a public good where more students created a kind of congestion effect in classes through negative externalities. He explains the negative externalities of teaching large classes with the disruptive behaviour among the students. He assumes that the teacher has to stop teaching when a student disrupts the teaching and he further assumes student's disruption to be uncorrelated thereby

making the total amount of disruption increasing progressively with the size of classes. On the other hand, the saved salaries to teachers only increase proportionally with class size and thereby define an optimal class size where the marginal cost and benefit per student of teaching are balanced.

Even through Lazears model is quite simple it gives a lot of testable implications. On the benefit side weaker students who disrupt teaching a lot call for smaller classes and this could explain that the sizes of school classes are smaller in areas with weak students. It also explains that class sizes are much larger in universities compared to the primary schools due to much more disruptions in classes with younger students. On the cost side, higher salary for teachers would increase the equilibrium class size and this could also explain the larger teaching classes in higher education where research based teaching is much more expensive than teaching in primary schools.

While it is easy to see that teaching is disrupted by misbehaved children in primary schools and child care, there are normally very few disciplinary problems in university teaching. One would rather expect positive externalities of teaching especially in smaller classes where students could interact with each other and the teacher. When good and thoughtful students ask deep questions it will highlight the subject for everybody in the class room and it can also stimulate the teacher to advance his teaching. The teacher has also more time to focus his teaching on each individual student in smaller classes which enhance the learning in smaller classes further.

One could therefore add these positive externalities to Lazear's model which of course would increase the optimal class size. The positive learning interaction among students may increase with more students as there are more possibilities of different interactions. However, as class size increases the probability that students rise questions on class decreases and the teacher has less time to focus his teaching to each individual student and the learning outcome then decreases. The positive externalities in class room teaching therefore also point to an optimal class size beyond which the teaching efficiency would be low.

The next section presents the data and the following section discuss the estimation results. In the primary school students normally follow the same class whereas in universities the students normally switch classes depending on the topics. This makes it much more easy to estimate the class size effects and gives more reliable results too.

2. Data description

This study uses a sample of 80,000 exam gradings of master students in economics and business administration at the Aarhus School of Business in the period from 1999

to 2010. Table 1 shows descriptive statistics for the variables used excluding grading of activities not related to class teaching like thesis writing.

Table 1: Descriptive statistics of variables

	Observations	Mean	Standard deviation
Grade	59754	5.6199	3.5116
Admission grade	46811	5.8983	1.6964
Class size (log)	59754	4.1323	0.7587
Age in years	59754	25.052	2.8598
Dummies			
Women	59754	0.3805	
Danish	59754	0.7752	
External evaluation	59754	0.3095	
Small courses (ECTS5)	59754	0.4336	
Written exam	59754	0.5410	
Oral exam	59754	0.0902	
Take-home exam	59754	0.1440	
Take-home + oral	59754	0.0521	
Group exam	59754	0.0257	
Group exam + oral	59754	0.0342	
Other exams	59754	0.1128	
New grading scale (From 2007)	59754	0.3210	
Re-exam 1	59754	0.0452	
Re-exam 2	59754	0.0087	

Achievements of students can be measured in different ways. The ultimate goals of education may be high earnings and high employment rates for students after they graduate. At universities high achievements can be measured using their grading levels, drop-out rates and the time used on their studies. This study uses grading and in 2007 a new grading scale with 7 steps (-3, 00, 02, 4, 7, 10, 12) and compatible with the ECTS-scale was introduced. A translation schedule from the former 10 step scale was agreed on and this has been used to adjust the old grading before 2007. The average admission grades from the bachelor degree are only available for Danish students and have been transformed to the new grading scale as well.

The files with the number of students enrolled in the different classes are lost after 2 years. Therefore class size has been defined from the number of gradings in the exam periods and excluding classes below 8 students. As there is no obligation for students to appear in class or take the exam afterwards, this definition probably gives a more reliable estimate of the number of active students in the classes. However, the data only include students attending the full time programs and besides some exchange students may attend the classes as well, on average below 10%.

Dummy variables have been defined to measure gender and national background of students as well as the different types of exam. Women make up 38% of the students and 77% are Danish. The most common exams are on-site written with 54% while group exams which are not allowed anymore account for 6%. 4% of the exams are a first time retake and less than 1% of the exams are a second retake.

Over the sample period from 1999 to 2010 a high degree of internationalization of the studies took place where the international student share of the grading increased from 5% to 36%, see table 2. Also the total number of gradings has more than doubled over the period and as teaching resources have not been allocated to the same extent the average class size has increased from around 60 in the beginning of the period to above 100 at the end of the period. The average grade of 5.6 has a weak upward trend over the period.

