Increasing Within-School Competition: a case for department level performance indicators?

Shqiponja Telhaj[#], Robert Coe[†], Peter Davies*, David Hutton[†] and Nick Adnett*

Corresponding author: Centre for the Economics of Education, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, Tel: 02078523578, e-mail: s.telhaj@lse.ac.uk)

† Curriculum, Evaluation and Management Centre, University of Durham

* Institute for Education Policy Research, Staffordshire University Business School

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Increasing Within-School Competition: a case for department level performance indicators?

Abstract

We investigate the size and stability of departmental effects in English secondary schooling during a period in which extensions of parental choice and annual publication of school performance tables had significantly increased competitive pressures on schools. Our database of nearly 450 English secondary schools enables us to investigate departments in terms of both their unadjusted and value-added average student performance in national examinations. We are interested in the nature of intra-school competition and concentrate upon two subjects, Geography and History, which were optional subjects in each of these schools. In general, we find that relative departmental performance varies significantly over time and that few departments manage to persistently out-perform the other subject in their school. We conclude that given the instability of relative departmental performance, the planned publication of department level performance indicators is unlikely to generate strong incentives for departments to raise their effort and effectiveness.

Key Words: school choice; performance indicators; curriculum; competition; student attainment

1. Introduction

There has been much recent discussion as to the effects of greater competition on overall school performance (summarised in Adnett and Davies, 2002 and Hoxby, 2002). Although the evidence-base is still relatively weak, many countries have introduced market-based reforms of public schooling to generate more inter-school competition. These market-based reforms have frequently promoted greater between-school rivalry by increasing parental choice and publishing school-level performance indicators on average student attainment in curriculum-based external examinations. Whilst the schooling system in England has been at the forefront of these reforms, over the last decade this country has also seen a rapid growth of specialist schools (Hansen and Vignoles, 2005 provide further details of the English education system). More than two-thirds of England's secondary schools now have specialist status, focusing on areas like arts, business, languages, sciences and sports. In theory, specialist schools reduce inter-school competition by enabling schools to create niche markets in which parents and pupils who want a particular specialist curriculum have only one choice. Specialist schools are supposed to work in partnership and act as a resource for neighbouring, and therefore competing, schools (Bell and West, 2003). It follows that, in general, the growth of specialist schools weakens the link between school performance and parental and pupil choice. Hence our interest in this new environment in examining whether increasing choice within schools provides an alternative means of producing the benefits claimed for increasing choice between schools. The announcement in September 2005 of the intention to publish performance tables at teacher and subject level in England provides a further reason to explore this issue (BBC, 2005).

In principle, greater within school competition would appear likely to produce many of the advantages claimed for greater between-school competition. As Adnett and Davies (2005) explain, providing increased choice within schools can generate yardstick competition between departments and promote increased allocative, productive and dynamic efficiency in the schooling system. The main mechanism through which these effects would be generated is that students are likely to choose the most effective teachers as evinced by previous student's external examinations

performance. Hence, subject teachers seek to preserve their sunk capital by ensuring that their own students perform relatively well in such examinations and hence attract future cohorts. However, greater within school competition is also likely to generate the same sort of dysfunctional effects, such as cream-skimming and syllabus switching, as those produced by increased competition between schools. For example, school heads may seek to manage student choice to improve relative performance in school league tables. In addition, unlike between-school competition, within-school competition is likely to reduce inter-departmental co-operation and the shared promotion of whole-school policies, whilst the resulting increased diversity of educational outcomes may raise labour market signalling costs.

Our unique database enables us to examine departmental performance across a number of years during a period when the extension of parental choice and publication of school performance tables was increasing intra-school competition (Johnson, 2004). We are thus able to examine whether departmental effects are stable over time. Since pupils in England typically choose their optional General Certificate of Secondary Education (GCSE), a curriculum-based external examination at the end of compulsory schooling, subjects at age 14, two years prior to taking their exams, they have to rely on relative departmental performance in year t to predict performance in year $t + 3$. If relative departmental performance is unstable over time then it would seem unlikely that greater student choice can provide an effective incentive mechanism to promote greater teacher effort and effectiveness. A priori there is no clear theoretical prediction regarding the behaviour of relative departmental performance over time. One could argue that a particularly effective department would be likely to sustain a relatively high unadjusted and value-added effect over time as reputation effects enable student and staff cream-skimming. However, in an increasingly results-orientated environment, such as the English secondary schooling system in recent years, teachers in less effective departments will continually face pressures to raise their own relative performance by, amongst other things, increasing their levels of effort and effectiveness.

