www.defra.gov.uk

Post – 16 Educational Choices in Rural Areas

A Report Prepared for the Department for Environment, Food and Rural Affairs by the Department of Economics, University of Sheffield

31 December 2009

department for
children, schools and families

defra
Department for Environment
Food and Rural Affairs

Further information about this publication is available from:

rural.research@defra.gsi.gov.uk

Published by the Department for Environment, Food and Rural Affairs
Nobel House
17 Smith Square
London SW1P 3JR

Tel: 020 7238 6000
Website: *www.defra.gov.uk*

Acknowledgements

This research was funded by the Department for Environment, Food and Rural Affairs (Defra) and is published jointly with the Department for Children, Schools and Families (DCSF). The data from the Youth Cohort Study (YCS) and the Longitudinal Study of Young People in England (LSYPE) were made available by the ESRC Data Archive at the University of Essex. We are also extremely grateful to Clare Baker, Michael Greer and Helen Wood at DCSF, and Pippa Gibson at Defra for the hard work they put in to supply us with required additional information. Finally, Tom Stafford and Robin Barnett at Defra, and Helen Wood, Michael Greer and Steve Gill at DCSF provided useful comments that improved this final version.

Andy Dickerson and Steven McIntosh
Department of Economics
University of Sheffield

Glossary of Education Terms

Key Stage 1:	education in reception year until Year 2 (age 4-7).
Key Stage 2:	education in Years 3-6 (age 7-11).
Key Stage 3:	education in Years 7-9 (age 11-14).
Key Stage 4:	education in Years 10 and 11 (age 14-16).
Primary Education:	education from ages 4-11 (Key Stages 1 and 2).
Lower Secondary Education:	education from ages 11-16 (Key Stages 3 and 4).
Upper Secondary Education:	education from ages 16-18.
Compulsory Education:	education until the end of the lower secondary phase (age 16).
Post-compulsory education:	education after the end of the lower secondary phase.
GCSEs:	public examinations taken at the end of compulsory schooling (Key Stage 4, the end of the lower secondary phase).

Executive Summary

Introduction

- Education and training beyond the compulsory stage (age 16) is increasingly regarded as important for the country's growth prospects, as well as the future labour market success of young people.
- Though current post-compulsory participation rates are high, with around four fifths of young people staying-on in post-compulsory education or training, this still leaves one in five who leave at age 16.
- The compulsory education and training participation age will be raised to 18 in 2015. While recognising that this raising of the participation age (RPA) policy also covers post-16 training, it is useful to understand what factors currently influence the decision to remain in post-compulsory education after the age of 16.
- Little is known about the influence of living in a rural location on the decision to participate in post-compulsory education.
- This study aims to document the difference in education staying-on rates between rural and urban areas, and determine the extent to which such differences can be explained by differences in the characteristics of the two types of areas and the people who live there.

Literature Review

- Existing literature on the post-compulsory participation decision has identified prior attainment and family background as key determinants of the decision to participate in post-compulsory education.
- There is some international research on rural-urban differences in post-compulsory participation, but any comparisons are hampered by differences in the definition and nature of rural areas in different countries. As far as the UK is concerned, no systematic evidence on this issue is available.

Data and Methods

- We use data from two cohorts of young people, namely the Youth Cohort Study (YCS, Cohort 12), who completed their compulsory schooling in 2003, and the Longitudinal Study of Young People in England (LSYPE), who completed

compulsory schooling in 2006. Both studies surveyed their respondents in the year immediately following the completion of compulsory schooling. To maintain consistency across data sets, only respondents living and attending school in England were considered in each case.

- We use multivariate econometric techniques to analyse the factors associated with participation in post-compulsory education and to determine the impact of living in a rural area, once other factors are controlled for (held constant).

Key Findings

- Looking at the raw data, participation rates in post-compulsory education are slightly higher in rural areas than in urban areas, by 3 percentage points in both data sets.
- Young people living in rural areas have various advantages compared to urban young people that might explain the higher participation in rural areas. In particular, young people in rural areas have significantly higher prior attainment at GCSE level. They are also more likely to have better-educated parents in more senior occupations, and to live in less deprived areas. Set against such advantages is the fact that young people in rural areas typically live further away from the nearest education institution than young people in urban areas and this greater distance may discourage participation. They are also less likely to be from ethnic minorities, who have higher participation rates.
- The results using the YCS data set show that, even after controlling for a range of individual, family and regional characteristics, young people in rural dispersed areas and rural village areas are significantly more likely to participate in post-compulsory education than young people in urban areas with identical characteristics. The magnitude of the differential is around 3 percentage points, suggesting that, relative to young people in urban areas, the positive and negative influences on participation approximately balance out. There is no difference in participation between young people in rural towns and urban areas.
- This conclusion is reinforced by a decomposition analysis, which shows in the YCS data that the observed characteristics cannot explain the small difference in participation rates between rural and urban areas, and which is therefore left unexplained and attributed to ‘ruralness’.

- These findings are not replicated in the LSYPE data set, however. In the estimated equations, rural-urban differences are statistically insignificant after controlling for other determinants of participation, and when the raw difference is formally decomposed, over half of this difference is 'explained' by the model. To a large extent, therefore, the LSYPE results suggest that the higher participation in rural areas is mostly due to the beneficial characteristics of the people who live there.
- In the YCS data, the participation gap between rural and urban areas is greater when attention is focussed on only academic participation. Young people in all types of rural areas (rural towns, villages and dispersed areas) are 5-6 percentage points more likely to participate in academic post-compulsory education than young people with identical observed characteristics in urban areas. There is some, weaker, evidence that individuals in rural areas are also less likely to participate in vocational post-compulsory education than individuals in urban areas.
- Again, these results are not replicated in the LSYPE data set, with no statistically significant differences in participation rates by type of education across areas being observed.
- Moving into the second year of post-compulsory study, a positive participation differential remains between rural dispersed and urban areas only, even after controlling for observed characteristics.
- There is some evidence that distance from nearest education institution has a small negative effect on post-compulsory education participation. When participation is differentiated between academic and vocational components, the negative impact of distance is stronger for participation in academic education.
- When considering the difference in findings between the YCS and LSYPE data, it is important to remember that there are differences in the sampling and data collection methodology between the two datasets which may potentially account for the difference. In addition, both are sample surveys, and while considerable care is taken to ensure their reliability, there will inevitably be sampling variability across data sets.

Table of Contents

List of Figures

List of Tables

1. Introduction

As the demand for skills in the labour market rises due to technological change and global competition, there has been increased emphasis on improving the skills of the UK population. As far as labour market entrants are concerned, this has manifested itself in a desire to raise post-compulsory education and training participation rates, so that as many young people as possible receive upper secondary education. The Leitch Review[1] argued that increasing qualifications at every level is necessary if the UK is to achieve its 2020 ambitions for jobs, productivity and skills. The 2020 targets include increasing the proportion of adults with at least Level 2 qualifications (lower secondary) to above 90%, and the proportion of adults qualified to at least Level 3 (upper secondary) to 68%. In this context, the compulsory education and training participation age is to be raised to age 17 by 2013, and to age 18 by 2015.

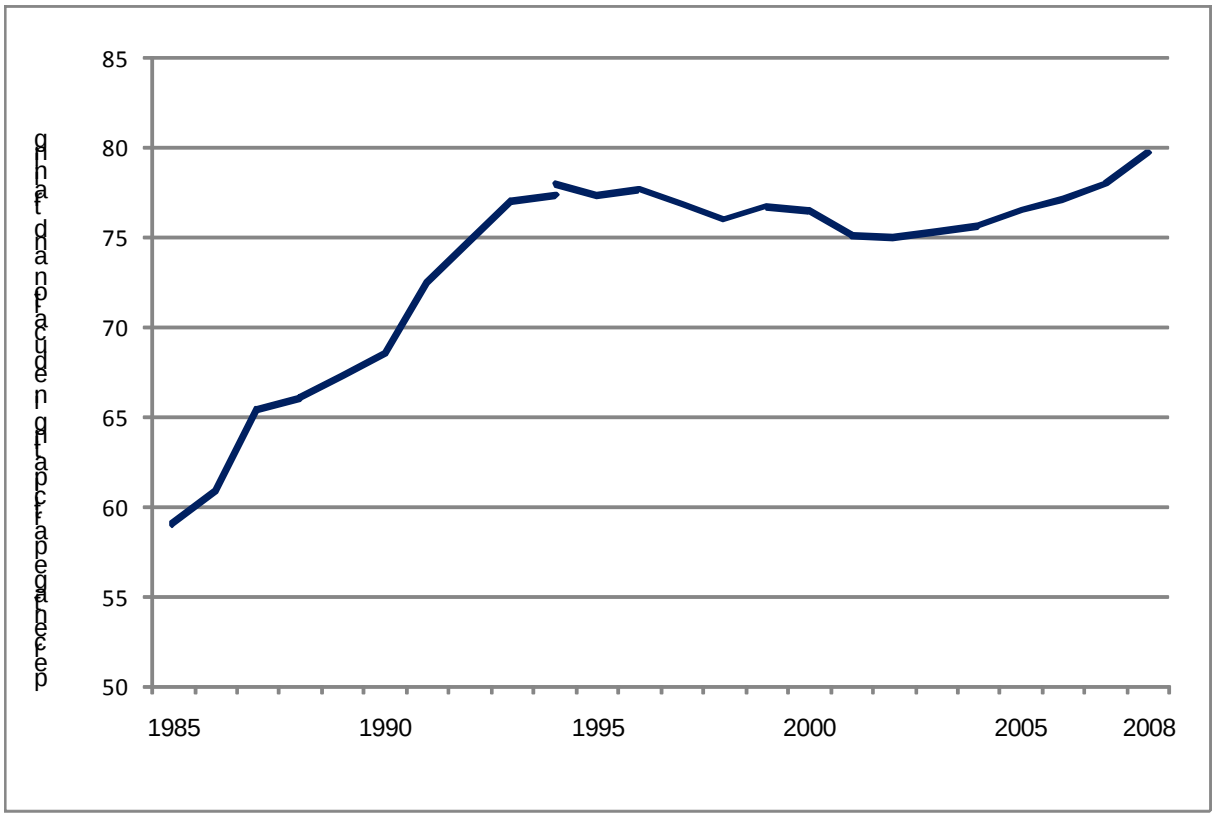
Figure 1 below displays the 16-18 year old participation rate in England for 1985 to 2008, and shows that there was a sharp increase in post-compulsory participation amongst 16-18 year olds in the second half of the 1980s and the first half of the 1990s. While the rate was then relatively static over the second half of the 1990s, there has been a renewed upward trend in the participation rate over the last 5 years, and the most recent (provisional) data for 2008 reveal a record level of participation of 79.7%.

This project utilises two cross-sectional data sets to examine post-compulsory participation across various areas of England in recent years. The aim of this is to document whether post-compulsory participation rates vary across areas, in particular examining whether there are any rural-urban differences in such participation. As well as documenting any such differences, the project also aims to explain them, in terms of observable characteristics of young people in each type of area, the characteristics of their families, and the characteristics of the regions themselves.

---

1 *Prosperity for All in the Global Economy – World Class Skills. Final Report of the Leitch Review of Skills*. London: HM Treasury, 2006. See also, *Ambition 2020: World Class Skills and Jobs for the UK*, UK Commission for Employment and Skills, UKCES, April 2009.

Figure 1: Participation of 16-18 year olds in education and training, England



Source: http://www.dcsf.gov.uk/rsgateway/DB/SFR/s000849/index.shtml, Table A13.

Notes:
1. The break in time series in 1994 is due to changes in the source of further and higher education data.
2. Data for 2008 are provisional.

The results of the analysis will show whether there are any problems with access to post-compulsory education in rural areas, by comparing participation in such areas with participation in urban areas, both in the raw data and when we compare individuals in the two areas with the same sets of observable characteristics. We can therefore say whether participation in rural areas is where we would expect it to be, given the characteristics of the people who live there. This information will help to inform the policy base for the future raising of the participation age (RPA), and also should be of use in monitoring progress towards PSA targets 10, 11, 14 and 16 which are concerned in particular with raising both participation and achievement in education and training of all children and young people[2].

[2] More specifically, these are: PSA10. Raise the educational achievement of all children and young people; PSA11. Narrow the gap in educational achievement between children from low income and disadvantaged backgrounds and their peers; PSA14. Increase the number of children and young

The next section of the report provides a brief literature review of this field, to reveal the variables that have been shown in previous studies as being able to explain variation in post-compulsory participation, either over time or across individuals. Section 3 describes the two data sets used in this study, with Section 4 providing the results of the study. A final section offers a summary of the findings and some conclusions.

people on the path to success; and PSA16. Increase the proportion of socially excluded adults in settled accommodation and employment, education or training. See http://www.hm-treasury.gov.uk/pbr_csr07_psaindex.htm for further information.

2. Literature Review

2.1 Introduction

A significant body of research has now emerged that considers the factors associated with the decision to remain in education beyond the current compulsory period. However, as will be seen, little of this work directly considers differences in participation between rural and urban areas. Previous research has typically used time-series data to explain changes in participation in post-compulsory education over time, or has used cross-sectional data to explain differences in participation across individuals at a particular point in time, and to identify the various factors associated with participation.

2.2 UK Time-Series Studies of Participation in Post-Compulsory Education

Some of the earliest studies of the factors associated with the decision to participate in post-compulsory education in the UK are time-series analyses at the aggregate level, in which the country's aggregate post-compulsory participation rate is modelled as a function of changing factors over time. Pissarides (1981) considers the period 1955-1978, estimating separate equations for the proportions of boys and girls remaining at school beyond the age of 16. He suggests that education is both a consumption and an investment good, and so predicts that the post-compulsory education rate will be positively affected by the level of consumer expenditure in the economy at each point in time, and the likely returns to education in terms of the wage premium received by graduates. These predictions are confirmed for both genders, with the wage effects being larger for boys. In addition, the level of adult unemployment has a positive influence on participation for boys, though not for girls.

Whitfield and Wilson (1991) study a different, though overlapping, time period of 1956-1985, and include some additional explanatory variables. They again show that the decision to continue in education beyond the age of sixteen is positively related to the rate of return to obtaining a degree and the adult unemployment rate, with the wage effects again stronger for boys. Of the new variables considered by

Whitfield and Wilson, social class, as measured by the proportion of the adult population in white-collar jobs, has a positive effect, and has the largest effect of all the variables on the decision to remain in education beyond the age of 16. On the other hand, a variable measuring success rates in GCE examinations does not appear to have any significant impact on participation rates. However, a similar analysis by McIntosh (2001), for the period 1961-1994 and thus including the switch to GCSE examinations in 1988, does find an important role for prior attainment in Year 11 examinations. Similarly McVicar and Rice (2001) consider time series data into the 1990s, and reveal the importance of prior attainment, stating that the introduction of GCSEs in 1988 and the consequent rise in attainment rates was the key determinant of the rise in post-compulsory participation at that time.

2.3 UK Cross-Section Studies of Participation in Post-Compulsory Education

A much larger literature exists that uses individual level cross-sectional data, rather than aggregate level time-series data, to study the factors associated with post-compulsory education participation. In the UK, the best data source for examining the participation decisions of individuals has been the Youth Cohort Study (YCS), and these data will also be used in the current report. The YCS has surveyed a succession of cohorts, in most cases at three points in time each, when the respondents are aged 16-17, 17-18 and 18-19, and is thus ideal for studying post-compulsory participation. Using these data sets, a number of reports were published in the 1990s examining a variety of issues, including participation.