Table 2: Developments in internalization, class size and grade level

Exam term	Grade average	International Students share	Class size	Number of grades
1999.6	5.23	0.05	70	1837
2000.1	5.12	0.07	66	2048
2000.6	5.35	0.07	58	1691
2001.1	5.34	0.12	74	2263
2001.6	5.48	0.14	66	1860
2002.1	5.25	0.19	78	2344
2002.6	5.50	0.20	71	2068
2003.1	5.68	0.21	78	2508
2003.6	5.27	0.20	61	2373
2004.1	5.53	0.25	88	3136
2004.6	5.56	0.20	62	2476
2005.1	5.35	0.29	81	2882
2005.6	5.55	0.21	60	2374
2006.1	5.53	0.33	81	2971
2006.6	5.81	0.26	51	2340
2007.1	5.39	0.22	114	3225
2007.6	6.23	0.22	48	2176
2008.1	5.92	0.19	94	3192
2008.6	6.05	0.21	49	2800
2009.1	6.06	0.27	117	4134
2009.6	5.65	0.35	61	3717
2010.1	5.89	0.36	143	5339
All	5.62	0.23	82	59754

3. Empirical evidence

Table 2 shows the results of four models estimating the grade premiums for students attending class teaching. The class size effect is negative and highly significant in all the OLS estimations which control for some student characteristics and types of exams. Introducing the admission grade from their bachelor degree in model 2 does not chance the class size effect in any significant way, but it comes close to doubling the R-square and the model's ability to explain the variation in grades. There is a clear positive effect from the average grade of their bachelor degree to their master grading. However, the coefficient is only 0.70 and significantly below one and therefore some regression to the mean exists. The implication of this is that students with a grade below average at the bachelor level have a higher probability of increasing their grade compared to the good student above average at the bachelor level. Part of this regression to the mean can be a result of the bounded grading scale where students with a 12 at their bachelor level have no possibility of increasing their grades like students in the bottom of the scale only have one way and that is up.

Admission grades only partly control for student heterogeneity and therefore model 3 introduces fixed effects for students which further double the model's ability to explain the variation in grades. The fixed effect estimation also controls for students sorting on class size as it only picks up the effect on the grades from the individual student attending different class size. The estimated class size effect is significantly larger than in the simple OLS estimation verifying that some sorting takes place with more low ability students in smaller classes. In the first year of the program the students have to follow mandatory courses but in their third semester they can choose among the elective courses offered and thereby sort themselves into smaller classes with a higher grade premium.

It is often claimed that delaying the master study a few years will enhance the students' ability to perform well as they become more experienced. This is not the case as the grades are reduced on average by close to 0.1 per year of delayed study. Women receive significantly higher grades but introducing the admission grade from their bachelor study in model 2 turns it insignificant. This indicates that they probably also earn higher grades in their bachelor study. Also students with a Danish background received significantly higher grades. As there is no information on the permission grade for foreign students it is unclear where they perform less well due to their bachelor background or their disability to learn in a Danish teaching environment.

Table 2: Estimation of grade premiums

	Model 1	Model 2	Model 3	Model 4 Good student
	Dependent variable: Grade of exam			
Intercept	7.4680 (0.1530)	3.4082 (0.1842)	Fixed effect for student	8.2539 (0.2140)
Class size (log)	**-0.3235** (0.0186)	**-0.33410** (0.0198)	**-0.4185** (0.0190)	**-0.1410** (0.0216)
Admission grade		0.7017 (0.0084)		
Age	-0.0960 (0.0047)	-0.0569 (0.0058)		-0.0580 (0.0073)
Women	0.2568 (0.0279)	0.0684* (0.0293)		-0.0309* (0.0320)
Danish	1.0393 (0.0338)			0.0473* (0.0448)
External evaluation	-0.2322 (0.0349)	-0.1080 (0.0353)	-0.0912 (0.0332)	-0.0195* (0.0409)
Small courses (ECTS5)	0.4292 (0.0291)	0.4006 (0.0303)	0.4847 (0.0276)	0.2484 (0.0338)
Oral exam	1.4486 (0.0546)	1.5262 (0.0535)	1.5113 (0.0517)	1.7775 (0.0661)
Take-home exam	1.4817 (0.0417)	1.7080 (0.0443)	1.3974 (0.0387)	0.9973 (0.0472)
Take-home + oral	2.6035 (0.0636)	2.8134 (0.0620)	2.3274 (0.0631)	1.7281 (0.0662)
Group exam	3.2991 (0.0872)	3.5612 (0.0840)	3.2663 (0.0791)	2.4657 (0.0945)
Group exam + oral	2.2667 (0.0760)	2.4778 (0.0769)	2.0381 (0.0687)	1.7353 (0.0817)
Other exams	1.3192 (0.0455)	1.5518 (0.0504)	1.3740 (0.0434)	0.7838 (0.0534)
Re-exam 1	-1.4875 (0.0651)	-0.8120 (0.0703)	0.1609 (0.0609)	-0.2263* (0.1248)
Re-exam 2	-2.1139 (0.1448)	-0.9715 (0.1605)	0.2287* (0.1341)	-0.8491* (0.4422)
New grading scale	0.5330 (0.0314)	0.3160 (0.0332)	0.4408 (0.0886)	0.5802 (0.0359)
R-square	0.1274	0.2395	0.4250	0.0832
Observations	59,754	46,811	59,754	30597

Note: All coefficients are highly significant except those marked with an *.