Our analysis is organised as follows. In section 2 we provide a review of the empirical evidence on the size and stability of school and department effects on student performance. In this section we also briefly outline the theory of within-school

competition, analysing its distinctiveness from conventional market-based reforms of state schooling. This is followed, in section 3, by an explanation of our empirical strategy including our choice of measures of relative departmental performance. In this study we concentrate upon performance in two common optional subjects: Geography and History. Section 4 describes our sample of nearly 450 English secondary schools with data on student GCSE performance analysed from 1998-2002. This unique database enables us to analyse both unadjusted and value-added performance. The following section reports our analysis of the stability of departmental rankings in these schools over time and the size and stability of relative departmental performance. Section 6 reviews our main findings and considers their policy implications.

2. A Theory of Within-School Competition

The market choice critique of publicly provided education with its advocacy of greater between-school competition has, as yet, been subjected to little empirical examination in England (Bradley and Taylor, 2004 provide a survey). Bradley and Taylor (2002) find that a 3% increase in the examination results in other local schools led to a 1% increase in a school's own performance: the impact being nearly twice as great in metropolitan as in non-metropolitan areas. This result is broadly replicated in Bradley and Taylor (2004) who also find that effects are much greater in metropolitan areas, though Gibbons et al. (2005) find that the degree of competition faced by primary schools in England had little effect upon their performance. However, school effectiveness research suggests that once adjustment is made for pupil characteristics then variations in pupil attainment levels between secondary schools in England are quantitatively small and unstable over time. For example, Yang and Woodhouse (2001) find that after allowing for prior student performance, gender, establishment type and region, individual establishments accounted for less than 5% of residual variance. On the stability of school effects, Mangan et al. (2005) find that English secondary school exam results are highly unpredictable and that any trends in performance are short-term. Kane and Staiger (2002) in their review of US evidence and Chay et al. (2005) examining the impact of educational reforms in Chile reach

similar conclusions about the unreliability of annual test scores as measures of performance differences across schools and over time.

There have been few attempts to break down these school-level effects and explore the contribution of departments to educational attainment. Though Davies et al. (2004a) found that there were large variations between the proportions of pupils taking examinations in individual subjects across English secondary schools. Only a small proportion of these differences can be explained by differences between the student cohorts attending schools or the size and type of school governance. Tymms (1992) suggests that the proportion of variance of pupil attainment attributable to schools is higher when performance is analysed on a subject by subject basis. He concludes that overall school performance is determined most strongly by the individual pupil, then by the department responsible, then by the school as a whole and, least of all, by the type of school. In the US, Rivkin et al. (2005) find large differences in teacher effectiveness whilst Nye et al. (2004) and Konstantopoulos (2005) find large teacher effects which exceed school effects. Together these results raise the question as to whether the policy emphasis should switch from school to department and teacher levels in the drive to raise pupil attainment.

Whilst greater within-school choice is likely to raise average student attainment levels through a better match of student's tastes and talents with their curriculum (Stables, 1997), we are primarily concerned here with resolving agency problems. In other words, the need to design mechanisms which encourage teachers to increase or divert their efforts towards improving student learning and increasing educational value-added. Incentive mechanisms in the form of whole school indicators, such as the annually published school performance tables in England, can create free-rider problems since the performance of individual teachers have only small effects on a school's rankings. Moreover, given the inertia in local hierarchies of schools, and the associated cream-skimming, many schools lack an incentive to improve their educational outcomes even with the advent of value-added tables (Adnett et al., 2002). Moreover the Tiebout motivated switching of pupils between schools championed by the advocates of school choice seem to create negative externalities, with pupils in schools with high turnover suffering a disadvantage (Hanushek et al., 2004).