Gray *et al.* (1993) describe the individual characteristics that they find to be significantly associated with post-compulsory education participation. Females, those whose family belongs to a higher social class and non-whites are all found to be significantly more likely to continue in education. However, the individual characteristic which completely dominates all these others is examination success at age 16; the better the individual performs in his or her GCE/GCSE examinations, the more likely he or she is to continue in education. The positive association between prior attainment and the post-compulsory participation rate could, of course, be a

reflection of ‘reverse causality’, whereby students who intend to stay on after the end of compulsory education work harder in school in order to be able to do so. Certainly, part of the observed positive relationship between attainment and post-compulsory participation will be attributable to this factor, rather than it all being due to the impact of ability, family background and schooling or other factors affecting attainment and thereby post-compulsory participation.

Gray *et al.* (1993) also consider the characteristics of the local labour markets in which the respondents live. They show that a lot of the difference in participation rates across local labour markets is the result of the characteristics of the individuals who live within them. Over and above this, however, the unemployment rate in a local labour market is significantly associated with participation, with a somewhat surprising negative effect. It seems likely that this variable is picking up the effects of more general social conditions in local labour markets, rather than saying that high local unemployment itself influences young people to leave school. The authors do not consider the urban/rural nature of local areas. Gray *et al.* (1994) replicate this unemployment result, and further include additional local labour market characteristics, such as industrial structure (the type of jobs locally available being hypothesised to influence the decision whether further education is required). They find only weak effects, however.

Further studies using the YCS have developed some of these findings. Ashford *et al.* (1993) examine whether the introduction of the GCSE examination could explain the large increase in participation that occurred in the late 1980s in the UK. They find that the new system did lead to an increase in the proportion of candidates obtaining good grades, which in turn leads to an increase in the probability of continuing in education. This was not the only cause of the increase in participation, however, since the authors point out that there was also an increase *within* all qualification groups.

Cheng (1995) considers the influence of school characteristics on the decision to remain in education after reaching the minimum school leaving age. He finds that the propensity to stay on is higher among students who have attended voluntary aided, rather than Local Education Authority maintained, schools. The size of the

school's sixth form (positively), and the degree of teacher turnover and the proportion of students eligible for free meals (both negatively) all also have a significant impact on a school's pupils remaining in post-compulsory education. Collecting their own data on 24 schools, Foskett *et al.* (2008) similarly identify the size, or presence, of a sixth form, and the socio-economic status of families as important in influencing post-compulsory participation. In addition, they refer to the importance of a school's career service, and the general ethos of the school.

Another issue considered by Foskett *et al.* (2008), that we have not mentioned yet, is whether individuals participate in academic or vocational education, once they have decided to continue their education. Their qualitative evidence suggests that schools in higher socio-economic status areas steer their pupils towards the academic stream, whereas the vocational route is more likely in schools in lower socio-economic status areas, suggesting that family background and aspirations are important.

This academic/vocational distinction is the main focus of research by Conlon (2005). He uses National Child Development Study (NCDS) data to suggest that, at least at higher qualification levels above Level 2, those young people who performed better in childhood reading and maths tests are more likely to take academic rather than vocational qualifications in the post-compulsory phase. Other factors that appear important include family background, and region of residence, with those living in London and other southern regions being more likely, other things equal, to undertake academic qualifications. Lenton (2005) explicitly tests whether the determinants of participation in academic and vocational education should be analysed separately, or whether they can be collapsed into a single overall participation rate. She finds that the choice of studying academic and vocational qualifications should be treated separately. As previously shown in the literature, the key determinants of choice of route are prior attainment and family background. Lenton suggests that one of the key reasons for the growth in participation has been the increased availability of vocational education, which has opened up post-16 education to those who achieved in the lower-middle range of GCSE attainment. Clark (2002) uses a different type of analysis, in particular producing a regional panel to analyse regional participation rates. His results suggest that when youth

unemployment rates are measured at a local level, rather than the national average as in the time series studies of Section 2.2, their positive effects on post-compulsory education participation are significantly larger[3]. Clark allows for the academic/vocational division by estimating a sequential model, whereby individuals decide first whether they want to undertake academic qualifications or not. If not, they then decide whether to study for vocational qualifications, or to look for a job. Clark's results show that the positive influence of the local youth unemployment rate is only present for the latter decision. It therefore appears that those on the academic track are always going to pursue this, given high prior attainment. However, those considering vocational qualifications are influenced by the state of the local labour market, and are more likely to study when jobs are scarce.

Clark (2002) therefore gives prominence to regional characteristics in his explanation of the rising post-compulsory participation rate. Other studies have similarly considered regional effects, though none in the UK have focussed on the rural-urban definition of an area. Rice (1999) also studies the impact of local labour market conditions, and agrees that the positive effect of local unemployment on the participation decision is strong for those young people who have low levels of prior GCSE attainment, though those with higher levels of prior attainment are not influenced. Her results also show that the labour market effects are stronger for boys than for girls. Similarly, Payne (1998) observes that regional effects are much stronger for those with lower prior attainment. Thus, she reports that the post-compulsory participation rate is a third higher in Greater London than in the North, with most of this difference caused by young people with average or poor GCSE results. Finally with respect to regional effects, though primarily focussed on the issue of migration, a recent paper by Bailey and Livingston (2008) describes how educational attainment is lower in more deprived neighbourhoods, a situation that is exacerbated by those who do achieve well in a deprived area tending to migrate out to more affluent areas.

[3] This positive effect of local unemployment contrasts with the negative effect observed by Gray *et al*. (1993, 1994). The explanation for the difference is that while the Gray *et al.* studies utilise cross-sectional data, Clark has panel data. With cross-sectional data (i.e. data recorded at a single point in time), local unemployment effects are masked by their correlation with the socio-economic characteristics of an area. However, when a time series element is introduced, as in panel data, then variation in local unemployment rates over time *within* areas reveals the positive impact of local unemployment on post-compulsory participation.

Before leaving this section, it is worth briefly mentioning that there is a large literature on intergenerational mobility which, while not having as its focus a full description of the determinants of post-compulsory participation, does nevertheless highlight the importance of family background. The fact that parents from a higher socio-economic status background are more likely to have children who participate in post-compulsory education is one cause of the positive intergenerational correlation in economic status. Blanden and Gregg (2004) consider the impact of parental income on a range of educational outcomes, including participation in post-compulsory education, and find a strong positive relationship. They then consider whether this is a direct causal mechanism, or whether there are more complex relationships at work that involve other variables. In other words, did the income that their parents had *directly* affect their children's educational outcomes, or is parental income correlated with other factors, such as parental education, parenting style, genetic ability etc, with these other factors being the actual determinants of the children's higher achievement? Blanden and Gregg use a range of econometric techniques to identify the causal mechanisms, and conclude that income itself does have a direct role to play. Similarly, Blanden and Machin (2004), though focussing on Higher Education rather than post-compulsory education, identify the importance of parental income for children's outcomes, and show that this intergenerational relationship has become stronger over time, so that parental background is becoming increasingly important. Micklewright (1989) also considers family background, and uses NCDS data to control for often unobserved characteristics of the children and their environment such as ability (measured by test scores) and school type, the omission of which could bias the relationship of interest. Micklewright shows that even after controlling for such factors, parental income is still positively related to the children's decision whether or not to participate in post-compulsory education.

2.4 Rural-Urban Differences in Post-Compulsory Education Participation: The Education Maintenance Allowance

Despite the existence of all of the UK-based research evidence described above on the decision to participate in post-compulsory education, we are aware of no study

which takes as its main focus the differences in such participation between rural and urban areas. Most do not even include rural-urban measures amongst the control variables, and thus they typically do not even indirectly consider this issue. An important exception is the analysis of the impact of the Education Maintenance Allowance (EMA), which was deliberately trialled in distinct rural and urban areas. Therefore, some evaluations of the EMA did examine differences in effects between the two types of area.

The EMA is a direct payment to young people aged 16-18 who remain in post-compulsory education, and whose parents earn less than some threshold amount[4]. It is therefore a direct attempt to reverse the effect observed in the studies described above, that individuals from less well-off backgrounds are significantly less likely to participate in post-compulsory education. The EMA was originally piloted in 10 Local Education Authorities in September 1999, allowing a formal evaluation of its impact by comparing outcomes in the pilot areas to outcomes in control areas where the EMA was not introduced. Of particular interest here is that the pilot areas were explicitly defined as either rural or urban. However, in the end there was only one rural pilot area included (Cornwall), with data collected from two rural control areas where the EMA was not introduced. It is therefore not clear whether any rural results obtained in the evaluations of the EMA are specific to Cornwall and perhaps a result of particular characteristics of that area, or whether they can be generalised to all rural areas.

The impact of the EMA on the participation rates by rural-urban area in the first year of post-compulsory schooling is best described in the second of the formal quantitative evaluations of the EMA pilots (Ashworth *et al.*, 2002). The methodology used was propensity score matching, whereby individuals in the EMA pilot areas were 'matched' with individuals in the control areas who looked most like them in terms of their observable characteristics, and then their participation rates were compared. The authors also matched on characteristics of the local areas in which individuals lived. All of the analysis was conducted only on those young people who

[4] For the academic year starting September 2009, the full EMA payment of £30 per week is paid to 16-18 year olds in full-time education whose household income (in the tax year 2008-09) is below £20,817, with the entitlement tapering with increasing household income (reducing from £30 to £20 to £10 per week) until being removed completely for household incomes above £30,810.

were eligible for the EMA, according to the household income criterion. The participation rates in the table below are therefore representative of the participation rates for the young people in this low income group in the chosen areas, rather than for all young people in these areas.

Table 1: Year 12 participation rates of matched EMA-eligible young people in EMA pilot and control areas

	Year 12 participation rates (%)		
	EMA Pilot Areas	Control Areas	Difference
Urban males	67.4	60.6	6.9
Urban females	72.3	67.5	4.8
Rural males	81.3	74.2	7.1
Rural females	85.7	80.5	5.2

Source: Ashworth *et al.* (2002), Tables 2.3-2.6.

The results in the table show that amongst young men in urban areas, those eligible for the EMA in the pilot areas are 6.9 percentage points more likely to participate in full-time education in Year 12 than those young men in the control areas who would have been eligible had the EMA been available in their area. The second row of results shows that young women in urban areas have a higher post-compulsory participation rate than young men. We might therefore expect the impact of the policy to be smaller for women given the higher numbers already participating in post-compulsory education in the absence of the policy, and this is precisely what the results show. The lower two rows show that participation rates in rural areas are significantly higher than in urban areas, for both males and females. We might therefore have expected the impact of the EMA policy to be lower in rural areas. This is not the case however, and the EMA raises participation in rural areas by slightly more than in urban areas for both males and females. Even when the authors estimate the impact of EMA for the full population in each type of area rather than just for the eligible population, and so take account of the fact that there is a lower proportion of low income families in rural areas as compared to urban areas to take advantage of the policy, the impact of the EMA is still larger in rural areas – EMA increases participation by 4.6 percentage points in rural areas as compared to 3.5

percentage points in urban areas (Ashworth *et al*., 2002, Table 2.8)[5]. Thus there is scope for participation to increase in both rural and urban areas, and payments to young people in low income families can help to improve participation in both types of area.

2.5 International Evidence on Rural-Urban Differences in Post-compulsory Education Participation

As noted above, there is a paucity of evidence on rural-urban differences in post-compulsory education participation for the UK. The situation is similar when we search for internationally comparative studies. Thus, while there is some research evidence on post-compulsory participation in a range of countries, this seldom has a rural-urban focus.

McIntosh (2001) examines changes in post-compulsory education participation over the last 20-30 years in five European countries, namely England and Wales together, Germany, Netherlands and Sweden. The most notable feature is the large international differences in participation and while England and Wales have been 'catching up' over the last two decades or so, their post-16 participation rates are still considerably lower than in the other countries. While this is an aggregate time-series study, the factors which influence participation, such as youth unemployment, clearly have a rural-urban dimension and thus could be informative for the current study. However, the key explanatory variable in determining post-compulsory participation in all countries is prior education attainment (i.e. achievement before the end of compulsory education), especially for females. This is a common finding. For example, Robertson (1990) for Scotland; Roussel (2000) and Le and Miller (2005) for Australia; and Yan (2003) for the US, all find that prior attainment is the most significant determinant of post-compulsory participation.

One paper which does explicitly examine the impact of rural-urban residence on participation is that by Le and Miller (2005). They find that students from rural areas

[5] The fact that the gap in the rural-urban impacts is now even wider suggests that the EMA affected the ineligible population, via spillover effects, differently in the two types of area.

in Australia have a lower probability of completing high school and therefore a lower probability of continuing in post-compulsory education. The most significant difference between students from rural and urban areas is the proportion of students who fail to finish high school in the rural areas, which means that they only complete at most 10 years of schooling[6]. The influence of students' expectations and aspirations in determining Australian post-compulsory participation is noted by James (2002). He shows that these expectations and aspirations are, in turn, affected by socio-economic background and rural-urban location. Students living in more rural and isolated areas and those from lower status socio-economic backgrounds attach a much lower priority to completing Year 12 of school and, consequently, they are less likely to participate in post-compulsory education. This 'poverty of aspirations' amongst certain groups is also likely to be important when attempting to understand differences in post-compulsory education participation in the UK.

Of course, there is an important caveat when considering international comparative work of this kind in terms of its relevance to the situation in the UK. Rural-urban differences in participation in post-compulsory education are likely to be influenced by the country-specific institutional environment, as well as the social structures, constraints and expectations, and thus any differences in rural-urban participation need to be considered in context. As a consequence, any generalisations beyond the specific country in which the study is undertaken need to be made with considerable caution.

2.6 Rural-Urban Differences in Post-compulsory Education Attainment

While this project is primarily focussed on developing our understanding of rural-urban differences in post-compulsory education *participation*, it is important to recognise that post-compulsory education participation may be influenced by students' expected *attainment*. Any rural-urban differences in education achievement may thereby influence rural-urban education participation rates – the incentive to participate in post-compulsory education in some areas may be reduced if students expect that their achievement is likely to be poor. Hence it seems important to briefly

[6] In most Australian states, high school is completed after 11 years of schooling (preparatory year plus years 1-10).

consider the evidence on rural-urban differences in educational attainment. A more comprehensive assessment of this issue has been separately undertaken contemporaneously with this study (Green, 2009). She finds that attainment in rural areas is higher at Key Stage 3 and Key Stage 4, though the differences are small, and statistically significant only for Key Stage 3. When other factors are controlled for, any differences in attainment in rural and urban areas disappear, showing that any higher attainment in rural areas is due to higher social status, less local deprivation and higher prior attainment at Key Stage 2 (itself probably a function of social advantages even earlier in life).[7]

Prior to Green (2009), there was little by way of direct evidence on rural-urban differences in attainment. A number of studies have emphasised the importance of geography for educational attainment, but rarely focus on rural-urban status. Instead, the focus tends to be on the impact of an area's social and economic deprivation and well-being on education attainment (and potentially other outcomes). Thus, for example, Garner and Raudenbush (1991), Raudenbush (1993), McCulloch and Joshi (2001), and Leckie (2008) all find that having taken into account pupils' ability, family background and schooling, there is a negative relationship between deprivation indices and education attainment in their various studies of Scotland and England, although the relative importance of school versus neighbourhood effects differs considerably. However, only Leckie (2008) also considers the separate impact of whether the area is rural or not (although the definition of rural is not given) and finds there to be a significant positive residual impact of ruralness on attainment having taken account of the other factors which affect individuals' education attainment. However, in general, it is difficult to know if these are genuine rural effects generating differences in attainment between rural and non-rural areas, or whether these differences are a reflection of differences in unobserved characteristics of the families living in these areas. Gibbons and Silva (2008) cleverly attempt to disentangle these explanations by exploiting the fact that students change urban density of their education institution when they transit from primary to secondary schools. Correlating the differences in attainment between primary and secondary school with the change in urban density suggests that the lower attainment in dense urban areas in England

[7] Note that when attainment at Key Stage 2 is included amongst the explanatory variables, the analysis is effectively considering the *change* in attainment between Key Stage 2 and Key Stage 3/4.

is not because of any neighbourhood disadvantage affecting attainment, but because the most disadvantaged students (in terms of average attainment) attend the most urban schools.

