Does attendance improve academic performance: empirical investigation of undergraduate students?

Umid Aliev

1. Introduction

Traditional approach in higher education assumes that substantial part of learning happens in classrooms, when students physically attend lectures and seminars. It is believed that attendance raises academic performance through providing access to learning materials and establishing useful learning environment where students benefit from in-class discussions with each other and a tutor. However, traditional view on education is seriously challenged by anecdotal evidence that student attendance, especially on undergraduate courses, is far below perfect. Indeed, if, with other things equal, attendance raises academic performance, and therefore post-university career prospects, then why would students miss classes? It is especially interesting given that once tuition fees paid students do not get refund for missing a class. Does it mean that students do not find attending classes academically useful? Are these students right in devoting their time in other activities than attending classes in their universities? Does attendance *really* improve academic performance? If we observe two students with equal cognitive ability, motivation, and efforts will the student with better attendance have better knowledge and skills? Answers to these questions may have important policy implications especially in current context of the UK. Increase in tuition fees from academic year 2012-2013 means that students will require better return for their payments forcing universities to arrange more efficient utilisation of learning resources including classroom time. Optimal allocation of learning resources will benefit from better understanding of learning impact of classroom time.

Another challenge for the traditional approach to education comes from rise of information communication technologies that makes provision of distance-learning cheaper and more efficient. Moreover, nowadays students do not have to be physically present in classrooms to get access to lecture materials. Most of the universities practice dissemination of lecture materials through virtual learning environment networks like Moodle or Blackboard which implies that students usually can get access to class materials without turning up to the classes. Therefore, one of the potential gains of attendance is no longer valid. Does it, in this case, imply that traditional approach to education has to be reconsidered or does physical attendance boosts students' learning? That's the main question we try to answer in this paper. We investigate attendance-performance relationship using a random sample of the first and second-year undergraduate full-time students registered at the University of Roehampton Business School.

Major problem arising in measuring impact of attendance on learning is endogeneity of attendance because students usually can *choose* whether to attend and the choice is affected by students' individual characteristics such as cognitive ability, motivation and interest to subject, and efforts to succeed in learning (e.g. self-study). In general, better able, more motivated and hard-working students tend to attend better. At the same time these students tend to perform better academically, too. It implies that in measuring attendance's impact on academic performance we have to account for the effect of the students' individual characteristics otherwise their effect will be falsely attributed to attendance. In other words, omitted variable bias arises. However, accounting for the effect of the students' individual characteristics can be a challenging task given that they are not usually directly observable or measurable. We deal with the omitted variable bias by utilising panel-data features of our dataset. For each student we observe attendance and performance he/she had in at least two (in the most of observations three) different modules. Panel-data features allow us using fixed effect transformation to account for the effect of unobservable individual characteristics without needing to measure them directly because the student's individual characteristics are unlikely to change substantially from one module to another.

Panel-data features of our dataset also allow us to claim that we observe students' attendance-performance relationship over a range of modules, while other studies are often limited to one module (e.g. Romer (1993), Stanca (2006), Arulampalam *et al.* (2012)) which makes it difficult to generalize the findings.

We also examine if there are differences in patterns of attendance-performance relationship between first- and second-year students. Usual practice in British universities is that only second- and third-year grades are considered for counting final degree classifications. It results that many first-year students are inclined more towards satisficing than maximising and are often content with mere passes, while the second-year students are usually more motivated to maximise their performance. If there are substantial differences in patterns of attendance-performance relationship it may indicate that the current practice when largely similar methods are used for teaching first- and second-year students should be reconsidered.

A further contribution of our paper is that we estimate non-linear specifications that argue that the impact of attendance varies across different levels of attendance and performance. This adds to current literature which tends to consider attendance-performance link as a linear relationship.

In the next section we present literature review. In Section 3 we describe the dataset and variables used in this paper. Section 4 provides detailed description of the methodology where the central point is dealing with the omitted variable bias. Section 5 presents discussion of the main results, while in Section 6 we demonstrate the outcomes of the robustness check. In Section 7 we investigate if there are differences in the patterns of attendance-performance relationship between the first- and second-year students. Final section concludes.

2. Literature review

Existing literature presents considerable evidence that there is a statistically significant positive relationship between attendance and academic performance. For example, Newman-Ford *et al.* (2008) in their large-scale investigation of 22 undergraduate modules within four separate award programmes report strong positive correlation between students' attendance and their academic performance. They also found that the attendance of students in the first and upper-second class degree classifications was significantly higher than those who received lower classifications or failed their modules. Thatcher *et al.* (2007) examined attendance and performance of 289 students in a second-year psychology class and found that number or attended classes was positively related to the students' marks. Particularly, the group that always attended obtained significantly higher marks than the groups that never or seldom attended. Romer (1993) reports that fraction of lectures attended in intermediate macroeconomics module has a positive impact on students' final grades and attendance alone accounted for nearly a third of the variance in performance. Stanca (2006) examines attendance of the students of introductory microeconomics course at the University of Milan between 2001 and 2004. The main finding of this study is that missing one lecture was associated with about a half percentage point drop in test score. Importance of the attendance has also been stressed by the estimation that the return to each hour spent attending lectures was substantially higher than the return to each hour of self-study. Park and Kerr (1990) employ a multnomial logit model and establish that number of times a student was absent or tardy to class affected negatively the probably of obtaining higher final grades. The model was estimated for students of undergraduate money and banking course. Durden and Ellis (1995) report non-linearity in the relationship between attendance and academic performance. Particularly, they found that low levels of absenteeism had no effect on the student's average course score, while excessive absenteeism has a large and significant effect.