Given these weaknesses, the alternative mechanism of greater within-school choice appears theoretically attractive. Stronger internal markets in schools create a situation in which future enrolments in each subject are dependent, in part, on their recent relative examination performance. This linkage can, in principle, generate powerful incentives as teachers seek to ensure the continuing usefulness of their sunk capital. However, this within-school choice argument relies on students rewarding relatively successful departments by opting for their courses. There is once again very limited empirical evidence on this issue for England. Adey and Biddulph (2001) found that only about 1 in 12 of the fourteen year olds surveyed claimed that previous examination results had influenced their choice of Geography or History at GCSE. In addition, resources are often allocated to departments on the basis of the numbers of students taught. Such allocation mechanisms create incentives for heads of departments to target raising student enrolments as a means of expanding their power base (Goodson, 1994). Davies et al. (2004b) found that absolute departmental performance, as measured by the percentage of students achieving grades A*-C, was negatively associated with the probability of a students being entered for examination in four of the six subjects they studied. However, these researchers found evidence of a significant positive relation between relative subject performance and the probability of a student being entered in that subject.

We have argued that for greater within-school competition to be an effective mechanism for raising educational attainment a pre-requisite is that students respond to relative departmental performance. If students are to use relative departmental performance as an indicator of the quality of teaching they will receive in the future then this relative performance needs to be stable over time. Hence our interest below is in the stability of relative performance over the four year period relevant for those making GCSE choices. The choice of which measure of relative performance students should consult when making their choice depends upon the relative size of peer and departmental effects. Unadjusted indicators contain some evidence of peer group effects which should have been cleansed from value-added measures. Hence in the empirical work below we utilise both unadjusted and value-added measures of relative department performance.

3. Empirical Strategy

Measuring the competition between departments within a school is problematic, especially in the absence of appropriate data. For this study we consider the difference of departmental performance between History and Geography departments since these two subjects are traditionally considered to be close competitors in English schools. Therefore, our task is to get a suitable measure of relative departmental performance in GCSE examinations, purged of any confounding influences. This is why we use two different indicators of the within school competition between History and Geography. First, we employ an unadjusted indicator of the within school competition where we compare the performance of the two departments within a school. Here, the departmental unadjusted average performance is simply the average GCSE performance of pupils in History and Geography given as:

$$\overline{SubjectScore}_s = \sum_{i=1}^{n} SubjectScore_i \Big/ n \qquad (1)$$

where n is the number of pupils that enrol in a particular subject. We take the difference in this performance between the two departments to be our first indicator of relative departmental performance. So,

$$Indicator1_s = \overline{History}_s - \overline{Geography}_s \qquad (2)$$

However, as we argued earlier, being an unadjusted measure of within-school competition, this relative departmental performance measure may be unreliable. This is because there may be different factors that have an impact on the departments' performance such as student characteristics or peer group effects. As an alternative to the first indicator, we use a value-added departmental performance. Student's value-added is provided in the YELLIS database and is measured by the amount of 'progress' made by a pupil when compared to that made by all other students of similar ability, so the value-added or pupil level residual is the difference between the attained grade and that predicted[1]. Therefore,

$$res_{ids} = GCSE_{ids} - \hat{\alpha} - \hat{\beta} YellisScore_{ids} \qquad (3)$$

[1] Details on how Yellis pupil value-added is calculated can be found at http://www.yellisproject.org

where *res* is the pupil's *i* residual in department *d* in school *s*.