Another important exception to the lack of direct evidence on the importance of rural-urban area classification is a recent paper by Ulubasoglu and Cardak (2007) which examines for a range of 56 developed and developing countries (not including the UK), the differences in rural-urban education attainment as measured by the ratio of the average years of schooling in rural areas and urban areas.[8] Rural years of education are, on average, only around half of those received by those in urban areas, but this overall mean disguises considerable variation between countries and over time. Richer counties as measured by per capita income have both higher and more equal participation rates in rural and urban areas. The rural-urban educational inequality ratio is then related to a range of factors which might be expected to impact upon individuals' objectives, opportunities and constraints (Haveman and Wolfe, 1995). Factors that are positively associated with lower rural education participation and greater rural-urban inequality are countries with a colonial past, lower political stability, with legal systems based on the French rather than the British legal system, those which are landlocked and/or are large, where investment in human capital is riskier and there is limited credit availability. One problem is that the range of countries is so diverse economically, politically, socially and culturally (running from Afghanistan to Zambia through France, Norway and the US) that it is difficult to know how to interpret the findings in terms of the factors that are likely to be important for the UK, particularly given the interrelationships between the factors which are found to be important. Moreover, few, if any policy prescriptions would appear to follow from this largely descriptive characterisation of the rural-urban participation differential.

[8] While this paper is couched in terms of attainment, the measure of attainment is actually based on years of participation in schooling, and so is arguably more directly related to the existing study than most other studies of attainment, for which post-compulsory participation is a pre-requisite for any post-compulsory attainment.

2.7 Summary

Much research has shown that education is important for economic outcomes, both at the national level – in terms of higher GDP, productivity etc – and at the level of the individual in terms of their earnings and job opportunities. There are also many wider benefits of having a more educated population in terms of better health, lower crime, greater community participation, financial inclusion etc. It is therefore important to understand the factors which determine differences in education participation (and attainment) between rural and urban areas, in order to help reduce any inequality in participation (and achievement).

In theory, the education participation decision is affected by three groups of interdependent factors: objectives; opportunities; and constraints (Haveman and Wolfe, 1995). Objectives are set at a national level in terms of funding for education versus competing demands, and also at the family level with respect to the value placed on education, and the benefits that it can bring. Opportunities to partake in education depend on the availability of local institutions delivering the desired programme of study. Opportunities to engage in alternative activities, principally working, also influence the education participation decision. Finally constraints to engaging in post-compulsory education include aptitude and ability, as well as financial constraints.

The available evidence surveyed above is largely consistent with this theoretical framework. Particularly with respect to constraints, studies have repeatedly shown that prior attainment and family socio-economic background are prime factors associated with the decision to participate in post-compulsory education. Opportunities for further study in terms of educational institution previously attended have also been shown to be important. Regional factors such as local unemployment rates are significant in a number of studies. What is less clear from the available evidence is how far these various factors are inter-related, with possibly multiple directions of causality running between them. For example, region of residence is not necessarily an exogenous variable, since location is often deliberately chosen for education or labour market reasons. In turn, therefore, location will be affected by socio-economic status and the financial ability to relocate

in desired areas. Similarly, prior attainment in compulsory schooling will be a function of upbringing and family background, as well as innate ability (which itself will be inherited to a certain extent and so dependent on one's parents). Finally, prior attainment in compulsory schooling can also be influenced by aspirations regarding future post-compulsory schooling participation.

In terms of rural-urban differences in participation, little research has been directed at this issue, and no UK study has this as its explicit focus of investigation. Some international research exists on rural-urban participation differences, but the extent to which this is relevant to the UK is not clear. The concept of a rural location in Australia, for example, is very different to a rural area in the UK. This project therefore examines differences in post-compulsory education between rural and urban areas, and attempts to account for these differences in terms of the various factors that have been identified in the research reviewed above as being relevant to explaining differences in participation. In particular, we examine whether any post-compulsory education participation differentials between rural and non-rural areas can be accounted for by differences in the characteristics of people living in rural and non-rural areas, or whether 'ruralness' *per se* affects individuals' participation.

3. Data

We utilise two major survey data sets in this project, namely the Youth Cohort Study (YCS) Cohort 12, and the Longitudinal Study of Young People in England (LSYPE). These are both briefly described in the following two sub-sections.

3.1 YCS

The YCS was designed primarily to provide information on young people's transitions from compulsory education to further and higher education, and/or the labour market. The YCS is a continuing series of representative cohort surveys which started with Cohort 1 in 1985, and has now reached Cohort 13 which was first interviewed in 2007. Each cohort is surveyed by postal questionnaire on a number of occasions (called 'sweeps'), with the first sweep in the Spring in the year after completing compulsory education when individuals are (mostly) aged 16. Individuals are then re-interviewed on an annual or biennial cycle, with most cohorts interviewed three times in total.[9] Coverage for Cohorts 1 to 12 is England and Wales, although a change in methodology from postal to face-to-face interviewing, driven at least in part by falling response rates, means that Cohort 13 covers England only.

We use the latest available full data set, which is Cohort 12 (YCS12). Survey participants were interviewed annually in 2004, 2005, 2006 and 2007. We focus mainly on sweep 1 which was carried out in Spring 2004 for individuals who completed their compulsory education 8 months earlier (i.e. were eligible to leave school for the first time in Summer 2003). Postal questionnaires were sent and, for the first time in YCS12, a web completion option was also provided towards the end of fieldwork (due to lower response rates than expected). Non-respondents were recontacted by telephone whenever possible in order to boost the response rate, which was just 47% in YCS12 sweep 1 (it was over 70% for cohorts which

[9] The exceptions are: Cohort 3 (mostly aged 16 at January 1987) which had a follow-up sweep when the respondents were aged 23 in 1994; Cohort 7 (mostly aged 16 in January 1994) which had just two sweeps at age 16 and age 18; Cohort 9 (mostly aged 16 at January 1998) which had a fourth sweep in Autumn 2000; Cohort 10 (mostly aged 16 at January 2000), which was surveyed twice (Spring and Autumn) in 2000; and Cohort 11 (mostly aged 16 in 2002) and Cohort 12 (mostly aged 16 in 2004) which both had four annual sweeps carried out at age 16, 17, 18 and 19.

commenced in the late 1980s/early 1990s). Consequently, in common with most other surveys, data from the YCS needs to be weighted in order to make it representative of the population of young people in England and Wales, since non-response is not random (and in particular, is much lower for those with poor educational attainment at Year 11).

The YCS collects data on education and labour market activity, qualifications gained and sought, details on current employment, as well some background socio-economic information about families and their attitudes. YCS12 sweep 1 was the first YCS sweep designed with the intention of collecting most Year 11 (i.e. GCSE) attainment information from administrative sources rather than asking the survey respondents themselves. Individuals were matched with their records in the National Pupil Database (NPD) which provides information on individuals' attainment and on their schools[10]. Since this administrative data source is only available for England, and the coverage of our other data set described below also covers England only, we restrict our attention to YCS respondents in England only.

3.2 LSYPE

The Longitudinal Study of Young People in England is also focussed on young people's transitions from education into the world of work. It is a single cohort study (which is still ongoing), tracking a sample of young people from age 13/14 (Year 9) in Spring/Summer 2004 in order to better understand their development from their early teens while still in education (as compared to the YCS which only starts post-16). Interviews (known as 'waves') are taking place annually, initially face-to-face at home addresses using Computer Aided Personal Interviewing (CAPI), and in later waves by telephone, or on-line, as well as face-to-face. Interviews are conducted with the young person and with parents and guardians living at the same address. The response rate at wave 1 was 74%, much higher than YCS, and at least in part because of the face-to-face survey methodology. We primarily focus on wave 3

[10] The linking of respondents to their individual attainment records and information on their schools in the NPD serves to reduce respondent burden as well as to increase the accuracy of the information in the YCS. However, information on schools can only be obtained for maintained schools since independent schools are not in the NPD (although their pupils' attainment records will appear in the NPD when they complete their GCSEs).

conducted in 2006 to obtain data on the explanatory variables about the young people and their families. At this point, the respondents were (mostly) aged 16, and were coming towards the end of their compulsory education, or in some cases had just completed it. The post-compulsory education participation variable in LSYPE was derived from information on respondents' current activity in wave 4 of the survey, conducted in 2007, around one year after the end of their compulsory schooling.

The LSYPE questionnaires cover a broader range of topics than the (shorter) YCS questionnaire[11]. Thus, in addition to the areas covered in YCS, the LSYPE also includes: attitudes to school and involvement in education; parental expectations and aspirations; risk factors (absences, truancy, police contact, bullying) and a range of parental questions. We make use of these additional questions in the analysis we present in the next section. As with the latest YCS, individuals can be matched with administrative data sources, such as the NPD, which provides Key Stage test results as well as GCSE attainment etc.

The above descriptions of the two data sets to be used make clear that they are from two different points in time, being three education years apart – 2004 for YCS12 and 2007 for LSYPE. Various education policies were introduced or changed during these three years[12], meaning that the education system faced by respondents in LSYPE is not exactly the same as that faced by respondents in YCS12. Moreover, the aggregate post-16 participation rate was higher in 2007 than in 2004 as shown in Figure 1 above. Such differences must be borne in mind throughout this study, and could potentially explain some of the different results found across the two data sets as reported below.

3.3 Rural-Urban Definition

As the main focus of this report is on rural-urban differences in post-compulsory education participation, it is important to clarify the definitions of such areas at this juncture. The rural-urban definition utilised is an 8-point classification, based on

[11] See http://www.esds.ac.uk/longitudinal/access/lsype/L5545.asp for further information.
[12] For example, the introduction of the Education Maintenance Allowance (EMA), Entry to Employment (E2E), re-launched apprenticeships and the abolition of GNVQs.

types of settlement and population densities. Four types of area are defined; in declining order of density these are ‘urban’, ‘small town and fringe’, ‘villages’ and ‘hamlets and dispersed households’. For each of these types of area, the context in which they are found (i.e. the surrounding area) is defined as ‘sparse’ or ‘less sparse’, with the majority of settlements being defined as ‘less sparse’. These four types of area with two contexts within each type produce the 8-fold rural-urban definition. For both surveys, the rural-urban definition was matched in to the data set by the postcodes of the individual respondents.

Table 2 below records the weighted proportion of individuals in each of our two data sets that live in each of the eight area types.

Table 2: Allocation of survey respondents to type of area

	YCS (%)		LSYPE (%)	
	Less sparse	Sparse	Less sparse	Sparse
Urban	79.4	0.1	80.1	0.1
Rural town	9.6	0.6	8.7	0.7
Rural village	6.6	0.4	6.8	0.4
Rural dispersed	2.8	0.4	2.9	0.5
Total	98.5	1.5	98.4	1.6

As this table makes clear, the vast majority (around 80%) of the population live in urban areas, with the proportions in the various categories of rural areas declining, as expected, with the size of such settlements. In addition, the figures show that only a very small proportion (around 1.5%) of the population lives in areas set in a ‘sparse’ context. Comparing across data sets, the proportions in each area type are very similar, with slightly more respondents in urban areas, and slightly fewer in rural towns, in LSYPE as compared to the YCS.

The focus of the report throughout will be on rural-urban differences in participation. At various points in the report, the six types of rural area will be identified separately, whilst at others, rural towns, rural villages and rural dispersed areas will be considered as three separate groups. Sometimes, all rural areas will be aggregated

together so that a simple rural-urban dichotomy will be used. The groups being used will always be made explicit.

4. Results

4.1 Descriptive Statistics

Table 3 shows how post-compulsory education participation rates vary according to the 8-fold rural-urban definition as described in Section 3.3 above.

Table 3: Post-compulsory participation rates by type of area

	YCS (%)			LSYPE (%)		
	Less sparse	Sparse	Total	Less sparse	Sparse	Total
Urban	74.2	62.1	74.1	71.1	87.8	71.1
Rural	77.2	68.1	76.6	74.2	74.5	74.2
Rural town	*73.5*	*62.3*	*72.9*	*71.4*	*69.7*	*71.3*
Rural village	*79.2*	*79.0*	*79.2*	*77.3*	*67.2*	*76.8*
Rural dispersed	*85.0*	*64.0*	*82.5*	*75.3*	*87.0*	*77.0*
Total urban and rural	74.8	67.6	74.7	71.6	75.2	71.7

Participation rates differ across the different area types, and also between data sets. However, recall from Section 3.3 that the sample sizes in the 'sparse' areas are extremely small, so that we may expect rather imprecise estimates of participation rates in such areas. Focussing only on the 'less sparse' areas (which account for around 98.5% of respondents as shown in Table 2 above) the two data sets both record that participation rates are higher in rural village and rural dispersed areas, compared to rural town and urban areas which have very similar participation rates.

In each of the four 'less sparse' areas, the estimated post-compulsory education participation rate in YCS is higher than in LSYPE, despite the aggregate participation rate being higher at the time that LSYPE was surveyed. A potential explanation of the lower participation rate in LSYPE compared to YCS is that most LSYPE respondents are surveyed in July or August, whilst most YCS respondents are surveyed in March. LSYPE respondents are therefore surveyed around four months later than YCS respondents, so allowing more time for course completion or dropout, and so lower participation, in LSYPE. Given that A-levels have a two year duration, so that

respondents in neither data set will have completed such courses during the first year of post-compulsory schooling, we might expect this difference in participation rates between surveys to disappear when we focus exclusively on participation in A-level study. This is exactly what is observed, with the A-level participation rate in LSYPE actually slightly higher than the equivalent rate in YCS (47.6% in LSYPE versus 47.0% in YCS).[13]

Table 4 below reports descriptive statistics for all of the variables used in the analysis from both data sets, presented separately for respondents living in rural and urban areas. The first row shows the participation rates in full-time post-compulsory education[14] in the first year after the end of compulsory schooling (age 16/17). In both data sets, young people living in rural areas are more likely to be participating in full-time education than those in urban areas, although the differences are not large. In the YCS, the participation rates are 77% and 74% in rural and urban areas respectively (the difference being statistically significant), with the equivalent figures in LSYPE being 74% and 71% (with the difference again statistically significant).

The remaining rows of Table 4 show the range of explanatory variables available in the two data sets, and their sample means and proportions. Panel A shows the variables available in the YCS, together with equivalent variables constructed from the LSYPE data. In Panel B the wider range of additional variables available only in LSYPE is shown.

[13] The participation rate gap between data sets is larger for the 'rural dispersed – less sparse' category, where the estimates are particularly hampered by small sample sizes and so are less robust.

[14] It should be made clear that only full-time participation in an educational institution such as school or Further Education college is being considered here. This definition of participation therefore differs from the DCSF's definition of participation as far as raising the participation age (RPA) to age 18 is concerned since their definition also includes part-time learning while in full-time employment, training in firms, apprenticeships etc.