Other authors that investigate relationship between attendance and academic performance include Schmidt (1983), Durden and Ellis (2003), Devadoss and Foltz (1996), Launis (1997), Thomas and Higbee (2000), and Martinez (2001). Usually attendance is measured through a number of taught sessions attended or percentage of them relative to the total number of the sessions. Traditional indicator of academic performance is a score or mark awarded for an exam or coursework. Alternatively, it can be an overall score/mark for a module or course. All these studies report statistically significant positive relationship between attendance and academic performance, although the size of the impact is found to be moderate in some studies.

It seems that there is considerable evidence suggesting existence of a positive impact of classroom attendance on academic performance. Nevertheless, the debate about contribution of attendance to learning does not look like to be over. The main reason for it is that the reported empirical findings are often arguable because measuring the impact of attendance is complicated by the fact that attendance levels are rarely, if ever, exogenous. Usually a student can *choose* whether to attend lectures and seminars and this choice is likely to be affected by individual characteristics such as ability, effort, and motivation. It is almost certain that better students, who are more able academically, more interested in the material, and work harder, tend to have higher attendance level, other things being equal. At the same time, student's ability, motivation and efforts directly affect his/her academic performance. In fact, Park and Kerr (1990) report ability and motivation to be more important determinants of academic performance than classroom attendance. In the conditions, when attendance and performance both are affected by students' individual characteristics it is difficult to establish if the observed statistical association between these two variables is primarily due to a genuine effect of attendance to learning or due to that attendance simply serves as a proxy for the students' individual characteristics. In other words, the estimates are subject to omitted variable bias unless the effects of the student's individual characteristics are taken into account.

Different studies take different approaches in dealing with this problem. Some authors simply brush it aside, although they recognise that their estimates are likely to be overstated (e.g. Newman-Ford et al. (2008) and Thatcher et al. (2007)). Unfortunately, this approach usually produces the results that are the most arguable and least convincing regarding the own impact of attendance on performance. More popular approach is to try to disentangle the impact of attendance on performance from individual characteristics by including additional proxy variables for ability, motivation, and efforts into regresson equations. Academic ability of the students is usually proxied by student's grade-point averages (GPA), scores on college entry exams, high school grades, and dummy variables for passing certain tests/exams (e.g. numeracy and calculus). Number of self-study hours and homework assignment completion serve as proxies for the student's efforts, while subject and teacher evaluations or fraction of problems sets completed in an exam are used to measure student's motivation. Examples of the authors using this approach include Schmidt (1983), Park and Kerr (1990), Romer (1993), Durden and Ellis (1995), Devadoss and Foltz (1996), and Stanca (2006). The findings of these studies indicate in two ways that it is very important to control for the effect of individual characteristics. Firstly, the proxy variables expose statistically significant effect this implies that student's ability, motivation, and efforts have direct impact on academic performance. Secondly, the inclusion of the proxy variables into the regression equation results in reduction of the estimated coefficient of attendance. It means that the regression or correlation analysis produces positively biased estimates if the effects of the student's individual characteristics are not controlled.

However, inclusion of the proxy variables corrects the omitted variable bias to certain extent only. Usually the proxy variables are imperfect measures of student's ability, motivation, or efforts. As a consequence ordinary least squares (OLS) estimates of the effect of attendance obtained from the equations including the proxy variables may still be biased and inconsistent to the extent that they incorrectly attribute to attendance the effects of components of ability, motivation, and efforts that were not captured by the included proxy variables. Stanca (2006) demonstrates that inclusion of the proxy variables offers only a partial remedy to the omitted variable bias in performance-attendance equations.

Moreover, use of the proxy variables that imperfectly measure student's individual characteristics may actually increase the bias. Todd and Wolpin (2003) show that this happens if the proxy variable for ability is correlated with the component of motivation and/or efforts not captured with the corresponding proxy variable. Student's academic ability is very much likely to affect his/her efforts to study. In this case GPA score used as a proxy for the student's ability may be correlated with the part of variation in efforts that is not controlled by the number of self-study hours. In this case inclusion of GPA into the regression equation may lead to a greater bias than the specification without the proxy variables (see Todd and Wolpin (2003) for more details).

