

This article was downloaded by: [Cardiff University]

On: 19 July 2013, At: 00:23

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Comparative Education

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/cced20>

### Devolution and geographies of education: the use of the Millennium Cohort Study for 'home international' comparisons across the UK

Chris Taylor <sup>a</sup>, Gareth Rees <sup>a</sup> & Rhys Davies <sup>a</sup>

<sup>a</sup> WISERD , Cardiff University , Cardiff , UK

Published online: 05 Jul 2013.

To cite this article: Comparative Education (2013): Devolution and geographies of education: the use of the Millennium Cohort Study for 'home international' comparisons across the UK, Comparative Education, DOI: 10.1080/03050068.2013.802927

To link to this article: <http://dx.doi.org/10.1080/03050068.2013.802927>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Versions of published Taylor & Francis and Routledge Open articles and Taylor & Francis and Routledge Open Select articles posted to institutional or subject repositories or any other third-party website are without warranty from Taylor & Francis of any kind, either expressed or implied, including, but not limited to, warranties of merchantability, fitness for a particular purpose, or non-infringement. Any opinions and views expressed in this article are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor & Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms &

Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

## Devolution and geographies of education: the use of the Millennium Cohort Study for ‘home international’ comparisons across the UK

Chris Taylor\*, Gareth Rees and Rhys Davies

*WISERD, Cardiff University, Cardiff, UK*

Following political devolution in the late 1990s and the establishment of the governments for Wales and Scotland, the education systems of the four home countries of the UK have significantly diverged. Consequently, not only does that mean that education research in the UK has to be sensitive to such divergence, but that the divergence of policy and practice provides an important opportunity to undertake comparative research within the UK. Such ‘home international’ comparisons between the four home countries of the UK also provide the opportunity to undertake ‘natural experiments’ of education policy and practice across similar socio-economic contexts. By drawing specifically on the UK Millennium Cohort Study (MCS) – a recent longitudinal birth cohort study specifically designed to provide the potential for geographical analysis – the paper finds considerable variation in child development by country of the UK, with no single story of ‘success’. However, the paper finds that literacy development amongst children in England is, particularly in London, on average, greater than for children elsewhere. The paper concludes by arguing that ‘home international’ comparisons must take seriously issues of scale and geography when interpreting the influence of ‘national’ education systems and policies on educational outcomes.

### Introduction

Educational research has been rather slow to absorb the implications of the ‘spatial turn’ that has been so influential across the social sciences (and the humanities) in recent years (for example, Warf and Arias 2008). To be sure, there is a long and distinguished tradition of scholarship in comparative education, which has been concerned overwhelmingly with comparisons between different states (Crossley, Broadfoot, and Schweisfurth 2007). More recently, interest has begun to grow in the analysis of educational issues at the supra-national level (for example, Ozga et al. 2011). However, the more complex geographies of spatial variation *within* national states, whilst *implicit* in much educational research, have much less frequently been analysed systematically (Taylor 2009; Thiem 2009).

This relative neglect of the regional and more local scales in the geographical analysis of educational issues is surprising. Educational outcomes (such as levels of educational attainment, for example) exhibit distinctive *spatial* distributions (at these scales), as well

---

\*Corresponding author. Email: taylorcm@cardiff.ac.uk

as the much better recognised differentiations between social groups (defined most frequently in terms of gender, ethnic background or social class) or distinctive institutional contexts (such as type of school attended). Indeed, these different dimensions of educational differentiation are closely interlinked. Hence, at one level, these local and regional distributions of educational phenomena themselves reflect the spatial patterning of social and economic conditions; the inequalities that are known to shape access to educational opportunities and consequent outcomes are themselves significantly differentiated between different regions and localities (for example, Butler and Hamnett 2007). Moreover, there are important questions as to the extent to which the effects of geographical area on educational outcomes amount to more than the impacts of the social and economic conditions which characterise each area (Atkinson and Kintrea 2001).