Table 4: Sample means and proportions for YCS and LSYPE data sets

	YCS		LSYPE	
	Rural	**Urban**	**Rural**	**Urban**
PANEL A: YCS specification				
% participating in full-time education	76.6*	74.1*	74.2**	71.1**
A1. Young person characteristics variables				
% female	47.9	50.1	51.1	49.2
% respondent from ethnic minority	2.1**	15.7**	1.7**	15.7**
% whether reached Level 2 in Year 11	60.9**	53.5**	66.5**	56.3**
% 5+ A*-C GCSEs including Maths and English	52.0**	43.2**	55.1**	44.3**
% has health problem or disability	4.9**	3.7**	3.3	3.3
% played truant in Year 11	29.1**	33.2**	21.4**	25.9**
% expelled or suspended in Years 10 or 11	9.6	10.0	5.7**	7.5**
A2. School variables				
% at grammar/independent school in Year 11	14.6**	9.7**	13.3**	10.6**
A3. Family background variables				
% live in owner-occupied house	85.7**	79.1**	80.6**	70.8**
% lives in a council house	7.8**	13.6**	12.7**	23.9**
% live with father only	5.0	5.5	2.6	2.7
% live with mother only	13.3**	16.6**	16.8**	24.7*
% live with neither parent	3.1*	3.9*	0.1**	1.3**
% only mother employed	10.9	11.9	16.5**	19.2**
% only father employed	16.3	15.6	14.7	14.8
% both parents employed	65.1**	58.2**	61.5**	49.2**
% father's occupation in SOC 1-3	40.6**	33.3**	40.3**	29.8**
% mother's occupation in SOC 1-3	31.9**	25.6**	33.5**	27.1**
% at least one parent with a degree[2]	32.5**	24.9**	21.8**	17.1**
% at least one parent with A-levels[2]	21.8	20.3	28.7**	23.5**
A4. Regional variables				
distance from academic institution in km	5.2**	1.5**	5.31**	1.5**
distance from vocational institution in km	11.6**	4.5**	11.00**	4.3**
regional index of local deprivation	0.131**	0.225**	0.127**	0.246**
PANEL B. Additional variables available in LSYPE				
B1. Young person characteristics variables				
key stage 3 average points score[3]			35.7**	33.5**
% parent mentions child has special needs			8.8	10.2
% child currently has statement of special needs			2.8**	4.4**
% had work experience place while at school			33.2**	38.0**
% whether has a job during term time			45.5**	26.5**
% whether has caring responsibilities			4.9*	6.1*
attitude to school[4]			33.0**	32.3**
B2. School variables				
% independent school			7.6	7.3
% foundation school			19.1**	14.7**
% voluntary aided/controlled school			10.6**	14.2**

	YCS		LSYPE	
	Rural	Urban	Rural	Urban
% of 15 year olds at school with SEN			12.1**	14.7**
% reaching Level 2 at school			63.0**	53.4**
KS3-GCSE value added at school[5]			994.6**	990.6**
% unauthorised absence rate at school			0.7**	1.3**
% eligible for free meals at school			6.4**	15.9**
% at school first language not English			1.4**	10.4**
B3. Family background variables				
% father not present			17.4**	26.0**
% mother not present			3.2	4.0
% father in professional/managerial job			40.3**	29.8**
% father in intermediate job			24.5**	19.3**
% father in routine job			11.8**	15.3**
% mother in professional/managerial job			33.5**	27.1**
% mother in intermediate job			23.2**	19.5**
% mother in routine job			21.3	21.6
% mother has a degree			12.9**	9.9**
% mother has other HE			14.3**	10.9**
% mother has A-levels			15.1**	11.9**
% father has a degree			12.9**	9.7**
% father has other HE			10.3**	6.9**
% father has A-levels			14.3**	11.8**
% high family income[6]			64.1**	52.1**
% lives with both natural parents			67.4**	61.0**
% lives in a single parent family			17.5**	25.8**
number of siblings living in household			1.3**	1.4**
number of risk factors faced[7]			1.5	1.5
B4. Parental attitude variables				
% whether attend parents' evenings			85.2**	82.9**
% arranged special meetings with teachers			22.1**	28.1**
% parent very / fairly involved in child's school life			73.4*	76.1*
% whether paid for private tuition in last year			13.4	14.4
% parent wants child to continue in pc education			80.3	79.9
% parents will pay expenses of pc education			89.3**	86.3**

Source: YCS12 and LSYPE, authors' calculations.

Notes:

1. Asterisks denote the results of a test of the difference in means between rural and urban areas for each variable. * denotes difference is statistically significant at 5%; ** denotes difference is statistically significant at 1%.
2. The qualifications held by parents are measured as their highest qualification. The reason for the higher level of parental qualification in the YCS is not clear. The YCS question does include step-parents, whilst the LSYPE question refers to just parents, potentially giving more people amongst whom to find a higher qualification in the YCS.
3. The average Key Stage 3 score variable has a minimum value of 15 and maximum value of 53 in the data, with a standard deviation of 6.8.
4. The 'attitude to school' variable is derived from the responses to a series of statements in LSYPE, with which respondents are asked whether they agree or disagree. Examples of the statements include 'I am happy when I am at school', 'school is a waste of time for me', 'school work is worth

doing', 'I work as hard as I can at school' etc. The variable has a minimum value of 0 and a maximum value of 48 in the data, with a standard deviation of 8.4.
5. The value added 'score' for each pupil is the difference between their GCSE/GNVQ total point score and the median GCSE/GNVQ point score for all pupils with a similar average Key Stage 3 score. These are then aggregated to give a 'score' for the school. This indicates the value the school has added, on average, for those pupils between Key Stage 3 and GCSE/GNVQ. In the data the variable has a minimum value of 930.4 and a maximum value of 1063.9, with standard deviation of 18.0
6. 'High family income' is defined as above the median wage. However, because the family income data is grouped into 13 bands, and the median falls part way into one band, then some individuals in this particular band are classed as 'high income' when in fact their unknown precise family income will lie just below the median. This gives an overall proportion, across all urban and rural respondents, of 54% reporting above median family income, rather than the anticipated 50%.
7. The 'number of 'risk factors' variable counts the number of risk factors the young person has experienced from a range of variables relating to cigarette, alcohol, cannabis usage and experience of graffiti, vandalism, shoplifting or fighting. It has a minimum value of 0 and a maximum value of 8 in the data, with a standard deviation of 1.5.

Starting in Panel A with the variables found in both data sets, the descriptive statistics show that, in terms of most characteristics, rural respondents are more 'advantaged' than their urban counterparts. The pattern of results is very similar in the two data sets. The literature review in Section 2 above revealed that prior attainment has been shown to be one of the main determinants of post-compulsory participation. As can be seen in Panel A1, young people in rural areas have a clear advantage here. For example, in the YCS, 61% of rural respondents had reached Level 2 (5 or more GCSEs at grade C or above, or their equivalent) by the end of compulsory schooling, compared to just 54% in urban areas. There is increasing emphasis on young people acquiring Maths and English amongst their GCSE successes. When the GCSE 'success rate' is measured as those individuals with 5 or more GCSEs at grade C or above including both Maths and English, 52% of rural respondents are successful according to this definition, compared to 43% of urban respondents. Similar advantages, with all levels of attainment slightly higher, are observed in LSYPE[15]. Other individual characteristics show that rural respondents are significantly less likely to have ever played truant or been excluded from their school. There are substantially fewer young people from ethnic minorities in rural areas (2%) than in urban areas (16%). The proportion reporting health problems or disabilities is small in both data sets, though the difference between rural and urban areas is statistically significant in the YCS.

[15] One reason for the higher attainment in LSYPE is that this cohort sat their GCSEs in 2006, three years after the YCS cohort, and so their achievement figures will reflect the overall increase in attainment rates over time.

The only characteristic of the school attended in Year 11 (the final year of compulsory schooling) recorded in the YCS is whether it is a grammar or independent school. As shown in Panel A2, rural respondents are significantly more likely to have attended such a school than urban respondents.

More information is available on family background in both data sets as can be seen in Panel A3, and here a rural 'advantage' is again observed on all measures of family background. Thus, relative to urban respondents, rural respondents are significantly more likely to live in a owner-occupied home and less likely to live in a council home, more likely to be living with both their parents, more likely to have both parents in employment, more likely to have parents in more senior occupations (SOC 1-3 which include professional, manager and associate professional occupations), and more likely to have parents with degrees and A-levels.

The final variables which are available in both data sets are presented in Panel A4 and are 'regional' variables. These include a measure of the distance from each survey respondent's home address to the nearest educational institution offering academic or vocational education. The distance from respondents' home addresses to their nearest educational establishment offering post-compulsory education is about 3-4 times greater for rural respondents than for urban respondents. Given that we hypothesise that distance might dissuade some individuals from participating in post-compulsory education, this therefore represents a negative influence on participation in rural areas. Note that, because far fewer institutions offer vocational qualifications than academic qualifications, individuals on average live further away from a vocational institution.[16] The other 'regional' variable is the index of multiple deprivation (IMD) for each respondent's local area[17]. Given the summary of characteristics in Panel A3, it is perhaps not surprising that the IMD is significantly higher for urban than for rural respondents.

[16] In classifying education institutions, it was assumed that only Further Education Colleges and Sixth Form Centres offer vocational education.

[17] The IMD is a weighted average of 7 'domains' or separate indicators of deprivation (income deprivation; employment deprivation; health deprivation and disability; education, skills and training deprivation; barriers to housing and services; living environment deprivation; and crime). See http://www.communities.gov.uk/communities/neighbourhoodrenewal/deprivation/deprivation07/ for further details.

Turning to Panel B, there are many more additional variables available in the more comprehensive LSYPE. These confirm the general pattern of rural advantage shown in Panel A. Panel B1 presents the information matched in from the National Pupil Database (NPD) which contains SAT scores for every pupil and shows that rural respondents were also outperforming their urban counterparts at Key Stage 3 (age 14)[18]. When compared to urban respondents, rural respondents are less likely to have taken part in a formal work experience programme whilst at school, though much more likely to have had a term-time job. They are also less likely to have caring responsibilities. There is much more information about the characteristics of the school attended at the end of compulsory education in LSYPE than in the YCS. Panel B2 of Table 4 shows that the schools attended by rural respondents have much lower proportions of pupils: with special educational needs (SEN); eligible for free school meals; not having English as their first language; unauthorised absence rates. In contrast, they have a higher value added score between Key Stage 3 and GCSE, and much higher proportions of pupils reaching Level 2 by the end of compulsory schooling.

The extended range of family background variables reported in Panel B3 continue the pattern, with rural respondents more likely to live with both parents, have parents in higher level occupations, have parents with higher qualifications, and to live in a high (above median) income family.

Finally, in terms of reported parental attitudes to education in Panel B4, here the differences between rural and urban respondents are less clear cut, and on three of the six measures (two of the differences being statistically significant), urban parents report a more positive attitude to education than rural parents, for example by getting more involved in their children's education.

[18] We also had information on Key Stage 2 (age 11) test scores. However, after controlling for Key Stage 3 and GCSE scores, Key Stage 2 scores were found to have a zero effect on the post-compulsory participation decision. In addition, this variable had significantly more missing values than the other measures of prior attainment, and so it was not used in any of the analyses reported here.

Summary

- The post-compulsory education participation rate is higher in rural than in urban areas, though the difference is small.
- Disaggregating areas further, the highest participation rates are observed in dispersed rural areas, and rural villages.
- The raw data reveal a number of differences between rural and urban areas, and between the individuals who live there.
- The characteristics of young people and their families living in rural areas suggest numerous advantages which could potentially explain the higher participation in rural areas. Young people living in rural areas are more likely to have highly educated, professionally-employed parents, more likely to live with both of their parents, and more likely to have higher family income. They are less likely to live in deprived areas, and more likely to attend higher achieving schools. As a result of such factors, they have higher levels of attainment, on average, during compulsory education.
- Young people in rural areas do, however, live further away, on average, from post-compulsory education institutions, which may negatively influence their post-compulsory education participation decision.
- The net result of these different positive and negative influences on post-compulsory participation is investigated in the multivariate analysis in the following sections.

4.2 Determinants of Post-Compulsory Participation at Age 16/17

Table 5 reports the first set of multivariate results, examining the factors associated with the likelihood of participating in full-time education in the first year after the completion of compulsory education. The estimating equation uses the variables in Panel A of Table 4, so that a common specification for the two data sets can be employed. The equations were estimated by probit analysis, to take account of the fact that the dependent variable is a dichotomous variable taking only the values 0 or 1. The numbers reported in Table 5 are the marginal effects on the probability of participating in post-compulsory education, rather than the actual probit coefficients themselves. This means the interpretation of the reported numbers is the change in

the probability of participating in post-compulsory education (measured in percentage points) if the situation denoted by the relevant explanatory variable holds relative to the omitted category for that variable (for discrete yes/no variables) or changes by one unit (for continuous variables). Thus, the first panel of results shows the marginal effects for the rural area indicators, where the full set of six rural area identifiers has been used. The omitted category in this case is all urban respondents[19] so that, for example, the reported marginal effect for 'rural dispersed – less sparse' respondents in the YCS 'without controls' specification shows that respondents in such areas are (0.080=) 8 percentage points more likely to participate in post-compulsory education than urban respondents. This effect is statistically significant at the 1% significance level as indicated by the **. The fact that there are no control variables included means, however, that none of the differences between urban and rural respondents as discussed above in Section 4.1, are taken account of in this comparison. It is therefore a 'raw' difference in participation rates. The only other statistically significant effect in the first column of results shows that respondents living in less sparse rural villages are 4.7 percentage points more likely to participate in post-compulsory education than urban respondents, again not controlling for any differences in characteristics between such individuals. It should be noted, however, that these results are not robust across data sets, and the third column of Table 5 shows that in LSYPE there are no statistically significant differences in post-compulsory participation rates across any of the categories of the rural-urban definition.

[19] We have amalgamated all urban respondents into the base category since the number of urban respondents in the less sparse areas is very few as shown in Table 2 above.

Table 5: In full time education at age 16/17: Full rural definition specification

	YCS		LSYPE	
	Without controls	With controls	Without controls	With controls
Rural definition variables				
Rural town – less sparse	0.000 (0.018)	0.010 (0.011)	-0.027 (0.016)	-0.006 (0.016)
Rural town – sparse	-0.089 (0.054)	-0.040 (0.048)	-0.043 (0.058)	0.053 (0.042)
Rural village – less sparse	0.047 (0.012)**	0.024 (0.013)	0.021 (0.017)	0.006 (0.019)
Rural village – sparse	0.033 (0.047)	0.069 (0.026)	-0.077 (0.077)	-0.064 (0.077)
Rural dispersed – less sparse	0.080 (0.016)**	0.037 (0.016)*	-0.007 (0.026)	-0.018 (0.027)
Rural dispersed – sparse	-0.070 (0.061)	0.026 (0.040)	0.092 (0.052)	0.067 (0.050)
Young person characteristics variables				
female		0.040 (0.006)**		0.055 (0.008)**
respondent from ethnic minority		0.100 (0.006)**		0.160 (0.008)**
whether reached Level 2 in Year 11		0.128 (0.012)**		0.111 (0.012)**
5+ A*-C GCSEs incl. Maths and English		0.112 (0.011)**		0.155 (0.011)**
has health problem or disability		0.013 (0.015)		-0.061 (0.024)**
played truant in Year 11		-0.074 (0.008)**		-0.094 (0.010)**
expelled or suspended in Years 10 or 11		-0.108 (0.015)**		-0.121 (0.019)**
School variables				
at grammar/indep. school in Year 11		0.078 (0.009)**		0.111 (0.012)**
Family background variables				
live in owner-occupied house		0.016 (0.013)		0.007 (0.019)
lives in a council house		0.003 (0.014)		-0.019 (0.018)
live with father only		-0.004 (0.014)		-0.076 (0.033)*
live with mother only		0.011 (0.010)		0.000 (0.014)
live with neither parent		-0.068 (0.021)**		-0.081 (0.044)*
only mother employed		-0.038 (0.015)**		-0.016 (0.015)
only father employed		-0.014 (0.013)		-0.002 (0.016)
both parents employed		-0.030 (0.011)**		-0.050 (0.016)**
father's occupation in SOC 1-3		0.014 (0.007)		0.043 (0.010)**
mother's occupation in SOC 1-3		0.028 (0.007)**		0.012 (0.010)
at least one parent with a degree		0.049 (0.008)**		0.089 (0.010)**
at least one parent with A-levels		0.029 (0.007)**		0.013 (0.009)
Regional variables				
distance from academic institution (km)		-0.003 (0.002)		0.000 (0.002)
distance from vocational institution (km)		-0.001 (0.001)		-0.002 (0.001)*
regional index of local deprivation		-0.098 (0.023)**		-0.099 (0.026)**
Number of observations	12,872	12,139	11,226	11,226

Notes:
1. Estimation by probit. Table reports marginal effects.
2. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.