Our approach in dealing with the omitted variable bias is based on the use of panel properties of the data due to the fact that we observe the same students across multiple modules. We exploit this feature of the dataset to control for such unobservable characteristics of students as ability, motivation, and efforts. Detailed description of the dataset is given in the next section.

3. Data and variables

The data was collected from the random sample pulled from the population of the first- and second-year full-time undergraduate students of the University of Roehampton Business School. The sample contains information on 47 first-year and 56 second-year students. For all second-year students we have information on their attendance and performance in the following three modules: Managing Organisations (MO), Business Research (BR) and Human Resource Management 1 (HRM). The modules have been chosen if they are either compulsory for all students (MO and BR) or very popular optional ones (HRM), so sufficient number of students are registered across all three modules to form a sample of a decent size. For the first-year students we have information on their attendance and performance in compulsory Business Skills (BS) and People and Organisations (PO) as well as in optional Business Economics[1]. However, it should be noted that for 12 first-year students we have information only on the compulsory modules.

Class attendance was measured for spring semester of academic year 2011-2012. The attendance was recorded from the registers for lectures or seminars, whichever was available for a module. For example, BR, BS, and BE were taught in seminars-only mode and there were no lecture groups. HRM was apparently taught in lectures-only mode because no record of seminar groups was found. MO had both lecture and seminar groups but filled registers were available for lectures only. Similarly, PO had both lecture and seminar groups but attendance in lectures was not recorded so we used registers for the seminars[2]. Attendance is measured as a proportion of taught classes attended.

In our sample students, on average, attended 56 percent of the taught classes with 10 percent of students attending less than 10 percent of classes, and about 10 percent of students attending more than 90 percent of classes. Average attendance of the first-year students is 62.5 percent of all taught sessions and it is higher than that of the second-year students, 51.8 percent. More details are given in Table 1.

Table 1. Descriptive statistics

Variable	Mean	SD	Min.	Max.
Total sample				
Classes attended (%)	56.47	28.58	0.00	100.00
Overall mark	57.85	11.50	2.00	82.00
End-of-semester mark	55.71	12.81	2.00	82.00
First-year students				
Classes attended (%)	62.51	25.90	0.00	100.00
Overall mark	53.54	12.51	2.00	75.00
End-of-semester mark	51.00	14.33	2.00	75.00
Second-year students				
Classes attended (%)	51.84	29.73	0.00	100.00
Overall mark	61.16	9.43	20.00	82.00
End-of-semester mark	59.32	10.16	20.00	82.00

[1] For 11 students attendance and performance in Business Economics are proxied by the ones in Marketing and Enterprise (ME) or Introduction to Business Law (IBL).

[2] We used seminar registers to measure attendance in ME as there was no attendance record for the lectures available. IBL was taught in lectures-only mode.

We use two indicators to measure students' academic performance: 1) overall module mark that is a weighted average of all assessment components taken by the students throughout both autumn and spring semesters; and 2) a mark of the final piece of module assessment taken by the students at the end of the spring semester. In Table 1 overall module mark is titled as overall mark and the mark of the final assessment component is referred as end-of-semester mark. These are the titles we will use for the rest of the paper. We expect that end-of-semester mark will be more closely related to the spring semester attendance than the overall mark. This expectation is caused by the fact that attendance in our analysis is measured for the spring semester only and the end-of-semester mark is the one awarded for the exam taken at the end of the spring semester, while the final mark reflects student's academic performance throughout both autumn and spring semesters.

Table 1 shows that, in general, overall mark is somewhat higher than the end-of-semester one. This may indicate that students tend to get higher marks on coursework than in unseen time constrained exams, which the usual form of the final assessment component; and it seems to be true for both, first- and second-year, students. We can also see that in our sample second-year students tend to be awarded higher marks than the first-year ones, although on average the second-year students attended smaller proportion of taught classes than the first-year ones. It may indicate that there is certain difference is learning patterns between the first- and second-year students, although proper tests are needed to make valid inferences as information on Table 1 is of descriptive nature only.

4. Methodology

The empirical specification used in this paper is based on a notion that learning is a cumulative process depending on academic and student inputs. Academic input broadly refers to teaching (lectures, classes, seminars, office hours, etc.). Student input includes a number of personal factors, among which the three main ones are cognitive ability, motivation, and efforts. We employ a simple additive model and describe the relationship as follows:

$$y_i = \beta_0 + \beta_1 x_{1i} + \boldsymbol{\beta_2 x_{2i}} + \varepsilon_i \qquad (1)$$

where y_i is learning for individual *i*, x_{1i} is a vector of variables measuring academic input and x_{2i} is student input, e.g. ability, motivation, and efforts. ε_i is an error term.