Given equation (3) our measure of raw departmental residual would be:

$$\overline{\text{Re}\, s_{ds}} = \sum_{i=1}^{n} res_{ids} \Big/ n_{ds} \tag{4}$$

This departmental value-added estimate is standardised to be able to make comparisons between the History and Geography departments (i.e. the spread of the raw residual is put on the same scale). After this, we take the difference between the two departments value-added as our second indicator of relative performance:

$$Indicator2_s = \overline{\text{Re}\, sHistory_s} - \overline{\text{Re}\, sgeography_s} \tag{5}$$

There may still be some unobserved heterogeneity that may bias our estimates of performance such as family characteristics or other pupil characteristics that are not controlled for in our estimation of (4). There are some other potential problems that need to be addressed here. The first one is that only a small number of pupils may enrol in these subjects within a school. If there are only few pupils enrolled in History/Geography, then the average departmental performance could be subject to possible errors in estimation due to random factors. A standard error of a school performance in a particular is approximately: $sd(performance) \approx sd(\varepsilon_{is}) \Big/ \sqrt{n_s}$ where n_s is the number of students in a school enrolling in a particular subject. This means that the magnitude of the error diminishes as the number of pupils enrolling in a particular subject increases implying that it is difficult to produce reliable estimates of performance in very small departments. We address this issue by reducing the sample to only medium to large departments (i.e. we drop the data on departments that enrol less than 30 pupils).

A second problem that needs to be addressed is the statistical significance of the difference between the performance/value-added between History and Geography departments. This is important if one wants to appropriately judge the performance between departments over time. We address this issue by using a means comparison t-test to check whether the difference between the two departments in a school is statistically significant and how it varies over the years.

4. Data

Our sample is nearly 450 schools from the University of Durham's Curriculum, Evaluation and Management Centre's YELLIS database covering the period 1998-2002. This data provides us with information on student characteristics, including a measure of student ability, and student expected and actual attainment across a range of optional subjects. This panel has been chosen to include only schools in which Geography and History were optional subjects at GCSE in each of these years. We did so by only including schools which in each year had between 5% and 75% of their cohort enrolled in GCSE examinations in each of these subjects. Small schools (those with a total roll of less than 500 pupils) were also excluded from the panel. As reported in Telhaj et al. (2004), our panel accurately represented the total population of secondary schools with two exceptions: independent and selective grammar schools are slightly over-represented as are schools obtaining high Office for Standards in Education (OfSTED) grades for school leadership and management.

5. Results

We start our empirical analysis by presenting some simple statistics for the whole set of schools in our sample. Table 1 presents some simple statistics of our panel data where we describe the difference in performance between History and Geography departments between schools, and within schools over the years. We have generated a dummy variable that is 1 where the difference between History and Geography departments is positive and 0 otherwise.

Place Table 1 around here

Columns 2 and 3 in Panel A summarise the results for the whole 1998-2002 period. In 55% of the school-years (or 1225 school-years) the difference between History and Geography departments was positive. In column 4 we indicate the overall stability of the relative performance between History and Geography: a result critical for our study. The overall stability of the unadjusted relative performance is 57.9 percent which is normalised between the weighted average of the within percents and tells us that the difference fluctuates over time. For example, if a school has a positive difference between History and Geography departmental performance (i.e. History department performs better than Geography), this department will continue to perform better only 61.4 % of the time, and if this difference is negative, it will continue to be so only just over half of the time that we study (54.1 percent of the period 1998-2002). Looking at the last column of Panel A one can conclude that if a History department is performing relatively better than the Geography one, it will be slightly more stable (i.e. perform in the same way) than in a case where Geography department performs relatively better. However, in both cases almost half of the time departments experience a reversal of their previous relative ranking. The same picture appears in Panel B where departmental value-added difference is considered, though here the differences discussed above are even smaller.

Table 1 indicates the overall picture of instability, but given our research questions it is of interest to investigate this variation on a year by year basis. That is does this year's measure of relative departmental performance provide a reliable indicator of next year's relative performance? Transition frequency matrices (and transition probabilities) are a means of exploring the dynamics of the relative departmental performance. Table 2 shows overall transition probabilities and Table 3 illustrates probabilities by years.

Place Table 2 around here

In Table 2 transition probabilities represent the presence of transitions made from state i to state j given state i at the time t. The rows reflect the initial values and the columns reflect the final values. As can be seen (Panel A in Table 2) each year in 60 percent of schools where Geography had a better unadjusted GCSE performance

Geography remained the better performer, while in the other 40 percent of the schools History overtook Geography to become the better performer. Conversely, each year in 69 percent of schools that had better History department GCSE performance this relative performance was unchanged, while in 31 percent of schools the Geography department replaced History as the better average performer.