Column 2 adds the range of available control variables in the YCS to the equation. The results for the area indicators in the top panel show that the higher rates of participation in rural villages – less sparse, and rural dispersed – less sparse areas are both reduced by around half once these other factors determining post-compulsory participation are taken into account, and only the latter remains statistically significant (although only at the 5% level). This suggests that when holding constant all the individual, school, family and regional characteristics listed in the table and so comparing like-with-like, individuals living in rural dispersed – less sparse areas are 3.7 percentage points more likely to participate in post-compulsory education than individuals living in urban areas. Comparing the results in the first and second columns, however, does show that around half of the higher participation rate observed in column 1 for young people in rural villages – less sparse and rural dispersed – less sparse areas is due to their favourable characteristics, rather than specifically due to the area where they live.[20] In LSYPE, there are no statistically significant differences in participation rates across any of the six rural categories, with or without control variables.

Briefly considering the control variables, their effects on participation are mainly consistent across the two data sets, and also mostly agree with the existing literature reviewed in Section 2 above. Prior attainment is a key determinant, with the acquisition of 5 or more good GCSEs (Level 2) being associated with an 11-13 percentage point higher post-compulsory participation rate. There is a further 11-16 percentage point increase in participation if these GCSEs include Maths and English. Young people who ever played truant during compulsory schooling or were excluded at some point are, as expected, less likely to participate in post-compulsory schooling. This effect is not necessarily causal, however, and is more likely to reflect unobserved characteristics that influence both inappropriate behaviour and dropping-out of education at age 16. Amongst the remaining individual characteristics, females are more likely to participate in post-compulsory education than males, by 4-6 percentage points, whilst ethnic minority young people are much more likely to participate, by 10-16 percentage points, after controlling for other characteristics.

[20] The extent to which the rural-urban difference in participation is 'explained' by the differences in the characteristics of the individuals living in those areas, and which characteristics in particular account for the difference is discussed in the decomposition analysis presented in the next sub-section.

Young people who have had health problems are significantly less likely to participate in the LSYPE sample, though not in the YCS[21].

Attending a grammar or independent school is associated with a higher post-compulsory participation rate by 8-11 percentage points. Parental background is also important in general, with those young people whose parents have a higher level of qualifications, or who work in more senior occupations, being more likely to stay-on in full-time education. After controlling for all other factors, however, there is no additional gain from having both parents in employment, with the coefficients on such variables actually negative and statistically significant in both data sets.

Finally, with respect to the regional variables, even after controlling for their own family situation and so holding that constant, young people who live in areas with a higher level of local deprivation are significantly less likely to participate in post-compulsory education. The variables measuring distance to academic and vocational educational institutions have negative coefficients, as hypothesised, so that those who live further from an education institution supplying post-compulsory education are less likely to remain in full-time education, though only the distance to a vocational institution in the LSYPE equation attracts a statistically significant coefficient. Given the potential policy relevance of this influence on participation[22], it was decided to investigate it further. In particular, it is assumed in the specification in Table 5 that the marginal effect of distance on the probability of participation is constant, such that each additional kilometre distance from an education institution has the same impact on the probability of post-compulsory participation. However, the effect of distance may be non-linear, such that the negative influence of distance may have an increasingly larger effect the greater the distances involved, or may only have an impact for large distances. Thus, an extra kilometre may not influence

[21] The questions used to define this characteristic are actually quite different in the two surveys and this may explain this difference in the findings between the data sets. The question in the YCS asks a yes/no question whether respondents have an ongoing health problem or disability, whereas the LSYPE asks individuals how their health has been in the previous year (at wave 3), with those responding that their health has not been very good or not good at all classified here as having a health problem. In particular, it should be noted that the YCS measure includes people with a disability, which is not the same as a specific health problem.

[22] If distance does influence the participation decision, then potential policies include subsidised travel for those living in more remote areas to facilitate access. This effect is probably easier to influence by policy – at least in the short term – than, for example, the effects on participation due to prior attainment or family background.

the participation decision much when short distances are involved (i.e. whether an individual lives 1 or 2 kilometres from an education institution does not greatly affect their decision to remain in post-compulsory education), but at further distances, an extra kilometre may just tip the balance in the decision not to participate (i.e. living 7 kilometres away rather than 6 does have more of an effect on the staying-on decision). To investigate whether there are such non-linear effects, we replaced the continuous distance measures as used in Table 5 with dummy variables indicating 2-8 km from the nearest education institution and more than 8 km from the nearest education institution (with the base category being individuals who live less than 2 km away), separately for institutions providing academic and vocational qualifications.[23] The marginal effects on these variables will measure the difference in the participation rate of young people living at such distances, relative to young people in the comparison group who live less than 2 kilometres from their nearest institution, controlling for the other characteristics in the estimated equation.

The marginal effects for these new distance variables are reported in Table 6 below. All other control variables used in Table 5 were again included in the estimated equations, but their results are not reported as they are virtually identical to those reported previously.

Table 6: Marginal effects of distance category variables on full time education participation at age 16/17

	YCS	LSYPE
Distance measures		
2-8 kilometres from nearest academic institution	-0.008 (0.008)	0.009 (0.010)
8+ kilometres from nearest academic institution	-0.016 (0.019)	0.010 (0.023)
2-8 kilometres from nearest vocational institution	-0.002 (0.008)	-0.017 (0.010)
8+ kilometres from nearest vocational institution	-0.017 (0.011)	-0.028 (0.014)*

Notes:
1. Estimation by probit. Table reports marginal effects.
2. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.
3. Control variables were included as listed in the 'with controls' specifications in Table 5.

[23] The choice of cutoffs at 2km and 8km is essentially arbitrary, although 2km is the mean distance across all individuals in the two data sets, and less than 2km seemed to represent a 'walking distance' category. 8km was chosen as it represents an 'extreme' distance that is relevant to just under 5% of all respondents. A higher cutoff point would therefore have been relevant to too few individuals. However, we did experiment with changing these cutoff points, and the results were not qualitatively affected.

While the results in Table 6 do suggest that greater distances have a larger impact on lowering participation, only one of the estimated coefficients is statistically significant.

Tables 7 and 8 present the results for alternative classifications of the rural-urban distinction. Some of the categories in the 8-point classification have few observations, particularly in those areas categorised as 'sparse'. These small sample sizes will result in high standard errors, and thus tend to give rise to estimates which are not statistically significant or robust. Table 7 therefore amalgamates all of the rural less sparse areas, and all of the rural sparse areas together into two categories, whilst Table 8 combines the various types of rural area (town, village and dispersed) together into three separate categories (i.e. merging sparse and less sparse areas within a particular area type together). The base category for comparison in both cases is all urban respondents as before. The 'with controls' specifications reported in Tables 7 and 8 also include all of the control variables used previously in Table 5. Since their coefficients are essentially identical to those reported in Table 5, they are not reported in Tables 7 and 8.

Table 7: In full time education age 16/17: Sparse and less sparse rural definition specification

	YCS		LSYPE	
	Without controls	**With controls**	**Without controls**	**With controls**
Rural definition variables				
Rural – less sparse	0.030 (0.008)**	0.017 (0.009)	-0.006 (0.011)	-0.004 (0.013)
Rural – sparse	-0.043 (0.032)	0.016 (0.026)	-0.009 (0.036)	0.027 (0.035)

Notes:
1. Estimation by probit. Table reports marginal effects.
2. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.
3. Control variables were included as listed in the 'with controls' specifications in Table 5.

The results in Table 7 show that, in the raw YCS data, respondents in less sparse rural areas are 3 percentage points more likely to participate in post-compulsory education than those in urban areas. There is no difference in participation rates for sparse rural areas, nor for either less sparse or sparse rural areas in the LSYPE data. Much of the positive differential for rural less sparse areas in the YCS data is

due to the favourable characteristics of the respondents in such areas as shown above since once these characteristics are controlled for, the observed effect of living in rural less sparse areas is much reduced and is not significantly different from zero.

Table 8: In full time education age 16/17: Type of rural area specification

	YCS		LSYPE	
	Without controls	With controls	Without controls	With controls
Rural definition variables				
Rural – town	-0.005 (0.012)	0.007 (0.011)	-0.028 (0.015)	-0.004 (0.016)
Rural – village	0.046 (0.012)**	0.027 (0.013)*	0.015 (0.016)	-0.002 (0.019)
Rural – dispersed	0.063 (0.016)**	0.035 (0.016)*	0.007 (0.024)	-0.011 (0.025)

Notes:
1. Estimation by probit. Table reports marginal effects.
2. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.
3. Control variables were included as listed in the 'with controls' specifications in Table 5.

When looking at type of rural area, as in Table 8, it is clear that where rural participation is higher than in urban areas in the YCS sample, it is in village and dispersed areas, with rural towns being very similar to urban areas in terms of their post-compulsory participation rate. Once again, these statistically significant raw differences are only observed in the YCS data set. When the characteristics of individuals are taken into account (i.e. as in the 'with controls' column), the magnitudes of these participation differentials are almost halved, although positive differences of 2.7 and 3.5 percentage points do remain between participation rates in rural village and rural dispersed areas, respectively, relative to urban areas, and these differences are statistically significant at the 5% significance level. Thus, even after controlling for the advantages enjoyed by individuals in these two types of rural areas, it would appear that their post-compulsory participation rates are slightly higher than those found in urban areas in the YCS data set. However, in contrast, in the LSYPE data there are no differences in participation between types of rural area and urban areas.

Table 9 decomposes the simple raw rural-urban difference in participation rates into that part due to differences in the characteristics of the individuals in the two areas (the 'explained' part) and that due to differences in the coefficients (the 'unexplained' part, or that caused by differences in the effects associated with a given set of

characteristics across the different types of area, for unknown reasons). Formally, this decomposition can be written:

$$D = \bar{X}_R \beta_R - \bar{X}_U \beta_U = (\bar{X}_R - \bar{X}_U)\beta_U + \bar{X}_R(\beta_R - \beta_U),$$

where *D* is the participation differential between rural and urban areas, $\bar{X}_R$ and $\bar{X}_U$ represents the mean values of the characteristics of the individuals in rural (*R*) and urban (*U*) areas respectively, while β_R and β_U represent the coefficients on these characteristics. The first term on the right hand side therefore represent the 'explained' part of the decomposition, that is the part of the overall differential, *D*, that is due to differences in the characteristics of individuals living in rural and urban areas. The second term on the right hand side is therefore the 'unexplained' component, i.e. that part of the overall differential that is due to how individuals with identical characteristics fare in the two types of regions (and therefore is due to differences in the coefficients). This latter part could be due to the inherent effect of something intangible such as 'ruralness', or it could be due to unobserved differences in the characteristics of individuals in the two areas that have not been controlled for. Since the unexplained gap is positive in favour of rural areas (see results in Table 9 below), one way to think about the unexplained portion is that it shows the predicted fall in the likely participation rate of a person with typical rural characteristics, if they were to be moved to an urban area rather than continue living in a rural area.[24]

The results in Table 9 show that the raw difference in participation rates in the YCS between rural and urban areas is 2.6 percentage points[25] and is statistically significant at the 5% level. However all of this difference is left 'unexplained' by the decomposition. The detailed breakdown of the explained part of the differential shows that rural young people have some positive characteristics that serve to increase their post-compulsory participation (principally prior attainment, but also school and family background) but also negative characteristics that act to reduce

[24] Note that it is possible to write the decomposition with the difference in characteristics evaluated in terms of the rural coefficients and the difference in coefficients to be evaluated in terms of the mean urban characteristics. This would give a different, and equally valid, result for the decomposition. It was decided to use the equation as set out in the text, as the principal interest here is in rural affairs, and it was felt that the unexplained portion being in terms of individuals with rural characteristics and how they would fare differently if they were in an urban area, was the most appropriate.

[25] This 2.6 percentage point difference between rural and urban participation rates was first seen in Tables 3 and 4 (subject to rounding).

their post-compulsory participation (namely individual characteristics[26] and regional characteristics). The YCS results in Table 9 show that the net impact of these positive and negative influences on participation almost exactly balance out, meaning that on the basis of the observed characteristics alone, young people in urban and rural areas should have virtually the same post-compulsory participation rates. The fact that they do not, and the participation rate is significantly higher in rural areas is therefore left unexplained, and attributed to some unobserved effect of 'ruralness'.

Table 9: Decomposition of rural-urban difference in full time age 16/17 participation rate

	YCS	LSYPE
Raw rural-urban difference	0.026 (0.011)*	0.032 (0.012)**
Due to characteristics:	-0.002 (0.015)	0.018 (0.016)
Prior attainment	*0.023 (0.003)***	*0.026 (0.003)***
Individual characteristics	*-0.018 (0.003)***	*-0.016 (0.003)***
School characteristics	*0.002 (0.001)***	*0.004 (0.001)***
Family characteristics	*0.005 (0.002)***	*0.008 (0.003)***
Regional characteristics	*-0.013 (0.014)*	*-0.005 (0.016)*
Due to coefficients:	0.028 (0.017)	0.014 (0.018)

Notes:
1. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.
2. Control variables were included as listed in the 'with controls' specifications in Table 5.

Table 9 also presents an identical decomposition analysis for the LSYPE data. As already seen earlier in Tables 3 and 4, the raw rural-urban participation gap in LSYPE is 3.2 percentage points. In this case, the decomposition analysis shows that differences in characteristics can 'explain' around half of this difference. The disaggregation of the explained part again shows that the main advantage of rural young people is their significantly higher prior attainment, offset by negative individual characteristics (again, a lower proportion of young people from ethnic minorities who are more likely to participate). The smaller offsetting negative effect of regional characteristics (essentially, the greater distances to education institutions) observed in LSYPE compared to YCS is the main reason for the overall positive

[26] Further inspection of the data, not shown in the table, reveals that the key individual characteristic of the rural respondents reducing their participation rate is ethnicity. Table 5 showed that young people from ethnic minorities are much more likely to participate in post-compulsory schooling than those from white ethnic groups, and Table 4 showed that only a very small proportion of rural young people belong to an ethnic minority compared to a significantly higher proportion in urban areas.

explanatory power of the characteristics in the LYSPE decomposition. Overall, the LSYPE results suggest that rural young people are mostly participating as their characteristics would predict, with a smaller unexplained differential than in the YCS.

Summary

- Section 4.1 showed that, in the raw data, post-compulsory participation rates are slightly higher in rural areas than in urban areas.
- The results from the analysis of the YCS data show that, disaggregating rural areas by type, the higher rural participation rates are observed in rural villages and rural dispersed areas, but not in rural towns which have similar participation rates as urban areas.
- When we control for the characteristics of individuals, their families and their local areas, these rural-urban differentials are reduced, but remain positive and statistically significant. The YCS results therefore suggest that the higher participation rates in rural village and rural dispersed areas are not solely due to the observed advantages of the young people who live there (in terms of their higher prior attainment or more advantaged family background). Thus, even when we compare like-with-like across areas, the post-compulsory participation rates in such rural areas are still around 3 percentage points higher than in urban areas.
- This result is confirmed by the decomposition analysis of the rural-urban differential in participation rates using YCS data. The observed characteristics cannot explain the observed differential, suggesting some small unexplained positive influence of living in rural areas for participation.
- These findings are not robust across data sets. Using the LSYPE data, there were no statistically significant differences in post-compulsory participation between urban areas and any of the rural areas when controlling for observed characteristics.
- One factor that may work against young people participating in post-compulsory education in rural areas is the distance to their nearest education institution. While there is some suggestion that distance can reduce participation, in both linear and non-linear specifications, the estimated effect of this factor on participation is small and is not statistically significant in either data set, with the exception of distance to vocational institutions in LSYPE.