Learning is measured by academic performance (exam and/or overall score) and academic input by class attendance. Student input is more difficult to measure as individual characteristics like ability, motivation, and efforts are usually not either observable or measurable. This would not be a problem for estimating β_1 (which is our prime interest) if x_{2i} and x_{1i} were uncorrelated. However, student's individual characteristics are likely to be positively correlated with attendance: anecdotal and empirical evidence suggest that those students that are more able, motivated, and work harder, tend to have higher attendance levels. As a consequence, omission of x_2 from equation (1) would mean that OLS estimations of β_1 are biased and inconsistent because they attribute to attendance the direct effect of the student's individual characteristics. In other words, the regression results will be subject to the omitted variable bias.

Our approach in dealing with the problem of omitted variable bias is based on utilising the panel feature of our dataset: we observe performance and attendance across two or three modules for all the students in our sample. The maintained assumption in this context is that the effects of student characteristics – such as ability, motivation, and efforts – are common across the three modules observed in our sample. This assumption would be less reasonable if students were taking modules from academically disparate directions, like arts and science. Instead, in our sample students are observed across the modules that are all offered within one school and related to each other. For this reason we believe that it is reasonable to assume that effect of the student's unobservable characteristics do not vary from one module to another. On the basis of this assumption we transform the equation (1) into

$$y_{ij} = \beta_1 x_{ij} + a_i + u_{ij} \qquad (2)$$

where y_{ij} is a student *i*'s mark in module *j* with $i = 1, \dots, N$ and $j = 1, \dots, J$. On the same way x_{ij} indicates attendance. a_i accounts for the effect of student's unobservable characteristics like ability, motivation, and efforts, that is assumed to be constant across the modules and potentially correlated with x_{ij}. u_{ij} represent idiosyncratic disturbances that are uncorrelated with x_{ij} and a_i and it is assumed $u_{ij} \sim (0, \sigma_u^2)$.

Because a_i is unobservable we use fixed-effect (FE) transformation to eliminate it. The FE transformation is obtained by subtracting from each variable its mean value averaged over $j = 1, 2, 3$. In other words, FE transformation implies

$$y_{ij} - \bar{y}_i = \beta_1 \left(x_{ij} - \bar{x}_i\right) + \left(u_{ij} - \bar{u}_i\right) \qquad (3)$$

where $\bar{y}_i = J^{-1} \sum_{J=1}^{J} y_{ij}$, $\bar{x}_i = J^{-1} \sum_{j=1}^{J} x_{ij}$, and $\bar{u}_i = J^{-1} \sum_{j=1}^{J} u_{ij}$. Wooldridge (2002) shows that OLS estimation of equation (3) produces unbiased and consistent estimates of β_1 as long as the transformed errors terms and independent variables are uncorrelated. We employ this method to evaluate the impact of attendance on academic performance.

5. Results

In this section we discuss the estimation results of various specifications of equations (2) and (3). Dependant variable is academic performance measured by the overall mark obtained by a student for a module. This mark is usually a weighted average of the marks obtained for a number of assessment components throughout the academic year. Alternatively, academic performance is also measured by the mark obtained by a student for the final assessment component that was taken at the end of the spring semester. The independent variable of our prime interest is attendance. Attendance is measured as a proportion, between 0 and 1, of taught classes attended by a student. It should be noted that for regression analysis we normalise both academic performance and attendance by dividing them to their means by modules. Need for the normalisation stems from the fact that average academic performance and attendance may be affected by module-specific factors such as quality of teaching or complexity of subject and others alike. Consequently, failing to control for their effect may result in obtaining overstated coefficients when the effect of the module-specific factors will be falsely attributed to attendance. Usually, these factors are unobservable and they are not eliminated by FE transformation because they are not necessarily constant across modules. Therefore, in the regressions we use academic performance relative to average in the module and attendance relative to average in the module.

Table 2 contains results of various specifications of equations (2) and (3). Upper part of the table presents the results for the specifications where the dependant variable is a relative overall mark for a module. The lower part does it for the specification where the dependant variable is a relative end-of-semester mark.

Table 2. Relationship between attendance and academic performance

Variable	(1)	(2)	(3)	(4)
Dependent variable: overall mark (relative)				
Classes attended (relative)	0.138***	0.259***	0.324***	0.148**
	(0.027)	(0.098)	(0.097)	(0.070)
Classes attended (relative) squared		-0.065***	-0.099**	-0.074**
		(0.046)	(0.045)	(0.033)
Wald tests: coefficients of module dummy variables are zero [p-value]	Not included	Not included	4.52 [0.0006]	4.49 [0.0022]
Adjusted R^2	0.1051	0.1130	0.1728	0.0868
Hausman test: statistic				24.159
[p value]				[0.0001]
Dependent variable: end-of-semester mark (relative)				
Classes attended (relative)	0.144***	0.318***	0.336***	0.175**
	(0.029)	(0.101)	(0.103)	(0.072)
Classes attended (relative) squared		-0.094**	-0.103**	-0.074**
		(0.047)	(0.048)	(0.037)
F-stat of the Wald test: coefficients of module dummy variables are zero [p-value]	Not included	Not included	2.21 [0.0534]	3.47 [0.0106]
Adjusted R2	0.1003	0.1150	0.1428	0.0759
Hausman test: statistic				13.601
[p value]				[0.0087]

Notes: number of observations is 297. Columns (1) - (3) present results of OLS estimations of untransformed variables. Column (4) presents the results of fixed-effects estimation. Numbers in round parentheses are standard errors robust to heteroskedasticity and serial correlation across modules. Robust form of Hausman statistic is used.
*** significant at 1%; ** significant at 5%; * significant at 10%.