The results when departmental value-added performance is used to compare History and Geography departments are similar, with around 30 percent of school departments changing ranking between the two states. As can be seen in Panel B, in 33 percent of schools History department overtook the performance of the Geography department, while in 31 percent of schools where the History departmental performance was initially superior these rankings were reversed.

Given our interest in whether current relative departmental performance is a good predictor of future relative performance, we are interested in the variation of relative departmental performance over the GCSE cycle. This is why we now focus on transition probabilities across a four year period to see how these fluctuate. Table 3 demonstrates the transition probabilities between *t* and *t-3* period, comparing the relative departmental performance between 1998 and 2001 and between 1999 and 2002.

Place Table 3 around here

The results presented in Table 3 demonstrate that the instability of relative performance is even higher between time periods *t* and *t-3*: in more than 47 percent of the schools the History department overtook the previously higher performing Geography department between 1998 and 2001 while in around 39 percent where the History department initially had superior performance the rankings were reversed. Overall, this analysis suggests that the relative performance of History and Geography departments is very unstable over time. However, one might ask: given this instability to what degree can one make predictions? How reliable are these results? To test temporal dependences in sequence we calculate the *phi* coefficient that measures the degree of association in the transition matrices introduced above. It informs us about the degree of association between the relative departmental performances between time *t* and *t- k* (where *k* in our case is *k=1, 3*). As reported in Table 3, the *phi* coefficients are very low, the relative unadjusted performance between 1998 and 2001 is 0.27, while between 1999 and 2002 it is 0.34. This indicates that there is little association in relative departmental performance between *t* and *t-k* (i.e. if, say, a History department was performing better/worse than the Geography department it is unlikely that this relative performance remained the same at the end of this period).

The volatility of test score measures could in general weaken measures of school or departmental performance. Kane and Staiger (2002) find the sampling variation becomes particularly striking when the number of pupils in a department/school is small. The amount of variation stemming from idiosyncrasies of a particular sample of pupils being tested is often large relative to the total amount of variation observed between departments/schools. Secondly, they also found that other factors produce non-persistent changes in performance in addition to sampling variation (such as a disruptive pupil in a class, bad weather etc.). Additionally, schools, and particularly departments differ little in their rate of change in test scores or in their mean value-added. Moreover, those differences that do exist are often non-persistent, either because of sampling variation, or other factors. A department's performance will vary from one year to the other simply because the particular sample of students in a given year differs. Schools with a small number of pupils in History and Geography department that we analyse, would make test scores fluctuate more, and this problem is exacerbated by the year to year fluctuation.

So far in our paper we have included all schools regardless of the number of pupils that enrol in each subject, or whether there was a statistically significant the difference between History and Geography over the years. However, as argued in Section 3, one of the potential problems is the small number of observation upon which the departmental average performance is generated (in some schools this number is as low as 4 for History and 7 for Geography). This means that the departmental performance indicators could be subject to possible errors in estimation due to random factors. Sampling variation will account for a larger share of the between department variance for small departments and smaller share than large ones. Thus, a small department has a higher probability of being wrongly categorised as being the "better" or the "worse" department compared to a large one. In an attempt to tackle the problem of sampling variation we exclude small departments (those that enter less than 30 pupils in GCSE examinations) from our analysis. This reduces our sample to 264 schools but also reduces the random noise. With this reduced sample, the correlation coefficient of relative unadjusted performance between t and t-k is 0.28 for between 1998 and 2001 and 0.41 for 2002 which still indicates that relative departmental performance is unstable.

Another problem that we need to overcome is to look at the statistical significance of the relative effectiveness and see how this changes over years. This is important if one wants to judge on the performance between departments. To do this, we compare departmental average performances in each school by using a comparison means t-test that is calculated as below for each academic year in our sample:

$$t = \frac{\overline{H} - \overline{G}}{\sqrt{s_h^2 / n_h + s_G^2 / n_G}}$$ where n_H and n_G are the number of pupils in each department

In Table 4 we present the number of History departments performing statistically better than Geography departments, those that perform worse, and those whose performance does not differ significantly. We again present both unadjusted and adjusted relative performance measures (i.e. Indicators 1 and 2 as described above).