4.3 Determinants of Post-Compulsory Participation at Age 16/17 by Academic/Vocational Status

So far, the analysis has only considered whether individuals participate in post-compulsory education, without taking any account of what they might be doing. It could be that the overall narrow gap in participation rates between rural and urban areas is masking larger differences in participation rates by type of education undertaken, for example if rural individuals were more likely to participate in 'academic' education (i.e. studying for academic qualifications) whilst those in urban areas were more likely to participate in 'vocational' education (i.e. studying for vocational qualifications). This sub-section therefore incorporates this distinction into the analysis. Participation by type of education is of interest, given that the existing evidence on returns to education suggests that the wage returns (and hence presumably the productivity returns) to academic qualifications are higher than to vocational qualifications.[27]

One way to investigate this issue would have been to estimate two separate probit equations similar to those presented in Table 5 above, with academic study and vocational study as the two dependent variables. However, it could reasonably be argued that these decisions are not independent but that, in fact, individuals make a simultaneous choice between three alternatives: academic study; vocational study; and non-participation. The outcomes of their decision-making should therefore be estimated jointly. Thus a multinomial logit equation was estimated, with the dependent variable taking one of three values to reflect the three possible outcomes of the decision-making process. This procedure estimates the impact of the explanatory variables on two of the outcomes relative to the third, omitted outcome, taking into account their inter-relationship. In the results reported in Table 10, the omitted outcome is not participating in post-compulsory education, so that effects are reported for participating in academic education and vocational education, all relative to this omitted outcome. As in the previous section with the probit equations, the table reports the marginal effects on the probability of participation rather than the actual estimated coefficients. The interpretation of the effects reported in Table 10 is

[27] For a recent and comprehensive analysis of wage returns to different qualifications in Britain, see Jenkins *et al.* (2007).

therefore as before, so that, for example, from the first column, young people living in rural town – less sparse areas are 4.7 percentage points more likely to engage in academic study than not participate at all, relative to young people in urban areas, having taken into account all of the other factors that are likely to influence their propensity to participate in post-compulsory education. Moreover, this difference is statistically significant at the 5% level.

The positive marginal effects on the rural identifier variables in the first column of results in Table 10 indicate that, in the YCS data, even when holding constant all of the other characteristics included in the table, young people living in all types of rural areas are more likely to engage in academic study than not participate at all, relative to young people in urban areas. The differentials are statistically significant for rural towns – less sparse and rural dispersed – less sparse areas (while rural village – less sparse only just fails to reach statistical significance at the 5% level). These results suggest that individuals in the various rural areas are 5-7 percentage points more likely to engage in academic post-compulsory study than not participate, relative to individuals in urban areas.[28]

Similarly, the second column of results suggest that rural young people are less likely to engage in vocational post-compulsory education than to not participate as compared to their counterparts in urban areas with the same observed characteristics, although the effects are only statistically significant for rural towns (both less sparse and sparse areas). However, in general, these results for the YCS data are not replicated in the LSYPE data set. Only rural town – sparse areas have significantly higher post-compulsory academic education participation, relative to urban areas. However, very few individuals reside in this area type.

[28] The 'sparse' rural areas have even larger marginal effects for academic study, but are not statistically significant as a consequence of their higher standard errors, in turn caused by lower sample sizes since relatively few people live in such areas.

Table 10: In full time education age 16/17 by academic/vocational status: Full rural definition specification

	YCS		LSYPE	
	Academic education	Vocational education	Academic education	Vocational education
Rural definition variables				
Rural town – less sparse	0.047 (0.019)*	-0.030 (0.013)*	-0.005 (0.026)	-0.001 (0.020)
Rural town – sparse	0.123 (0.078)	-0.135 (0.063)*	0.213 (0.093)*	-0.132 (0.073)
Rural village – less sparse	0.045 (0.023)	-0.017 (0.016)	0.003 (0.030)	-0.000 (0.023)
Rural village – sparse	0.129 (0.077)	-0.027 (0.048)	-0.107 (0.103)	0.038 (0.082)
Rural dispersed – less sparse	0.066 (0.033)*	-0.011 (0.021)	-0.009 (0.041)	-0.006 (0.032)
Rural dispersed – sparse	0.079 (0.082)	-0.044 (0.054)	0.006 (0.097)	0.067 (0.071)
Young person characteristics variables				
female	0.037 (0.010)**	0.005 (0.007)	0.068 (0.012)**	-0.006 (0.009)
respondent from ethnic minority	0.171 (0.016)**	-0.020 (0.010)*	0.266 (0.016)**	-0.044 (0.011)**
whether reached Level 2 in Year 11	0.288 (0.015)**	-0.134 (0.010)**	0.317 (0.017)**	-0.155 (0.012)**
5+ A*-C GCSEs incl. Maths and English	0.256 (0.013)**	-0.144 (0.010)**	0.353 (0.016)**	-0.192 (0.013)**
has health problem or disability	0.021 (0.026)	-0.004 (0.016)	-0.104 (0.034)**	0.034 (0.024)
played truant in Year 11	-0.080 (0.011)**	0.009 (0.007)	-0.105 (0.015)**	0.008 (0.011)
expelled or suspended in Years 10 or 11	-0.104 (0.020)**	0.012 (0.013)	-0.199 (0.031)**	0.061 (0.020)**
School variables				
at grammar/indep. school in Year 11	0.255 (0.023)**	-0.147 (0.019)**	0.252 (0.032)**	-0.098 (0.027)**
Family background variables				
live in owner-occupied house	0.022 (0.020)	-0.006 (0.013)	-0.007 (0.026)	0.009 (0.019)
lives in a council house	-0.020 (0.024)	0.015 (0.015)	-0.051 (0.028)	0.022 (0.020)
live with father only	0.011 (0.022)	-0.009 (0.014)	-0.173 (0.045)**	0.074 (0.031)*
live with mother only	0.019 (0.016)	-0.006 (0.011)	-0.042 (0.023)	0.033 (0.016)*
live with neither parent	-0.036 (0.028)	-0.018 (0.018)	-0.032 (0.058)	-0.036 (0.042)
only mother employed	-0.025 (0.022)	-0.006 (0.014)	-0.018 (0.023)	0.002 (0.016)
only father employed	0.005 (0.020)	-0.012 (0.013)	-0.011 (0.025)	0.008 (0.017)
both parents employed	-0.023 (0.018)	-0.004 (0.011)	-0.095 (0.025)**	0.033 (0.018)
father's occupation in SOC 1-3	0.051 (0.011)**	-0.030 (0.008)**	0.077 (0.016)**	-0.025 (0.012)*
mother's occupation in SOC 1-3	0.046 (0.012)**	-0.015 (0.008)	0.038 (0.016)*	-0.022 (0.012)
at least one parent with a degree	0.090 (0.013)**	-0.032 (0.009)**	0.161 (0.020)**	-0.048 (0.016)**
at least one parent with A-levels	0.041 (0.012)**	-0.009 (0.008)	0.008 (0.015)	0.007 (0.011)
Regional variables				
distance from academic institution (km)	-0.012 (0.003)**	0.007 (0.002)**	-0.008 (0.003)*	0.007 (0.003)*
distance from vocational institution (km)	-0.001 (0.001)	-0.000 (0.001)	0.002 (0.002)	-0.004 (0.001)**
regional index of local deprivation	-0.080 (0.037)**	-0.014 (0.024)	-0.203 (0.042)**	0.073 (0.030)*
Number of observations	12,139		11,226	

Notes:
1. Estimation by multinomial logit. Table reports marginal effects.
2. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.

The remaining coefficients in Table 10 reveal that individual characteristics again dominate in the determinants of post-compulsory participation when now also explaining the type of post-compulsory education chosen. Thus, individuals who have reached Level 2 in compulsory education are around 30 percentage points more likely to engage in academic post-compulsory study rather than not participate at all, compared to someone who failed to reach that level, with an additional, similar-sized effect, if they also have Maths and English amongst their 5 good GCSEs. Higher level prior attainment is also associated with significantly lower participation rates in vocational education relative to non-participation.

Young people with a history of truancy or exclusion, whom we know from earlier are less likely to participate overall, are less likely to participate in academic post-compulsory education, though they are *not* less likely to engage in vocational post-compulsory education as well. Females and individuals from ethnic minorities are more likely to engage in academic study than males and those from white ethnic groups respectively, with ethnic minorities also being less likely to engage in vocational study. If the respondent has attended a grammar or independent school, then they are significantly more likely to progress into post-compulsory academic study, and significantly less likely to continue into vocational study, than not to participate.

In terms of family background, young people with well-educated parents and/or parents in senior occupations are significantly more likely to engage in academic study, and less likely to engage in vocational study. Finally, the regional variables indicate that respondents from more deprived areas have a lower likelihood of academic study. The distance variables shows that, in both data sets, the further an individual lives from an institution offering academic qualifications, the less likely they are to participate in academic study and the more likely they are to undertake vocational study. The estimated marginal effects suggest that each extra kilometre distance reduces the likelihood of academic participation by around 1 percentage point (1.2 percentage points in YCS and 0.8 percentage points in LSYPE) and increases the likelihood of vocational participation by 0.7 percentage points. Similarly, at least in the LSYPE data, the greater the distance an individual lives from an institution offering vocational post-compulsory learning, the less likely they are to

undertake vocational study, with the magnitude of the effect being about 0.4 percentage points per kilometre. This pattern of findings is consistent with the hypothesis that distance from appropriate provision can dissuade individuals from participation, at least at the margin.

It is of interest to note that a number of the estimated effects – such as the prior attainment and parental education and occupational variables – have different signs for academic and vocational study (i.e. are positive for academic and negative for vocational study, or vice versa). It might have been expected that the effects would be in the same direction in explaining both types of post-compulsory education, with perhaps the academic effects being larger in size. Recall that the effects are measured relative to the omitted category, which in this case is not participating in compulsory education at all. So we might have expected a person with, say, good GCSEs to be much more likely to study for further academic qualifications than not to study at all, but also at least somewhat more likely to study for vocational qualifications than not to study at all. The results show that this is not the case however, with individuals holding good GCSEs being *less* likely to undergo post-compulsory vocational education than to not participate in study at all. It therefore seems that the hierarchy of choice in post-compulsory education for young people with good GCSEs or from families with higher socio-economic status is: first, academic study; then second, no post-compulsory education; and only then, last of all, participating in vocational study.

As in the previous section and reported in Table 6 above, we again experimented with a non-linear specification for the distance variable within the context of the academic/vocational multinomial logit specification, using dummy variables to indicate individuals living 2-8 kilometres and 8+ kilometres from their nearest education institution. The results for these variables are shown in Table 11 below. As before, only the marginal effects for the distance dummy variables are reported, although all of the control variables included in Table 10 above were again included, with their marginal effects unaffected by the change in the definition of the distance variables.

Table 11: Marginal effects of distance category variables on full time education participation at age 16/17 by academic/vocational status

	YCS		LSYPE	
	Academic education	**Vocational education**	**Academic education**	**Vocational education**
Distance measures				
2-8 km from nearest academic institution	-0.039 (0.013)**	0.025 (0.008)**	-0.037 (0.016)*	0.039 (0.012)**
8+ km from nearest academic institution	-0.078 (0.029)**	0.048 (0.019)*	-0.060 (0.039)	0.063 (0.029)*
2-8 km from nearest vocational institution	0.003 (0.013)	-0.004 (0.009)	0.011 (0.015)	-0.028 (0.011)*
8+ km from nearest vocational institution	-0.004 (0.017)	-0.011 (0.011)	0.017 (0.022)	-0.040 (0.016)*

Notes:
1. Estimation by multinomial logit. Table reports marginal effects.
2. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.
3. Control variables were included as in the specifications in Table 10.

The results in Table 11 show that, in both data sets, an individual who lives 2-8 kilometres from their nearest education institution offering academic qualifications is almost 4 percentage points less likely to undertake academic post-compulsory study than an otherwise identical individual living less than 2 kilometres distant. For those who live more than 8 kilometres away from their nearest academic institution, their academic post-compulsory participation is 7.8 percentage points lower than those living less than 2 kilometres away in the YCS data set only. This is a sizeable effect and there is therefore quite strong evidence that individuals may be dissuaded from academic study by large travel distances, at least in one of the two data sets utilised in this report. Similarly, in the LSYPE results at least, greater distance from an institution offering vocational education is associated with a lower probability of participating in post-compulsory vocational education.

Tables 12 and 13 report the results when the rural categories are combined as in Tables 7 and 8 for the probit analysis of the overall participation rate. As before, the results in Tables 12 and 13 are derived from a specification where all control variables were included in the estimated equations, but their coefficients are not reported for reasons of space and because they are unchanged from those reported in Table 10.

Table 12: In full time education age 16/17 by academic/vocational status: Sparse and less sparse rural definition specification

	YCS		LSYPE	
	Academic education	Vocational education	Academic education	Vocational education
Rural definition variables				
Rural – less sparse	0.049 (0.016)**	-0.024 (0.011)*	-0.004 (0.021)	-0.001 (0.016)
Rural – sparse	0.098 (0.048)*	-0.065 (0.033)*	0.049 (0.063)	-0.019 (0.048)

Notes:
1. Estimation by multinomial logit. Table reports marginal effects.
2. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.
3. Control variables were included as in the specifications in Table 10.

Table 13: In full time education age 16/17 by academic/vocational status: Type of rural area specification

	YCS		LSYPE	
	Academic education	Vocational education	Academic education	Vocational education
Rural definition variables				
Rural – town	0.048 (0.019)*	-0.032 (0.013)*	0.004 (0.026)	-0.007 (0.019)
Rural – village	0.049 (0.023)*	-0.015 (0.015)	-0.010 (0.030)	0.004 (0.022)
Rural – dispersed	0.064 (0.032)*	-0.013 (0.021)	-0.016 (0.039)	0.007 (0.030)

Notes:
1. Estimation by multinomial logit. Table reports marginal effects.
2. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.
3. Control variables were included as in the specifications in Table 10.

The aggregated rural categories confirm the results from Table 10, whereby in the YCS, though not in LSYPE, there is evidence that young people in rural areas are significantly more likely to participate in academic post-compulsory education (and by up to 10 percentage points in rural – sparse areas), and less likely to participate in vocational post-compulsory education than young people in urban areas with the same observed characteristics. The greater propensity to participate in academic education applies to all types of rural areas – towns, villages and dispersed areas.

Summary

- When we distinguish between the types of post-compulsory education (academic or vocational), then the YCS evidence consistently shows that young people in all types of rural area are more likely to participate in

academic post-compulsory education, by somewhere between 5 and 10 percentage points depending on the type of rural area.

- There is also some, slightly weaker evidence in the YCS that young people in rural areas are less likely to participate in vocational post-compulsory education than their counterparts in urban areas.
- Again, the LSYPE results differ from the YCS results, with few statistically significant effects of location on type of participation observed in the case of LYSPE.
- When account is taken of the type of post-compulsory education. undertaken, there is some evidence that a greater distance to nearest education institution can particularly reduce participation in academic post-compulsory education.