In the basic univariate specification (column 1), the point estimate indicates that increase in attendance by 10% relative to average in a module will result in 0.014 increase in the student's final mark relative to module average. Subject to the estimated values of the means (see Table 1), it implies that attending 5% of more classes results in increase in final mark by 0.8 mark. The regression results suggest somewhat stronger effect of attendance on the end-of-semester mark. This is an expected outcome given that attendance in our analysis is measured for the spring semester only and the end-of-semester mark is the one awarded for the exam taken at the end of the spring semester, while the final mark reflects student's academic performance throughout both autumn and spring semesters.

In column (2) we present the result of the specification with non-linear relationship between attendance and academic performance. The estimated coefficients of attendance and its square are both statistically significant and have expected signs. The estimates suggest that attendance-performance relationship is non-linear, which implies that attendance has a positive effect on academic performance but the size of the effect diminishes with increase in attendance. Particularly, from the estimated values of the coefficients we can calculate that the effect of attendance on academic performance diminishes to zero when attendance exceeds average level by 70% (end-of-semester mark) or 90% (overall mark). Subject to the estimated values of the means (see Table 1), it implies that when a student attends 95% of all classes further increase in attendance does not result in better academic performance.

We also introduce dummy variables for each module to capture effect of any module-specific factors (e.g. level of student support outside classes) that could affect students' academic performance. The estimation results presented in column (3) confirm the correctness of this step because Wald tests reject the hypothesis that the coefficients of the control variables are equal to zero. The estimates of the coefficients of attendance do not change substantially and the inferences are largely unchanged from those presented on the previous paragraph.

In column (4) we present results of estimating equation (3), where the effect of the student's unobservable characteristics like ability, motivation, and efforts has been eliminated through FE transformation. In general, FE estimation results imply that if we observe two students with equal ability, motivation, and efforts the one that attends more classes is more likely to obtain a higher mark. However, because the relationship between academic performance and attendance is estimated to be non-linear it indicates that growth in attendance causes growth in academic performance but at decreasing rate. Particularly, the effect of attendance diminishes to zero when a student attends 60-70% of all taught classes. Our finding that attendance-performance relationship is non-linear is in line with the results of Durden and Ellis (1995) and Arulampalam et al. (2012)

The striking finding is that the value of the coefficients of FE estimation is nearly half of those of OLS estimation in column (3). It implies that nearly half of the effect on academic performance attributed to attendance in column (3) was indeed due to student's ability, motivation, efforts, and other similar unobservable characteristics. Moreover, the drop in the values of the coefficients in FE estimation suggests that there was a positive correlation between unobservable characteristics and attendance. This is also confirmed by the Hausman test statistic (see Table 2), that strongly rejects the null hypothesis that unobservable characteristics are uncorrelated with attendance. In general, FE estimation results indicate that our expectation of student's unobservable characteristics affecting both academic performance and attendance is not only theoretically plausible but also has been confirmed empirically. It also highlights the importance of controlling for the impact of unobservables, when otherwise the effect falsely attributed to attendance could be as large as its own genuine one.

One possible objection to our conclusion that attendance has a positive and non-linear impact on academic performance could arise from the fact that the academic performance is measured through exam and test scores. In general, attendance could affect exam performance because students learn how to do well on the exam without any actual effect of quality of learning (Stanca (2006)). This is true if lectures provide examinable material that is not covered in textbooks of offer information about topics and types of exam questions. However, this criticism does not seem to apply to the dataset we used for analysis. First of all, all students have full access to all lecture materials and past exam papers through the virtual leaning environment network. It means that attendance does not usually reveal any private information. Moreover, undergraduate teaching material usually closely follows the textbook and students can still answer correctly to exam questions if they decide to rely exclusively on textbooks instead of attending classes.

6. Robustness

Existing empirical literature indicates that sometimes values of the estimated coefficients may be substantially affected by presence of so-called outliers – observations that are numerically distant from the rest of the data. Presence of outliers in data may indicate measurement errors. We believe it is unlikely that there is a problem of measurement error in our data because it was collected from primary sources and all observations were entered with caution and checked thoroughly. However, outliers may arise by chance in any distribution. In our case an outlier could be a student who did not attend classes but still obtained a good mark. That could happen if the student is of very high capability or did similar module previously.