External evaluation of grading has a significant negative effect on the grade level as may be expected. Surprisingly small courses of only 5 ECTS reward students with about 0.5 higher grades on average compared to larger courses of typically 10 ECTS. Re-exam has a significantly lower grade in model 1 due to the fact that it is students with a grade below average which normally take the re-exam. However, controlling

for student fixed effect in model 3 the students actually improve their grade in the re-exam but only marginally and not significantly in their second re-exam.

The type of exam has surprisingly large and significant effects on the grade which are quite stable across the first 3 models. An oral exam and a take-home exam reward the student with about 1.5 higher grades compared with an on-site written exam which is the bench mark as it has been excluded as a dummy variable. Group exams are the most favourable type of exam for the student with a 3.2 higher average level which amounts to the amazing figure of 57% higher grade measured at the average grade level of 5.6 for class teaching. Surprisingly students do not seem to choose courses according to the size of the grading premiums given for the different types of exam as the estimated coefficients are not significantly larger in the fixed effects estimation controlling for student heterogeneity.

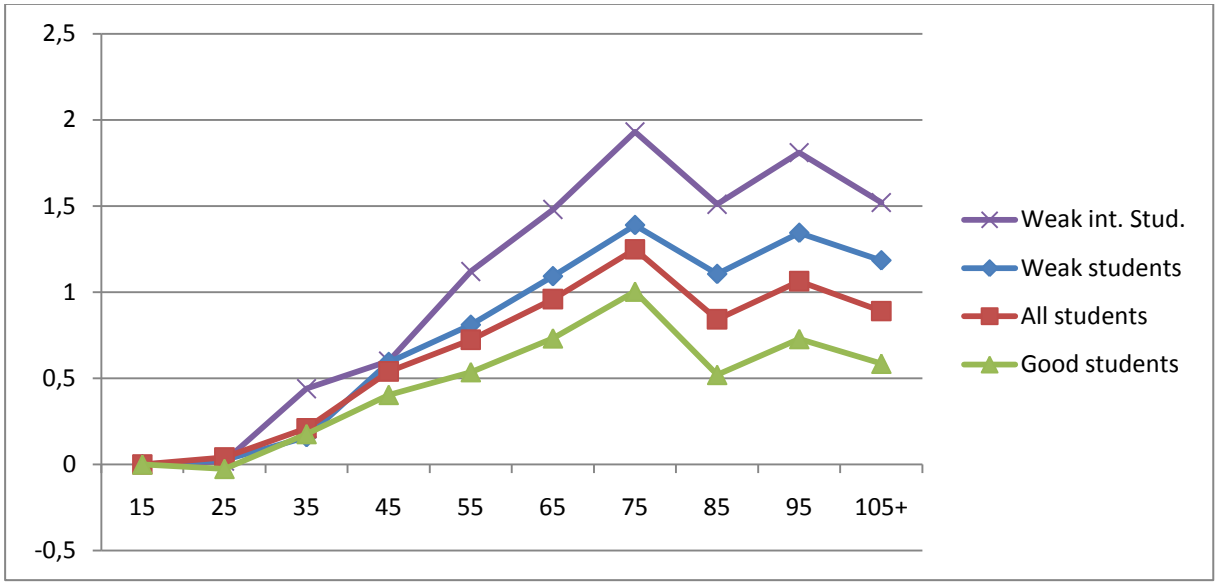
Model 4 estimates the grade premiums for the good students with an average grade above the overall average grade of 5.6 which amounts to half the students. The class size effects are significantly lower for the good student which mirrors that smaller classes are more helpful for students with less ability which therefore needs more supervision from the teacher or active interaction with other students in their learning process. It is also worth mentioning that in the group of good students, women and Danish students are not graded significantly different and external evaluation does not seem to have any influence on the grade level on average. Further, it seems that an oral exam is more beneficial for the good students whereas students with less ability have advantages with the take-home and group exam where they can more easily gather and share information with others.