4.4 Determinants of Post-Compulsory Participation at Age 16/17 – Full LSYPE Specification

The list of variables in the descriptive statistics in Table 4 showed that there are many more variables available in LSYPE than in YCS. So far, most of these have not been used, so that the results above for YCS and LSYPE were based on the same specification of the participation equation. This sub-section therefore adds the additional variables available in LSYPE to the estimated LSYPE participation equation. The results are contained in Table 14. Given that no statistically significant differences in participation rates across types of areas were observed earlier in the LSYPE analysis, it is unlikely that any would emerge now, with even more potential differences in characteristics between the areas now being held constant. Table 14 confirms that this is indeed the case, with none of the estimated rural marginal effects being statistically significant.

Table 14: In full time education age 16/17 – full LSYPE: Full rural definition specification

	LSYPE
Rural definition variables	
Rural town – less sparse	-0.004 (0.016)
Rural town – sparse	0.066 (0.039)
Rural village – less sparse	0.017 (0.018)
Rural village – sparse	-0.071 (0.082)
Rural dispersed – less sparse	-0.004 (0.027)
Rural dispersed – sparse	0.062 (0.050)
Young person characteristics variables	
female	0.028 (0.008)**
respondent from ethnic minority	0.090 (0.010)**
whether reached Level 2 in Year 11	0.037 (0.012)**
5+ A*-C GCSEs including Maths and English	0.084 (0.012)**
key stage 3 average points score	0.007 (0.001)**
parent mentions child has special needs	0.050 (0.013)**
whether child currently has statement of SEN	0.066 (0.018)**
ever suspended or excluded from school	-0.029 (0.017)
played truant in last year	0.015 (0.011)
health not very good or not good at all	-0.043 (0.024)
had work experience place while at school	-0.014 (0.008)
whether has a job during term time	-0.038 (0.009)**
whether has caring responsibilities	0.007 (0.015)
attitude to school	0.006 (0.001)**
School variables	
independent school	0.039 (0.030)
foundation school	0.008 (0.011)
voluntary aided/controlled school	0.002 (1.215)
proportion of 15 year olds at school with SEN	-0.002 (0.000)
proportion reaching Level 2 at school	0.041 (0.038)
KS3-GCSE value added at school	-0.001 (0.000)
unauthorised absence rate at school	-0.178 (0.313)
proportion eligible for free meals at school	-0.006 (0.050)
proportion at school whose first language not English	0.083 (0.031)**
Family background variables	
live in owner-occupied house	-0.005 (0.017)
lives in a council house	-0.017 (0.018)
father not present	0.006 (0.021)
mother not present	-0.031 (0.026)
father in professional/managerial job	0.009 (0.016)
father in intermediate job	-0.027 (0.016)
father in routine job	-0.013 (0.016)
mother in professional/managerial job	-0.024 (0.013)
mother in intermediate job	-0.021 (0.013)
mother in routine job	-0.022 (0.012)
mother has degree	0.048 (0.015)**
mother has other HE	0.033 (0.012)**
mother has A-levels	0.021 (0.012)

	LSYPE
father has degree	0.062 (0.015)**
father has other HE	0.003 (0.016)
father has A-levels	-0.004 (0.013)
high family income	-0.015 (0.010)
lives with both natural parents	0.037 (0.012)**
lives in a single parent family	0.028 (0.018)
number of siblings living in household	0.000 (0.003)
number of risk factors faced	-0.023 (0.003)**
Parental attitude variables	
whether attend parents' evenings	0.044 (0.012)**
parent arranged special meetings with teachers	-0.024 (0.009)**
parent very or fairly involved in child's school life	-0.000 (0.009)
whether paid for private tuition in last year	0.059 (0.010)**
parent wants child to continue in pc education	0.278 (0.015)**
parents will pay expenses of staying in pc education	0.062 (0.013)**
Regional variables	
distance from nearest academic institution (km)	0.001 (0.002)
distance from nearest vocational institution (km)	-0.001 (0.001)
regional index of local deprivation	-0.058 (0.031)
Number of observations	11,226

Notes:
1. Estimation by probit. Table reports marginal effects.
2. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.

With respect to the control variables in this full LSYPE specification, the impact of variables already analysed earlier in the 'YCS-equivalent' specification in Section 4.2 and presented in Table 5 remain qualitatively the same in the new fuller specification, and so need not be discussed again now. Of the new variables added, the results in Table 14 show that, even controlling for GCSE achievement, an individual's Key Stage 3 test score has an independent, positive impact on post-compulsory education participation. Each additional point scored, on average, at Key Stage 3 (with a mean of 36) is associated with a 0.7 percentage point higher post-compulsory participation rate. Term time working during compulsory education is associated with a lower participation rate[29], while more positive attitudes to school are, as expected, associated with a higher probability of participating in post-compulsory education.

[29] The cause of the negative term-time working effect is not that individuals' GCSE performance was impaired by devoting time to work rather than study. Looking at the raw data, 67% of those who worked during term-time in compulsory schooling achieved 5 or more good GCSEs, compared to 61% of those who did not work in term-time. Neither is it an indicator of individuals from poorer backgrounds who were compelled to work, since those who worked in term-time come from families with higher income on average. Perhaps this variable indicates those individuals who are more interested in working than further study, thus explaining the lower post-compulsory participation rate.

The school variables introduced into the LSYPE equation all, with one exception, have marginal effects on individuals' post-compulsory participation rates which are not statistically significant. It seems that, for this cohort and for this particular education outcome, peer effects are not very important. The one exception is a higher rate of post-compulsory participation amongst individuals who attended schools where a higher proportion of pupils had a language other than English as their first language.

The family background results confirm the previously noted positive effect of well-qualified parents on young people's post-compulsory participation. The new variables measuring parents' attitudes show that such attitudes are also of considerable importance, with four of the six included measures of parental attitudes attracting positive and statistically significant coefficients[30].

Tables 15 and 16 report the results for the more aggregated rural classifications, with the full list of control variables still included in the estimated equations (but not reported). Aggregating the rural categories to increase cell sizes and reduce standard errors has no effect on the statistical significance of the marginal effects of these variables, however, so there still seems to be little difference in participation across different types of rural-urban areas in LSYPE.

Table 15: In full time education age 16/17 – full LSYPE: Sparse and less sparse rural definition specification

	LSYPE
Rural definition variables	
Rural – less sparse	0.002 (0.014)
Rural – sparse	0.031 (0.035)

Notes:
1. Estimation by probit. Table reports marginal effects.
2. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.
3. Control variables were included as in the specifications in Table 14.

[30] A fifth, indicating parents who arranged special meetings with teachers, has a *negative* and statistically significant effect on post-compulsory participation. Perhaps the fact that special meetings with teachers were required indicates problems with the young person's education.

Table 16: In full time education age 16/17 – full LSYPE: Type of rural area specification

	LSYPE
Rural definition variables	
Rural – town	-0.002 (0.016)
Rural – village	0.009 (0.018)
Rural – dispersed	0.001 (0.025)

Notes:
1. Estimation by probit. Table reports marginal effects.
2. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.
3. Control variables were included as in the specifications in Table 14.

Table 17 performs a similar decomposition of the rural-urban difference in post-compulsory participation as presented in Table 9 above, but now with the full LSYPE specification.

Table 17: Decomposition of rural-urban difference in full time age 16/17 participation rate – full LSYPE specification

	LSYPE
Raw rural-urban difference	0.032 (0.012)**
Due to characteristics:	0.018 (0.016)
Prior attainment	*0.019 (0.003)***
Individual characteristics	*-0.015 (0.003)***
School characteristics	*0.000 (0.005)*
Family characteristics	*0.002 (0.003)*
Parental attitudes variables	*0.008 (0.004)*
Regional characteristics	*0.004 (0.015)*
Due to coefficients:	0.014 (0.017)

Notes:
1. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.
2. Control variables were included as in Table 14.

Despite the greatly expanded list of explanatory variables, the results of the decomposition are very similar to those reported in Table 9. Thus, just over half of the 3.2 percentage point gap in raw participation rates between rural and urban areas in LSYPE is still explained by the explanatory variables in the model, with the remainder (1.4 percentage points) unexplained. The further analysis in Table 17 shows that it is mainly differences in prior attainment that is explaining the higher

participation in rural areas, as was found earlier. Again as before, individual characteristics serve to reduce the participation gap, with this again being due to high-participating ethnic minorities being much less likely to live in rural areas. Other groups of characteristics play only minor roles in explaining the participation gap, though there is some impact of parental attitudes.

Summary

- Exploiting the much larger range of variables available in LSYPE does not alter the results obtained with that data set.
- The difference in post-compulsory participation rates between urban and rural areas remains very small and not statistically significant in the LSYPE data set, after controlling for other characteristics of young people, their families, schools and local areas.

4.5 Determinants of Post-Compulsory Participation at Age 16/17 by Academic/Vocational Status – Full LSYPE Specification

Table 18 reports the results of a multinomial logit analysis of the three-way decision between academic study, vocational study or no study, estimated with the extended specification including all of the LSYPE variables.

Table 18: In full time education age 16/17 by academic/vocational status – full LSYPE: Full rural definition specification

	LSYPE	
	Academic education	**Vocational education**
Rural definition variables		
Rural town – less sparse	0.000 (0.028)	0.000 (0.022)
Rural town – sparse	0.273 (0.111)*	-0.149 (0.083)
Rural village – less sparse	0.026 (0.033)	-0.002 (0.025)
Rural village – sparse	-0.154 (0.109)	0.076 (0.087)
Rural dispersed – less sparse	0.023 (0.045)	-0.018 (0.035)
Rural dispersed – sparse	-0.005 (0.102)	0.078 (0.076)
Young person characteristics variables		
female	0.052 (0.013)**	-0.017 (0.010)
respondent from ethnic minority	0.166 (0.020)**	-0.041 (0.014)**
whether reached Level 2 in Year 11	0.215 (0.019)**	-0.130 (0.014)**
5+ A*-C GCSEs including Maths and English	0.204 (0.019)**	-0.136 (0.015)**
key stage 3 average points score	0.025 (0.002)**	-0.014 (0.001)**
parent mentions child has special needs	0.020 (0.033)	0.019 (0.022)
whether child currently has statement of SEN	0.112 (0.051)*	-0.014 (0.033)
ever suspended or excluded from school	-0.107 (0.034)**	0.050 (0.023)*
played truant in last year	0.016 (0.021)	0.001 (0.015)
health not very good or not good at all	-0.098 (0.037)**	0.047 (0.027)
had work experience place while at school	-0.005 (0.013)	-0.008 (0.010)
whether has a job during term time	-0.072 (0.015)**	0.025 (0.011)*
whether has caring responsibilities	-0.010 (0.027)	0.014 (0.019)
attitude to school	0.009 (0.001)**	-0.002 (0.001)**
School variables		
independent school	-0.002 (0.053)	0.047 (0.041)
foundation school	0.055 (0.019)**	-0.040 (0.015)**
voluntary aided/controlled school	0.038 (0.021)	-0.030 (0.016)
proportion of 15 year olds at school with SEN	-0.045 (0.074)	0.035 (0.053)
proportion reaching Level 2 at school	0.091 (0.065)	-0.060 (0.049)
KS3-GCSE value added at school	-0.001 (0.000)*	0.001 (0.000)
unauthorised absence rate at school	0.933 (0.549)	-0.907 (0.399)*
proportion eligible for free meals at school	-0.165 (0.085)	0.105 (0.061)
proportion at school whose first language not English	0.248 (0.050)**	-0.120 (0.036)**
Family background variables		
live in owner-occupied house	-0.021 (0.029)	0.010 (0.021)
lives in a council house	-0.041 (0.031)	0.016 (0.022)
father not present	0.039 (0.039)	-0.025 (0.028)
mother not present	-0.048 (0.042)	0.008 (0.031)
father in professional/managerial job	0.010 (0.026)	0.001 (0.019)
father in intermediate job	-0.038 (0.026)	0.010 (0.019)
father in routine job	-0.046 (0.026)	0.024 (0.019)
mother in professional/managerial job	-0.017 (0.022)	-0.008 (0.016)
mother in intermediate job	-0.038 (0.021)	0.014 (0.016)
mother in routine job	-0.040 (0.020)*	0.014 (0.015)

	LSYPE	
	Academic education	Vocational education
mother has degree	0.098 (0.028)**	-0.042 (0.023)
mother has other HE	0.035 (0.022)	0.001 (0.017)
mother has A-levels	0.009 (0.020)	0.015 (0.015)
father has degree	0.128 (0.030)**	-0.049 (0.024)*
father has other HE	0.035 (0.027)	-0.029 (0.021)
father has A-levels	-0.003 (0.022)	-0.003 (0.016)
high family income	-0.047 (0.017)**	0.024 (0.012)
lives with both natural parents	0.069 (0.021)**	-0.025 (0.015)
lives in a single parent family	-0.011 (0.036)	0.035 (0.025)
number of siblings living in household	-0.000 (0.006)	0.000 (0.004)
number of risk factors faced	-0.030 (0.006)**	0.004 (0.004)
Parental attitude variables		
whether attend parents' evenings	0.050 (0.020)*	-0.000 (0.014)
parent arranged special meetings with teachers	-0.018 (0.015)	-0.006 (0.011)
parent very or fairly involved in child's school life	0.016 (0.016)	-0.013 (0.012)
whether paid for private tuition in last year	0.102 (0.019)**	-0.023 (0.014)
parent wants child to continue in pc education	0.365 (0.024)**	-0.094 (0.016)**
parents will pay expenses of staying in pc education	0.022 (0.022)	0.034 (0.015)*
Regional variables		
distance from nearest academic institution (km)	-0.008 (0.004)*	0.007 (0.003)*
distance from nearest vocational institution (km)	0.004 (0.002)*	-0.004 (0.001)**
regional index of local deprivation	-0.107 (0.053)*	0.033 (0.038)
Number of observations	11,226	

Notes:
1. Estimation by multinomial logit. Table reports marginal effects.
2. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.

In Table 10, there were no statistically significant effects of region type on either academic or vocational participation in the reduced 'YCS-equivalent' specification using LSYPE, with the exception of a positive rural town – sparse effect on post-compulsory academic education participation. As expected, in the extended specification, this is still very much the case, the continued single exception being that of the rural town – sparse areas which has a significantly higher participation rate in academic post-compulsory education as compared to similar urban areas.

With respect to the covariates, many of the comments made in Section 4.2 with respect to the probit equation for overall participation are mostly relevant again for the multinomial logit specification. Thus, many of the factors that influenced overall participation are also associated with a greater likelihood of academic study and a

lower likelihood of vocational study, with the measures of prior attainment once again the dominant explanatory variables. Parental variables such as education, occupation and attitudes seem to be of slightly less importance in explaining the academic/vocational split than in explaining overall participation, with the exception of parents being graduates.

The regional variables have statistically significant coefficients when explaining participation in post-compulsory study. In particular, young people living in more deprived areas are significantly less likely to engage in academic study. The distance variables show that, as before, the further a young person lives from an education institution offering academic qualifications, the less likely they are to undertake further academic study; and the further a young person lives from an education institution offering vocational qualifications, the less likely they are to undertake further vocational study. Furthermore, the further they live from one type of provision, the more likely they are to pursue to the other type of education, with the effects being statistically significant.

Tables 19 and 20 report the marginal effects on the rural variables when categories are aggregated together, for the multinomial logit estimated on the extended LSYPE specification. None of the marginal effects in either table are statistically significant, however.

Table 19: In full time education age 16/17 by academic/vocational status – full LSYPE: Sparse and less sparse rural definition specification

	LSYPE	
	Academic education	**Vocational education**
Rural definition variables		
Rural – less sparse	0.009 (0.024)	-0.002 (0.018)
Rural – sparse	0.037 (0.069)	-0.001 (0.052)

Notes:
1. Estimation by multinomial logit. Table reports marginal effects.
2. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.
3. Control variables were included as in the specifications in Table 18.