We identify outliers using added variable plots, and they indicate that several (quite few) observations in the dataset could be considered as outliers. We employ winsorising technique to eliminate the effect of the outliers. Winsorising is a technique where observations with extreme values (outliers) are replaced by a specified percentile of the data. For example, 90% winsorising means that all values below the 5th percentile are replaced with the 5th percentile, and all values above 95th percentile are replaced by the 95th percentile. We winsorise 5% of each tail for all the variables. FE estimation results for the winsorised data are presented in Table 3.

Table 3. Relationship between attendance and academic performance – 5% of observation from each have been winsorised. FE estimation		
Variable	(1)	(2)
Classes attended (relative)	0.158**	0.191**
	(0.081)	(0.076)
Classes attended (relative) squared	-0.079**	-0.081**
	(0.037)	(0.039)
Wald tests: coefficients of module dummy variables are zero [p-value]	5.29	3.91
	[0.0007]	[0.0054]
Adjusted R^2	0.1015	0.1019
Hausman test: statistic	25.608	10.420
[p-value]	[0.0000]	[0.0339]

Notes: number of observations is 297. Column (1) presents results for the specification with a dependent variable being a relative overall mark. Results for the specification with a relative end-of-semester mark as a dependant variable are given in column (2). Numbers in round parentheses are standard errors robust to heteroskedasticity and serial correlation across modules. Robust form of Hausman statistic is used.
*** significant at 1%; ** significant at 5%; * significant at 10%.

We can see that winsorising results in certain changes in estimated values of the coefficients. However, the changes are not large. More importantly, our conclusion that the relationship between academic performance and attendance is non-linear remains valid. Moreover, the regression coefficients obtained from the wisnorised data also indicate that the effect of attendance diminishes to zero when a student attends 60-70% of all taught classes.

In general, we conclude that winsorising did not result in any substantial change in our findings. Our conclusions remain intact. It implies that outliers are unlikely to be important in the main regressions presented in Table 2.

7. Is there any difference in the attendance-performance relationship between first- and second-year students?

In this section we examine if there is any difference in the pattern of the attendance-performance relationship between first- and second-year students. Anecdotal evidence suggest that the second-year students are more likely to be motivated to maximise their scores, while the first-year students may aim simply to pass because the scores obtained during the first year of study are not counted for degree classification.

We split our sample into two sub-samples, separate for each level. Table 4 presents the results for the estimations using sub-sample of the first-year students only. Table 5 presents the results for the second-year students.

Table 4. Relationship between attendance and academic performance – first-year students only		
Variable	(1)	(2)
Dependent variable: overall mark (relative)		
Classes attended (relative)	0.634***	0.030
	(0.204)	(0.268)
Classes attended (relative) squared	-0.187*	-0.047
	(0.100)	(0.127)
Wald tests: coefficients of module dummy variables are zero [p-value]	1.39	1.47
	[0.2521]	[0.2413]
Adjusted R^2	0.2809	0.0599
Hausman test: statistic		22.721
[p value]		[0.0000]
Dependent variable: end-of-semester mark (relative)		
Classes attended (relative)	0.767***	0.382*
	(0.216)	(0.215)
Classes attended (relative) squared	-0.235**	-0.204*
	(0.113)	(0.112)
F-stat of the Wald test: coefficients of module dummy variables are zero [p-value]	0.20	0.33
	[0.8186]	[0.7185]
Adjusted R2	0.2927	0.0414
Hausman test: statistic		29.386
[p value]		[0.0000]

Notes: number of observations is 129. Column (1) presents the results of OLS estimations of untransformed variables. Column (2) presents the results of fixed-effects estimation. Numbers in round parentheses are standard errors robust to heteroskedasticity and serial correlation across modules. Robust form of Hausman statistic is used.
*** significant at 1%; ** significant at 5%; * significant at 10%.

OLS estimation results (see column (1) of Table 4) suggest that there is strong, positive effect of attendance on academic performance for the first-year students, and the effect is non-linear. We also undertake FE estimation to account for the effect of the student's unobservable individual characteristics. The results of the FE estimation are given in column (3) and they indicate that once we control for the impact of the student's ability, motivation, and efforts there is no strong effect of attendance on academic performance for the first-year students. Particularly, for the specification where a dependant variable is the relative overall mark attendance coefficients do not expose statistically significant effect, and in the specification with the relative end-of-semester mark as a dependant variable they are just marginally significant at 10% level. It implies that nearly all of the statistically significant effect observed in the OLS estimation results in column (2) is due to unobservable individual characteristics such as ability, motivation, and efforts, are correlated to both attendance and academic performance. This conclusion is also supported by the results of Hausman test, which strongly rejects the null hypothesis that the unobservable characteristics are not correlated with attendance.

In general, the estimation results for the first-year students indicate that once we account for the effect of the student's ability, motivation, and efforts, there is no strong effect of attendance on academic performance of the first-year students. It also implies that nearly all of the observed association between attendance and academic performance of the fist-year students is due to a simple fact that more able, motivated and hard-working students tend to attend classes and also obtain higher marks.