Equally, educational policies which are responsible for shaping the availability of educational opportunities also have impacts that are differentiated at the local and regional levels. It is true that many educational policies originate at the national level; indeed, there has been a significant trend towards increasing centralisation of control over education policy in the UK (and in England especially) in recent decades (for example, Ball 2008). However, some educational policies are targeted at local or, much less frequently, regional scales, in order to focus provision on geographical areas that are characterised by particular social and economic conditions (usually high levels of social disadvantage) (Power, Rees, and Taylor 2005). Moreover, the *effects* of national-level policies continue to be mediated significantly at the local level, whether through local education authorities or individual schools and colleges. In part, this reflects the capacities of actors at sub-national levels to shape the implementation of policies in particular ways (Ball, Maguire, and Braun 2012). Equally, however, there are complex *interactions* between national policy initiatives and social and economic circumstances that produce outcomes that are inevitably differentiated between different local areas and, indeed, regions. As Byrne, Williamson, and Fletcher (1975) demonstrated some time ago, the same national-level policies (in relation to educational expenditure, for example) impacted on different localities and regions in highly distinctive ways, reflecting the variations in pre-existing conditions in these diverse places.

Clearly, sorting out the effects of these various dimensions of sub-national, geographical variation in education poses complex analytical issues. One innovative approach to addressing some of these issues (at least in the context of the UK) is that proposed by Raffe et al. (1999), involving what they termed ‘home international’ comparison. Here, the differences in policy approaches between the constituent countries of the UK – England, Northern Ireland, Scotland and Wales – provide a means of analysing the differential impacts of these policies on educational outcomes; thereby permitting an exploration of the determinants of differences in educational outcomes at the sub-national level. Whilst there are undoubtedly some differences in social and economic conditions between the ‘home countries’, these are significantly less marked than is the case in fully international comparisons; thereby introducing elements of ‘controlling’ these exogenous factors and enabling the generation of important new insights, both theoretically and empirically. In short, then, ‘home international’ comparisons offer an innovative *hybrid* in terms of geographical scales; ‘national’-level policies may be compared between the ‘home countries’, whilst the wider UK provides a significant element of commonality, thereby making the attribution of relationships between policies and educational outcomes somewhat clearer.

Moreover, since the advent of parliamentary devolution in 1999, the constitutional framework within which ‘home international’ comparisons can be made, has shifted

significantly. The UK government has retained responsibility for educational provision in England. However, in the other ‘home countries’, it has become the responsibility of devolved governments in Belfast, Cardiff and Edinburgh. This, in turn, has provided a context for the increasing divergence in policy approaches between the ‘home countries’; and, more specifically, between England and the other countries (Greer 2009). In the case of education policy, the UK government has adopted an increasingly radical programme of change that has led to important differences with the other ‘home countries’ across the range of educational provision, from early years to higher education and lifelong learning. This is not to suggest, however, that UK-wide commonalities have wholly disappeared. Indeed, the exact extent and significance of policy divergence between the ‘home countries’ has become a matter of considerable debate (for example, Raffe 2006; Rees 2007). Nevertheless, it is clear that ‘home international’ comparisons do hold out considerable promise for policy analysis, in terms of permitting a more robust exploration of the effects of policy interventions at a geographical scale below that of the UK state than would otherwise be possible.

Given this, it is undoubtedly unfortunate that these already complex methodological waters have been muddied by the use of extremely crude ‘home international’ comparisons by politicians and the media, in the context of debates about the efficacy of the educational policies adopted in the different UK jurisdictions. More specifically, such comparisons between England and Wales have been especially influential.<sup>1</sup> Indeed, in Wales, there is now a widespread view that the educational system is in ‘crisis’, largely on the basis of – at best – very partial comparisons between Welsh and English levels of educational attainment; and also what is perceived to be Wales’s ‘under-performance’ in international comparisons such as PISA, not only in relation to the other countries of the UK, but also more widely. Moreover, this popular account of the state of Welsh education has had tangible results in terms of significant shifts in educational policy, with new interventions in relation, for example, to raising ‘teaching quality’, evaluating the performance of schools, improving levels of literacy and numeracy, and shifting responsibility for implementing educational provision away from the local authorities (Rees 2012).

What is most immediately problematic about the use of these ‘home international’ comparisons in such political contexts is straightforwardly that their methodological basis is so crude. For example, the well-established limitations of PISA in providing a satisfactory basis for evaluating the effects of systems of educational provision are nowhere acknowledged (Goldstein 2008). Similarly, comparisons of levels of educational attainment are frequently made without recognising the difficulties in constructing comparable datasets for the different ‘home countries’ (Rees 2012); or taking adequate account of the effects of differences in social and economic conditions between them (Gorard 2000). Accordingly, if ‘home international’ comparisons are to be used as a method of policy evaluation – and, potentially, policy learning (Raffe et al. 1999) – it is essential that they are conducted on an appropriately robust methodological basis.