Table 20: In full time education age 16/17 by academic/vocational status – full LSYPE: Type of rural area specification

	LSYPE	
	Academic education	Vocational education
Rural definition variables		
Rural – town	0.010 (0.028)	-0.006 (0.022)
Rural – village	0.008 (0.032)	0.004 (0.025)
Rural – dispersed	0.010 (0.043)	-0.002 (0.033)

Notes:
1. Estimation by multinomial logit. Table reports marginal effects.
2. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.
3. Control variables were included as in the specifications in Table 18.

Summary

- Extending the number of control variables used in the multinomial logit analysis of post-compulsory participation by type of education, using the wider range of variables available in LSYPE, does not alter the results obtained with that data set.
- There are no statistically significant differences between rural and urban areas in LSYPE in either academic or vocational post-compulsory participation rates (with the exception of the 'rural town – sparse' category into which few individuals fall).

4.6 Determinants of Post-Compulsory Participation at Age 17/18

The final section of results in this report make use of the fact that the YCS surveys individuals three or four times after they complete compulsory education, and so respondents are observed after the first year of post-compulsory education. Table 21 therefore reports the results of estimating a probit equation for whether individuals are still participating in full-time education in the second year after they completed compulsory education, obviously for the YCS only[31].

[31] The next wave of the LSYPE will also correspond to the second year after compulsory education has been completed, and thus a similar analysis could be undertaken when that data becomes available.

Table 21: In full time education age 17/18: Full rural definition specification - YCS

	YCS	
	Without controls	**With controls**
Rural definition variables		
Rural town – less sparse	-0.005 (0.015)	0.010 (0.015)
Rural town – sparse	-0.079 (0.065)	-0.000 (0.054)
Rural village – less sparse	0.018 (0.016)	0.009 (0.018)
Rural village – sparse	-0.050 (0.071)	-0.021 (0.065)
Rural dispersed – less sparse	0.071 (0.021)**	0.063 (0.019)**
Rural dispersed – sparse	-0.023 (0.072)	0.055 (0.051)
Young person characteristics variables		
Female		0.035 (0.008)**
respondent from ethnic minority		0.119 (0.008)**
whether reached Level 2 in Year 11		0.186 (0.017)**
5+ A*-C GCSEs incl. Maths and English		0.124 (0.015)**
has health problem or disability		0.028 (0.019)
played truant in Year 11		-0.119 (0.011)**
expelled or suspended in Years 10 or 11		-0.093 (0.021)**
School variables		
at grammar/indep. school in Year 11		0.056 (0.012)**
Family background variables		
live in owner-occupied house		0.009 (0.018)
lives in a council house		-0.007 (0.021)
live with father only		-0.008 (0.019)
live with mother only		0.008 (0.014)
live with neither parent		-0.039 (0.028)
only mother employed		-0.032 (0.020)
only father employed		-0.034 (0.019)
both parents employed		-0.026 (0.015)
father's occupation in SOC 1-3		0.033 (0.009)**
mother's occupation in SOC 1-3		0.024 (0.010)*
at least one parent with a degree		0.056 (0.010)**
at least one parent with A-levels		0.025 (0.010)*
Regional variables		
distance from academic institution (km)		-0.004 (0.002)
distance from vocational institution (km)		-0.002 (0.001)*
regional index of local deprivation		-0.071 (0.031)*
Number of observations	9,282	8,836

Notes:
1. Estimation by probit. Table reports marginal effects.
2. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.

The results in the first column of Table 21 are from a specification with no control variables included, and so represent the raw differences in Year 2 post-compulsory

education participation across area types. It can be seen that individuals living in rural dispersed – less sparse areas are 7.1 percentage points more likely than urban individuals to still be participating. None of the other differences are statistically significant.

The second column of results shows that, even when a range of control variables are added to the specification, the rural dispersed – less sparse marginal effect remains statistically significant. Thus, young people living in rural dispersed – less sparse areas are 6.3 percentage points more likely than young people in urban areas, with the same observed characteristics, to be in full-time education two years after completing compulsory schooling.

Considering the control variables, many of the factors associated with post-compulsory participation in the first year remain associated with post-compulsory participation in the second year. Thus, prior attainment remains a key determinant, and females and those individuals belonging to ethnic minorities are more likely to still be participating as compared to males and whites. Parental education and occupation also remain important determinants of Year 2 participation.

Tables 22 and 23 aggregate the rural categories as before. The results in Table 23 show that young people living in rural dispersed areas are significantly more likely than young people in urban areas to be participating in full-time education two years after completing compulsory education, even when observable characteristics are held constant. The magnitude of the difference is around 6 percentage points. There are no differences in this participation rate between urban individuals and those with the same observed characteristics living in rural villages or rural towns.

Table 22: In full time education age 17/18: Sparse and less sparse rural definition specification - YCS

	YCS	
	Without controls	With controls
Rural definition variables		
Rural – less sparse	0.015 (0.010)	0.017 (0.013)
Rural – sparse	-0.054 (0.040)	0.007 (0.037)

Notes:
1. Estimation by probit. Table reports marginal effects.
2. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.
3. Control variables were included as in the specifications in Table 21.

Table 23: In full time education age 17/18: Type of rural area specification - YCS

	YCS	
	Without controls	With controls
Rural definition variables		
Rural – town	-0.009 (0.014)	0.011 (0.015)
Rural – village	0.014 (0.016)	0.008 (0.018)
Rural – dispersed	0.061 (0.020)**	0.063 (0.019)**

Notes:
1. Estimation by probit. Table reports marginal effects.
2. Standard errors in parentheses. * denotes significant at 5%; ** denotes significant at 1%.
3. Control variables were included as in the specifications in Table 21.

Summary

- Further sweeps of the YCS data set continue to track respondents beyond their first post-compulsory year, so can be used to analyse determinants of participation in the second post-compulsory year.
- The results reveal that young people in rural dispersed areas are significantly more likely than those in urban areas to still be in full-time education two years after completing compulsory schooling, even when controlling for observed characteristics and so comparing like-with-like.

5. Conclusions

This report has examined differences in participation rates in post-compulsory education between young people living in rural and urban areas. Using two data sets which have been compiled to explicitly examine young people's transitions from school to further and higher education and/or the world of work, we document the differences in their education participation rates, principally in the first year after they have completed their compulsory education. Overall participation rates are slightly higher in rural than in urban areas in both of our data sets. Our main objective in this report is to investigate the factors that determine post-compulsory education participation in order to understand the reasons underlying the small differences in rural-urban participation that are observed in the data.

Participation in post-compulsory education can largely be explained by a number of factors associated with individuals' personal characteristics, their family background and their schooling. Across both data sets, the single most important determinant of post-compulsory participation is prior attainment, as measured by students' GCSE performance in Year 11. Other important factors which are positively and statistically significantly related to post-compulsory participation in *both* data sets include: being female; from an ethnic minority; at a grammar or independent school in Year 11; and having at least one parent with a degree. Factors which are negatively and statistically significantly related to post-compulsory participation in both data sets include: having played truant or been suspended/expelled in Year 11; living in an area with a high level of deprivation; and having both parents employed (which is related to parents' education, which therefore seems to pick up much of its effect).

Looking at the characteristics of young people in rural areas and urban areas, those in rural areas are more likely to have many of the positive factors in the previous paragraph and fewer of the negative factors. In particular, rural young people have significantly higher prior attainment at GCSE level, have better educated parents working in more senior occupations, and live in less deprived areas.

The results using the YCS data show that even after controlling for these various advantages enjoyed by rural young people, those living in rural dispersed or rural village areas (though not rural towns) are still more likely to participate in post-compulsory education than those with the same observed characteristics living in urban areas, by around 3 percentage points (Table 8). In other words, comparing two otherwise identical individuals except for the fact that one lives in a rural setting and the other in an urban setting, we can expect the rural resident to have a 3 percentage point higher probability of participating in post-compulsory education. Rural young people are therefore participating in post-compulsory education with a slightly higher likelihood than their personal, family and regional characteristics would predict.

This positive participation differential in the YCS data set in favour of rural areas is larger when attention is focussed on participation in specifically academic post-compulsory education. Now young people in all rural areas, including towns in this case, enjoy a higher academic post-compulsory participation rate than young people with the same characteristics but living in urban areas, by around 5-6 percentage points (Table 13). There is also some weaker evidence that individuals in rural areas are less likely to participate in vocational post-compulsory education than their urban counterparts.

The results in the previous two paragraphs are for the YCS data set only. These results are not replicated in the LSYPE data set, where the similar sized rural-urban difference in participation in the raw data (Table 4) is explained to a greater extent by the characteristics of the young people, and of their families, schools and areas. Almost all rural-urban differences in participation in the estimated equations using LSYPE data throughout the report are therefore statistically insignificant. This conclusion is supported by the results of the decomposition analysis in Table 9, in which the raw rural-urban differential in participation rates is left unexplained in the YCS data, but more than half is explained in the LSYPE data.

In terms of policy implications of the results, few of the explanatory variables in the estimated models are easily affected by policy in the short-term. The key determinants of participation in full-time education have been shown to be prior

attainment and family background, neither of which can be directly and immediately altered, and so are not direct policy levers. It is therefore more the case that policy needs to recognise these influences on participation and work with them, for example by designing and offering post-compulsory education and training that is of interest and within the capabilities of those with lower prior attainment, by providing information about post-compulsory education to families containing no members with direct experience of it, and by addressing any financial reasons for low participation amongst less well-off families. Of course, the participation rate will be raised anyway to 18 by 2015, requiring individuals to continue in some form of education or training until that age. The results of this report still remain relevant in such a situation, however. The report has shown that certain characteristics are associated with a lower participation in full-time education. Simply forcing individuals with such characteristics into full-time education until age 18 would not address these reasons for their lack of desire to do so voluntarily, and would likely be of little value to such individuals or society. It is therefore important that alternative forms of age 16-18 learning, other than full-time education, are available for those groups who currently have a lower likelihood of participating in such learning.

Finally, as for rural issues, the results have shown that there is little need for an explicit rural policy on post-compulsory participation. The raw data show that participation rates are slightly higher in rural areas compared to urban areas, and even when we allow for the characteristics of rural areas and the families who live there, there is no evidence that young people in rural areas are participating less than their characteristics would have predicted them to. Indeed, in the YCS results at least, the reverse is true, and young people in rural areas have a slightly higher participation rate than would be predicted on the basis of their characteristics. The one factor influencing participation that could be considered for rural policy is the impact of distance to nearest education institution, which is on average considerably longer in rural areas, and which was shown by the results to reduce the likelihood of participation the greater the distance involved, holding other things constant.

References

Ashford, S., Gray, J. and Tranmer, M. (1993). 'The introduction of GCSE exams and changes in post-16 participation.' Employment Department Research Series, Youth Cohort Report no. 23.

Ashworth, K., Hardman, J., Hartfree, Y., Maguire, S., Middleton, S., Smith, D., Dearden, L., Emmerson, C., Frayne, C., and Meghir, C. (2002). 'Education maintenance allowance: the first two years. A quantitative evaluation.' Department for Education and Skills, Research Report 352.

Bailey, N. and M. Livingston (2008). 'Selective migration and neighborhood deprivation: evidence from 2001 Census Migration Data for England and Scotland.' *Urban Studies*, 45, 943-961.

Blanden, J. and Gregg, P. (2004) 'Family income and educational attainment: a review of approaches and evidence for Britain.' *Oxford Review of Economic Policy*, 20, 245-263.

Blanden, J. and Machin, S. (2004) 'Educational inequality and the expansion of UK higher education.' *Scottish Journal of Political Economy*, 51, 230-250.

Cheng, Y. (1995). 'Staying on in full-time education after 16: do schools make a difference.' Employment Department Research Series, Youth Cohort Report no. 37.

Clark, D. (2002). 'Participation in post-compulsory education in England: what explains the boom and bust?' Centre for the Economics of Education Discussion Paper 24.

Conlon, G. (2005). 'The determinants of undertaking academic and vocational qualifications in the United Kingdom.' *Education Economics*, 13, 299-313.

Foskett, N., Dyke, M. and Maringe, F. (2008). 'The influence of the school in the decision to participate in learning post-16.' *British Educational Research Journal*, 34, 37-61.

Garner, C.L. and Raudenbush, S.W. (1991). 'Neighborhood effects on educational attainment: A multilevel analysis.' *Sociology of Education*, 64(4), 251-262.

Gibbons, S. and Silva, O. (2008). 'Urban density and pupil attainment.' *Journal of Urban Economics*, 63, 631-650.

Gray, J., Jesson, D. and Tranmer, M. (1993). 'Boosting post-16 participation in full-time education. A study of some key factors.' Employment Department Research Series, Youth Cohort Report no. 20.

Gray, J., Jesson, D. and Tranmer, M. (1994). 'Local labour market variations in post-16 participation: evidence from the end of the eighties.' Employment Department Research Series, Youth Cohort Report no. 26.

Green, R. (2009). *Educational Attainment in Rural Areas*. Forthcoming Defra report prepared by Natcen.

Haveman, R. and Wolfe, B. (1995). 'The determinants of children's attainments: A review of methods and findings.' *Journal of Economic Literature*, 33, 1829-1878.

James, R. (2002). 'Socioeconomic background and Higher Education participation.' Commonwealth Department of Education Science and Training.

Jenkins, A., Greenwood, C. and Vignoles, A. (2007). *The Returns to Qualifications in England: Updating the Evidence Base on Level 2 and Level 3 Vocational Qualifications*. Centre for the Economics of Education Discussion Paper 89.

Le A. T. and Miller, P.W. (2005). 'Participation in Higher Education: equity and access?' *The Economic Record*, 81(253), 152-165.

Leckie, G. (2008). 'Modelling the effects of pupil mobility and neighbourhood on School differences in educational achievement.' CMPO Working Paper 08/189., January 2008.

Lenton, P. (2005). 'The school-to-work transition in England and Wales.' *Journal of Economic Studies*, 32, 88-113.

McCulloch, A. and Joshi, H. (2001). 'Neighbourhood and family influences on the cognitive ability of children in the NCDS.' *Social Science and Medicine*, 53(5), 579-591.

McIntosh, S. (2001). 'The demand for post-compulsory education in four European countries.' *Education Economics*, 9, 69-90.

McVicar, D. and Rice, P. (2001). 'Participation in further education in England and Wales: an analysis of post-war trends.' *Oxford Economic Papers*, 53, 47-66.

Micklewright, J. (1989). 'Choice at sixteen.' *Economica*, 56, 25-39.

Payne, J. (1998). 'Routes at sixteen. Trends and Choices in the Nineties.' Department for Education and Employment Research Report 55.

Pissarides, C.A. (1981). 'Staying-on at school in England and Wales.' *Economica*, 48, 345-363.

Raudenbush, S.W. (1993). 'A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research.' *Journal for Educational Statistics*, 18(4), 321-349.

Rice, P. (1999). ‘The impact of local labour markets on investments in further education: evidence from the England and Wales Youth Cohort Studies.’ *Journal of Population Economics*, 12, 287-312.

Robertson, C. (1990). ‘Participation in post-compulsory schooling in Scotland.’ University of Strathclyde, December 1990, mimeo.

Roussel, S. (2000). ‘Factors influencing participation in post-secondary education and training in Australia: 1989 to 1997.’ Research and Evaluation Branch, Department of Education, Training and Youth Affairs, Australia, REB Report 7/2000.

Ulubasoglu, M.A. and Cardak, B.A. (2007). ‘International comparisons of rural-urban educational attainment: Data and determinants.’ *European Economic Review*, 51, 1828-1857.

Whitfield, K. and Wilson, R.A. (1991). ‘Staying on in full-time education: the educational participation rate of 16-year-olds.’ *Economica*, 58, 391-404.

Yan, W. (2003). ‘Post-secondary enrolment and persistence of students from rural Pennsylvania,’, mimeo Indiana University of Pennsylvania.