Table 5 presents the results for the second-year students.

Table 5. Relationship between attendance and academic performance – second-year students only

Variable	(1)	(2)	(3)
Dependent variable: overall mark (relative)			
Classes attended (relative)	0.179**	0.172**	0.181***
	(0.091)	(0.065)	(0.070)
Classes attended (relative) squared	-0.055	-0.076**	-0.066**
	(0.043)	(0.029)	(0.031)
Wald tests: coefficients of module dummy variables are zero [p-value]	5.01	7.29	14.55
	[0.0077]	[0.0016]	[0.0007]
Adjusted R^2	0.1056	0.1290	0.1228
Hausman test: statistic		2.339	
[p value]		[0.3106]	
Dependent variable: end-of-semester mark (relative)			
Classes attended (relative)	0.139	0.138*	0.140**
	(0.091)	(0.073)	(0.071)
Classes attended (relative) squared	-0.040	-0.046	-0.043
	(0.043)	(0.035)	(0.031)
F-stat of the Wald test: coefficients of module dummy variables are zero [p-value]	4.50	6.77	13.94
	[0.0125]	[0.0023]	[0.0009]
Adjusted R2	0.0960	0.1447	0.1443
Hausman test: statistic		0.136	
[p value]		[0.9344]	

Notes: number of observations is 168. Column (1) presents the results of OLS estimations of untransformed variables. Column (2) presents the results of fixed-effects estimation. Column (3) presents the results of random-effects estimation. Numbers in round parentheses are standard errors robust to heteroskedasticity and serial correlation across modules. Robust form of Hausman statistic is used.
*** significant at 1%; ** significant at 5%; * significant at 10%.

On Table 5 we can see that significance of the coefficients varies between the specifications. Particularly, OLS estimation results (column (1) upper part) of the specification with the relative overall mark as a dependant variable imply presence of a statistically significant, positive relationship between attendance and academic performance. The relationship seems to be closer to a linear one because the coefficient estimate of the squared attendance is not statistically different from zero.

Lower part of column (1) of Table 5 presents OLS estimates of the coefficients for the specification with the relative end-of-semester mark as a dependant variable. These estimates indicate that attendance has no statistically significant effect on academic performance. This is quite surprising result given that we expect closer association between attendance and end-of-semester mark. However, we have to remember that OLS estimates are biased and inconsistent because they do not account for the effect of the student's unobservable characteristics such as ability, motivation, and efforts. It is known that as a remedy to this problem we run FE estimation. Its results are presented on column (2) of Table 5.

FE estimation results suggest that there is a positive and non-linear effect of attendance on the student's relative overall mark. The coefficient estimates are statistically significant. The relationship between attendance and student's relative end-of-semester mark is also positive but seems to be closer to linear one because the coefficient of squared attendance is statistically insignificant. In general, attendance-performance relationship in this specification does not look strong as the attendance coefficient is just marginally significant at 10% level. However, it should be noted that in the sub-sample with the second-year students only the assumption that the student's unobservable individual characteristics are correlated with attendance is not confirmed. Particularly, Hausman test indicates that there is no evidence to reject the null hypothesis that the student's unobservable individual characteristics are uncorrelated with attendance. Econometrics literature suggests that in this context the random effects (RE) estimator is more efficient than the FE estimator. RE estimator accounts for the effect of unobservable characteristics but, differently from FE estimator, sets an assumption that they are uncorrelated with observed independent variables – attendance in our case (see Wooldridge (2002) for more details).

Column (3) of Table 5 presents the results of RE estimation for the second-year students. We can see that attendance positively affects both overall mark and end-of-semester mark. The coefficients are statistically significant at 1% and 5% levels. The relationship between attendance and overall mark is non-linear and point estimates of the coefficients suggest that the effect of attendance on academic performance diminishes to zero if a student attends 70% of taught classes. At the same time, the relationship between attendance and end-of-semester mark is closer to linear, which suggest that attendance has continuous positive effect on performance in this specification. This finding is in line with our expectation and previous results that end-of-semester mark is more closely associated with attendance.

In general, the findings in this section suggest that attendance of the second-year students is not correlated with their individual characteristics such as ability, motivation and efforts. That's an interesting observation given that we find the contrary for the first-year students. It implies that second-year students tend to attend regardless of their ability, motivation, or efforts, while only better first-year students attend classes. Such difference in student behaviour could be caused by difference in motivation, when the second-year students try to maximise their scores and the first-year ones are content with mere pass. This difference in motivation and behaviour seems to be a plausible explanation for the difference in the attendance-performance relationship. We found that once we account for the effect of the student's unobservable individual characteristics attendance has very weak or no effect on academic performance for the first-year students, while there is positive and statistically significant impact of attendance on academic performance of the second-year students. This finding raises questions on appropriateness of using different teaching approaches for the first- and second-year students and may also imply that benefits of the current practice when the scores obtained during the first year of study are not counted towards degree classification should be reconsidered.