Place Table 4 around here

As can be seen in Table 4 in more than half of schools in our sample, (when we use the unadjusted indicator), there is no statistical significant difference in the performance of the two departments (148 of 264 schools in 1998) and that number seems to increase over the years. However, in 51 schools the Geography department performed statistically better than History department in 1998, with the relative number of "better" geography departments decreasing over the years. Table 4 also shows that when the adjusted indicator is used to compare departments we get similar results.

W also need to know whether the same department maintained their relative effectiveness over the other for the whole 1998-2002 period. This is a more difficult task since, as we argued earlier, test scores fluctuate much more from year to year, partly because of other factors that impact on the relative effectiveness. Although by considering large departments in our analysis we are able to reduce the importance of sampling variation, there might still be some other factors that produce non-persistent changes in relative performance over years that, given the data, we are not able to identify. We also do not know whether changes in relative effectiveness are the result of competition between History and Geography departments or other factors. In an attempt to see whether the better/worse relative performance of one department persists over time, we take our reduced sample of large departments that perform significantly better/worse and those that do not perform differently from the other department in 1998 and see whether there are any changes over the years. Results are presented in Table 5.

Place Table 5 around here

Table 5 shows the degree of persistence of the relative performance over time for both unadjusted and value-added within school competition indicators. Looking at this table, out of 65 schools where History was a better performing department than Geography only 6 schools retained this ranking throughout the period we are studying. A similar story is found in the case where the Geography department performed better, and also in those schools where there was not initially a statistical difference between History and Geography. This degree of non-persistence in relative performance could be because of some of the reasons that we discussed above.

In this section we have analysed the stability of relative departmental performance over time. Even after we took account of possible noise and the statistical significance of the differences, relative departmental performance as measured above is highly unstable over time. A noticeable additional and consistent finding has been that the value-added indicator, in general, provides a similar story regarding this instability. Our findings are therefore inconsistent with the hypothesis introduced above that relatively successful departments can sustain their leading position by cream-skimming students and staff. Overall, these results suggest two possible interpretations. Either relative departmental performance is being largely driven by random fluctuations in exam performance that we have not captured in either our unadjusted or value-added measures or competitive pressures within schools tend to ensure that differences in relative performance are not sustainable over time. This inability to retain an initial advantage may also be due to ceiling effects which assist the catching-up of lower performing departments or that managers concentrate upon the perceived weaker departments when seeking school improvements. We are unable to distinguish between these explanations in the analysis presented above.

6. Conclusions

We have explored the possible beneficial effects of promoting increased within school competition through increased optional choices in the secondary school curriculum. Potentially this could compensate for the reduced competition between schools resulting from the expansion of specialist schools in England. Our investigation was also motivated by the recent announcement of the intention to publicise school performance data disaggregated at departmental and teacher level. As the Secretary of State for Education explained this would mean that "if you have got a really poor French department, for instance, you will see that pupils between 12 and 13 aren't making sufficient progress in French. The school will need to do something about that

department – change the teaching staff, upgrade their professional development, bring in external support" (BBC, 2005). Our theoretical analysis examined the underlying rationale for such a reform. It suggested that greater competition within schools promoted by greater choice in the curriculum and supported by the provision of departmental and teacher performance indicators, could increase incentives for departments and teachers to raise the attainment levels of students taking their subjects.

Pupil and parental decision-making plays a key role in this theoretical rationale. A crucial requirement is that current departmental relative performance provides a reliable signal of departmental performance in the future when these pupils currently making optional choices actually take their external examinations. Our empirical analysis of English secondary schools indicated that this signal was unlikely to be reliable, regardless of the measure of departmental performance taken. Current relative performance, both unadjusted and value-added measures, was a generally poor indicator of relative performance in three years time. This instability of relative departmental performance suggests that increased choice within schools would be unlikely to create significant additional incentives for departments and teachers to raise their performance to retain or expand their recruitment. In other words identifying a "really poor French department" or trying to uncover the characteristics of a successful department may be a fundamentally flawed exercise.