The new grading scale which was introduced from the autumn term of 2007 has elevated the average grading level with about half a grading point. It is seen from model 4 that the new scale was especially beneficial for the good students and there is actually no significant effect on the grades among students with an average grade below the overall average. This underlines the widely accepted view among teachers that the transformation of the old grades was to the disadvantage of students with a high grade. Especially it has been claimed that a 9 in the old scale which was grouped with an 8 into a 7 in the new scale should instead have been transformed to a 10 in the new scale. One has to take this into account when comparing diplomas on the old grading scale with applicants graded on the new scale.

The large negative effects on grading from larger classes is estimated using a log transformation in table 2 which indicates that the class size effects are larger in smaller classes. To further study the class size effect, the class size variable has been categorized into groups of 10 and a dummy variable for each size group has been used in model 3 instead of the log transformation. Figure 1 lists the negative grade

premiums in the different class size categories for all students and for good students with an average grade above the averages and for weak students below the average and separately for weak international students.

Figure 1: Estimated negative grade premiums in different class size categories



Comments: Weak and good students have been defined as students with an average grade below and above the overall average.

The dummy variable for the class size group 8-20 students has been left out of the estimations and is then the bench march. The grading level in classes from 20 to 30 students is not significantly different from the grading level in the smallest classes below 20 students. But for classes above 30 students the negative grading effect is highly significantly and increasingly negative up to a class size about 70 students where the negative effect from class size levels off. These results are different from the class size effects found in the primary schools where the negative learning effects set in already at a class size of about 5 students, see Heinesen (2010). The larger optimal class size in universities verify the theory discussed above, where the positive interactions among students and the individual focused teaching from the teachers probably level off at 30 students in class and totally disappear at class size of about 70 students. At classes larger than 70 students there seems to be no diseconomies of teaching as it turns into a one way communication where the students are listeners and if the auditorium is equipped with loudspeakers the student can keep listening also in larger classes.

Figure 1 also shows that larger classes especially is a disadvantage for weak students which on average are graded 1.2 below the level in the small classes compared to only 0.6 below for the good students. This is especially true for weak international students where the disadvantage in larger classes comes close to 2 grading points. This shows

that the learning environment is more important for the weak students whereas the good students have higher ability to learn by themselves and therefore are more independent of the teaching environment.

As mentioned earlier the budget cuts for higher education during the period mainly through reallocation of university funding from basic funding to target research programs with a low containment of teaching has drained the resources away from research based teaching. As the basic funding has been fixed at a historical level, universities or studies with a high growth in student enrolment have been squeezed most by this budget method. As the enrolment in the master program at the Aarhus School of Business has doubled during the period the result has been an increase in the average class size of close to 100%, see table 2.

An estimate of the effects of the increases in class size on student achievements can be done using the estimated size coefficients in table 3. If the average class size has doubled during the period the average grade is reduced by 0.42 grading point. However, the new grading scale has reduced the average grade from the old grading scale before 2007 with 0.44 grading point according to the estimated coefficient in table 3. These two effects on the average grade mainly cancel each other out over the whole period and leave the average unchanged.

4. Conclusion

The production technology for higher education is not well known and this study highlights the relation between class size and student achievements using a large dataset of student from the Aarhus School of Business. The data set covers more than 80.000 gradings and the estimations of class size effects control for a number of student and class characteristics like gender, age, average grade from the admission education, nationality, type of exam and external evaluations among others.

The estimations show a large negative effect of larger classes on the grade level of students. It further shows that there is no significant class size effect on grading of classes below 30 students, but from 30 to 70 students the student achievements decrease significantly. For classes above 70 students there is no further negative effect from increases in class size. For weak students the class size effects are even larger.

The type of exam also has a large and significant effect on student achievements. An oral exam and a take-home exam reward the student with about 1.5 higher grades compared with the on-site written exam. Group exams are the most favourable type of exam for the student with a 3.2 higher average grade level compared to a written exam.

References

Heinesen, Eskil (2010): Estimating class-size effects using within-school variation in subject-specific classes. *Economic Journal*, 120, pp. 737-760.

Hanushek, Eric A. (1998): Conclusions and Controversies about the effectiveness of School Resources. *Federal Reserve Bank of New York Economic Policy Review*, March, pp. 11-27.

Hedges, L. V., R. Laine and R. Greenwald (1994): Does money matter? A meta-analysis of studies of the effects of differential school inputs on students outcomes. Education Researcher, pp. 5-14.

Krueger, Allan B. (1999): Experimental Estimates of Educational Production Functions. *Quarterly Journal of Economics*, pp. 497-532.

Krueger, Allan B. (2003): Economic Considerations and Class Size. *The Economic Journal*, pp. 34-63.

Lazear, Edward. P. (2001): Educational Production. *Quarterly Journal of Economics*, pp. 777-803.