8. Conclusion

In this paper we presented an empirical investigation of the impact of attendance on academic performance for the undergraduate students of the University of Roehampton Business School. We employed a sophisticated approach (fixed-effects estimation) to account for the omitted variable bias caused by the presence of the direct effect of students' unobservable individual characteristics (e.g. ability, motivation, and efforts) on their academic performance as well as attendance. Comparison of OLS and FE estimation results confirms correctness of the chosen approach.

In general, our results suggest that physical attendance has positive impact on academic performance. The impact is non-linear where the marginal effect of attendance is declining. It implies that attendance is particularly important for the students who have poor attendance and performance. Increase in attendance improves academic performance but the size of impact decreases with growth in attendance. It emphasizes the notion that physical attendance cannot and should not be considered as the only learning method. Nevertheless, our results indicate that physical attendance has its own direct impact on students' learning and therefore should be encouraged.

Another important finding in our examination is the stark difference in patterns of attendance-performance relationships between the first- and second-year students. We found that once we accounted for the effect of the unobservable individual characteristics attendance of the first-year students did not have statistically significant impact on their academic performance. The individual characteristics were found to be correlated with attendance indicating that observed association between the attendance and performance of the first-year students should be attributed to a simple fact that more able, more motivated and hard-working students were attending better and also performed better academically. Physical attendance did not have its own impact on learning of the first-year students. This finding may raise questions to efficiency of classroom teaching of the first-year students and appropriateness of the current practice when largely similar methods are used in teaching first- and second-year students.

In the case of the second-year students we observed the opposite. Attendance was found to have its own direct impact on academic performance and the student's individual characteristics were uncorrelated to his/her attendance. In other words, second-year students demonstrated absolutely different behaviour from that of the first-year ones. Possible explanation could be that the second-year students are more likely to try to maximise their grades while the first-year students could be content with mere passes due to a current practice when the grades on the first year of an undergraduate course are not counted towards final degree classification. However, on the example of the second-year students we can see that once students attempt to maximise their learning they benefit from classroom attendance.

We believe that our results shed more light in our understanding of the impact of physical attendance may have on student learning. This understanding is particularly important in deciding on efficient use of learning resources such as classroom teaching time. In turn, efficiency of use of learning resources has become especially important in the current context of the UK when substantially higher tuition fees were introduced from academic year 2012-13.

Bibliography

Arulampalam W., Naylor R., Smith J. 2012. Am I missing something? The effects of absence from class on student performance. *Economics of Education Review* 31, pp. 363-375.

Devadoss S., Foltz J. 1996. Evaluation of factors influencing student class attendance and performance. *American Journal of Agriculture Economics* 78 (3), pp. 499-507.

Durden G. C., Ellis L. V. 1995. The effects of attendance on student learning in principles of economics. *American Economic Review* 85 (2), pp. 343-346

Durden G. C., Ellis L. V. 2003. Is class attendance a proxy variable for student motivation in economic classes? An empirical analysis. *International Social Science Review* 78, pp. 22-34.

Launis M. H. 1997. College student attendance: attitudes and academic performance. *College Student Journal* 31, pp. 86-92.

Martinez P. 2001. *Improving student retention and achievement. What do we know and what do we need to find out?* London: Learning and Skills Development Agency.

Newman-Ford L., Fitzgibbon K., Lloyd S., Thomas S. 2008. A large-scale investigation into the relationship between attendance and attainment: a study using an innovative, electronic attendance monitoring system. *Studies in Higher Education* 33 (6), pp. 699-717.

Park K. H., Kerr P. M. 1990. Determinants of academic performance: a multinomial logit approach. *Journal of Economic Education* 21 (2), pp. 101-111.

Romer D. 1993. Do students go to class? Should they? *Journal of Economic Perspectives* 7 (3), pp. 167-174.

Schmidt R. 1983. Who maximizes what? A study in student allocation. *American Economic Review* Papers and Proceedings 73 (2), pp. 23-28.

Stanca L. 2006. The effects of attendance on academic performance: panel data evidence for introductory microeconomics. Journal of Economic Eduication 37, pp. 251-266.

Thatcher A., Fridjhon P., Cockcroft K. 2007. The relationship between lecture attendance and academic performance in an undergraduate psychology class. *South African Journal of Psychology* 37 (3), pp. 656-660.

Thomas P. V., Higbee J. L. 2000. The relationship between involvement and success in developmental algebra. *Journal of College Reading and Learning* 30 (2), pp. 222-232.

Todd P. E., Wolpin K. I. 2003. On the specification and estimation of the production function for cognitive achievement. *Economic Journal* 113 (485), pp. F3-F33.

Wooldridge J. M. 2002. *Econometric analysis of cross-section and panel data*. The MIT Press.