The limited nature of our conclusions needs clarification. Increased within-school choice may still raise average educational attainment through alternative processes, such as a better match between the curriculum and pupil abilities, interests and aspirations. Similarly, well-designed disaggregated performance indicators may further encourage school managers and teachers to target those outcomes publicised in the performance tables and hence raise average measured attainment. However, our analysis suggests that current relative departmental performance is a poor indicator of future relative performance, as a consequence the potential dysfunctional effects of publicising such an indicator are likely to be significant.

Acknowledgement

An earlier version of this paper was presented at the European Conference on Education Research in Dublin, September 2005. We are grateful for the helpful comments received from participants at that conference.

References

Adey, K. & Biddulph, M. (2001) The Influence of Pupil Perceptions on Subject Choice at 14+ in Geography and History, *Educational Studies*, 27(4): 439-50.

Adnett, N., Bougheas, S. & Davies, P. (2002) Market-based Reforms of State Schooling: Some unpleasant dynamics, *Economics of Education Review,* 21, 323-30.

Adnett, N. & Davies P. (2002) *Markets for Schooling: an economic analysis*, London: Routledge.

Adnett, N. & Davies, P. (2005) Competition Between or Within Schools? Re-assessing school choice, *Education Economics*, 13(1), 111-23.

BBC (2005) *Teacher Ratings 'bad for morale*. Retrieved September 12 2005 from http://news.bbc.co.uk/1/hi/education/4235342.stm.

Bell, K. & West, A. (2003) Specialist Schools: an exploration of competition and co-operation, *Educational Studies*, 29(2/3), 273-89.

Bradley, S. & Taylor, J. (2002) The Effect of the Quasi-market on the Efficiency-equity Trade-off in the Secondary School Sector, *Bulletin of Economic Research*, 54(3), 295-314.

Bradley, S. & Taylor, J. (2004) The Economics of Secondary Schooling, in G. Johnes and G. Johnes (eds) *International Handbook on the Economics of Education*, Cheltenham, Edward Elgar.

Chay, K., McEwan, P. and Urquiola, M. (2005) The Central Role of Noise in Evaluating Interventions that use Test Scores to Rank Schools, *American Economic Review*, 95(4): 1237-1258.

Davies, P., Telhaj, S., Hutton, D., Adnett, N. & Coe, R. (2004a) The Myth of the Bog Standard Secondary School: a school level analysis of students' choice of optional subjects, *Institute for Education Policy Research, Staffordshire University, Working Paper* No. 2004/13

Davies, P., Telhaj, S., Hutton, D., Adnett, N. & Coe, R. (2004b) Subject Choice and Performance of Departments in Secondary Schools, *Institute for Education Policy Research, Staffordshire University, Working Paper* No. 2004/14.

Gibbons, S., Machin, S. & Silva, O. (2005) Competition, Choice and Primary School Performance, *Centre for Economic Performance, London School of Economics Working Paper*.

Goodson, I. (1994) *Studying Curriculum*, Buckingham, Open University Press.

Hanushek, E., Kain, J. & Rivkin, S. (2004) Disruption versus Tiebout Improvement: The costs and benefits of switching schools, *Journal of Public Economics*, 88(9/10), 1721-46.

Hansen, K. & Vignoles, A. (2005) The United Kingdom Education System in a Comparative Context, in S. Machin and A. Vignoles (eds) *What's the Good of Education?: the economics of education in the UK*, Princeton: Princeton University Press.

Hoxby, C. (2002) *The Economic Analysis of School Choice*, Chicago: University of Chicago Press.

Johnson, P. (2004) Education Policy in England, *Oxford Review of Economic Policy*, 20(2), 173-97.

Kane, T. & Staiger, D. (2002) The Promise and Pitfalls of Using Imprecise School Accountability Measures, *Journal of Economic Perspective*s, 16(4): 91-114.

Konstantopoulos, S. (2005) Trends of School Effects on Student Achievement: Evidence from NLS: 72, HSB: 82 and NELS: 92, *IZA (Bonn) Discussion Paper* No. 1749.

Mangan, J., Pugh, G. & Gray, J. (2005) Changes in Examination Performance in English Secondary Schools over the Course of a Decade: Searching for patterns and trends over time, *School Effectiveness and School Improvement*, 16(1), 29-50.

Nye, B., Konstantopoulos, S. & Hedges, L. (2004) How Large are Teacher Effects?, *Educational Evaluation and Policy Analysis*, 26(3), 237-57.

Rivkin, S., Hanushek, E. & Kain, J. (2005) Teachers, Schools and Academic Achievement, *Econometrica,* 73(2), 417-58.

Stables, A. (1997) Perspectives on Subject Choice: the case for a humane liberalism in curriculum planning, *Journal of Curriculum Studies*, 29(2), 197-208.

Telhaj, S., Hutton, D., Davies, P., Coe. R. & Adnett, N. (2004) Competition Within Schools: Representativeness of the YELLIS Sample Schools in a Study of Subject Enrolment of 14-16 Year Olds, *Staffordshire University Institute for Education Policy Research Working Paper* No. 2004/11.

Tymms, P. (1992) The Relative Effectiveness of Post-16 Institutions in England (including Assisted Places Scheme schools), *British Educational Research Journal*, 18(2), 175-92.

Yang, M. & Woodhouse, G. (2001) Progress from GCSE to A and AS level: institutional and gender differences and trends over time, *British Educational Research Journal*, 27(3), 245-67.

Table 1: Relative Performance of History and Geography

Unadjusted Performance	(2)	(3)	(4)
PANEL A			
	Overall (between and within school) stability		**Within school stability**
	Freq.	Percent	*Percent*
Negative Difference (History-Geography)	989	44.6	*54.1*
Positive Difference (History - Geography)	1225	55.3	*61.4*
Total	2214	100.0	*57.9*

Relative Value-Added Performance

PANEL B			
	Freq.	Percent	*Percent*
Negative Difference (History - Geography)	1081	48.8	*59.1*
Positive Difference (History - Geography)	1133	51.2	*60.0*
Total	2214	100.0	*59.6*

Table 2: Overall Transition Probabilities

PANEL A

Relative Performance History_Geography – Unadjusted Performance

	Negative	Positive	Total
Negative Difference	475	317	792
%	59.97	40.03	100
Positive Difference	300	679	979
%	30.64	69.36	100
Total	775	996	1,771
%	43.76	56.24	100

PANEL B

Relative Performance History_Geography - Value-Added Performance

	Negative	Positive	Total
Negative Difference	585	287	872
%	67.09	32.91	100
Positive Difference	280	619	899
%	31.15	68.85	100
Total	865	906	1,771
%	48.84	51.16	100

Table 3: Comparison of Relative Performance between *t* and *t-3*

1998-2001

	Negative	Positive	Total
Negative	112	102	214
	52.34	47.66	100
Positive	89	140	229
	38.86	61.14	100
Total	201	242	443
	45.37	54.63	100
phi coeff=0.14			

1999-2002

	Negative	Positive	Total
Negative	109	80	189
	57.67	42.33	100
Positive	88	165	253
	34.78	65.22	100
Total	197	245	442
	44.57	55.43	100
phi coeff=0.22			

Table 4: Statistical significance of difference in performance between departments

	Indicator 1			Indicator 2		
Years	History Statistically better	Geography Statistically better	No statistical difference between depts	History Statistically better	Geography Statistically better	No statistical difference between depts
1998	65	51	148	67	61	136
1999	69	47	148	58	58	148
2000	60	40	164	60	54	150
2001	59	41	164	59	55	150
2002	58	35	171	61	53	150

Table 5 The Degree of Persistence of the Relative Performance Over Time

	Indicator 1					Indicator 2				
	1998	**1999**	**2000**	**2001**	**2002**	**1998**	**1999**	**2000**	**2001**	**2002**
History statistically better	65	33	18	9	6	67	28	13	5	4
Geography statistically better	51	18	9	4	3	61	30	12	7	7
no difference	148	92	64	43	34	136	87	53	32	22