One approach – but by no means the only one – to establishing this sort of robust methodological basis is through the conduct of ‘home international’ comparisons as what have been termed ‘natural experiments’. Here, ‘naturally occurring’ events or circumstances can be used to replicate some of the characteristics of an experiment in a wholly observational study (Dunning 2012). Hence, the adoption of a different policy in one ‘home country’ (the ‘treatment’ group) allows systematic comparison with another (the ‘control’ group), thereby enabling the delineation of the effects of

the policy innovation, assuming, of course, that other differences between the two ‘home countries’ are limited or can, in some way, be allowed for in the analysis (generally, by means of statistical applications). It is instructive, however, that relatively few such formal ‘natural experiments’ have actually been carried out in the context of ‘home international’ comparisons, especially in the area of education policy [see, however, Burgess, Wilson, and Worth (2010) for an interesting, but flawed example]. And one problematical issue here is the availability of appropriate data to enable this kind of formal analysis.

In this paper, we address this issue of the availability of appropriate data to make stringent ‘home international’ comparisons by drawing upon the Millennium Cohort Study (MCS). This is a large-scale, birth cohort study of children born across the UK during 2000–2001, with data currently available up until they are seven years old. As with other birth cohort studies, it has the advantage of tracking individuals over time, thereby providing a more robust means of measuring the effects of educational changes than is provided, for example, in the repeated cross-sectional datasets that are more frequently used in analysis of this kind (as, for example, in comparing the GCSE attainments of successive cohorts of young people) (compare, for example, Goldstein 2004). In addition, the MCS is unusual in its use of geographical criteria in the selection of its sample of respondents. It therefore provides an excellent basis on which to carry out geographical comparisons of children growing up in the different ‘home countries’ of the UK and, indeed, in different areas within these ‘home countries’ (a consideration which is especially relevant for England, given its much greater population size than the other countries).

We also attempt to develop the methodological basis on which ‘home international’ comparisons can be made. Here, we apply novel statistical techniques – propensity score matching – which, we argue, produce *much* more robust comparisons than those – beloved of politicians and the media – which are based upon ‘raw score’ measures of educational attainment. They also have some advantages over more conventional statistical approaches (based on various forms of regression analysis), especially in relation to the clarity with which results can be presented publicly.

Finally, we explore some of the complexities that arise when ‘home international’ comparisons take account of regional variations *within* England. Considerable interest has been aroused recently in such regional variations, especially as a result of the dramatic improvements in levels of educational attainment that have been recorded over recent years in London (Cook 2012), with the inevitable speculations that these improvements are the result of specific policy interventions, such as London Challenge, or of national policy initiatives that have been particularly prevalent in the UK capital, such as academies and Teach First (Whitty and Anders forthcoming). In fact, of course, the ‘London effect’ is more appropriately understood as one example of a much wider phenomenon of geographical variation at a sub-national scale. As the analyses which follow demonstrate, exploring variations in educational issues at such sub-national scales, whilst methodologically complex, opens up the potential for important new insights, in relation both to the impacts of policies, as well as more fundamental theoretical concerns.

## **Methodology**

A key limitation to any comparative study is the issue of ‘equivalence’. For example, Gorard (2000) demonstrated that important differences in the number of pupils eligible

for free school meals could account for many of the ‘apparent’ differences in educational achievement between England and Wales. This is an important observation to make, particularly for policy-makers who may be concerned about the extent to which differences in the education systems are responsible for differences in educational achievement. However, the issue of equivalence in comparative geographical studies extends well beyond just accounting for the number of pupils from low-income families. For example, to what extent should such comparisons take in to account other factors that may help determine educational achievement?

Related to this is the availability of data or information relating to the characteristics of children. One of the benefits of ‘home international’ comparisons is that this reduces a number of major differences between the comparison groups that may otherwise have had to be accounted for. However, even then there is the constraint of having access to common or equivalent outcome measures. For example, Scotland has a very different qualifications system to the rest of the UK which means it rarely gets compared to elsewhere in the UK.

In order to address many of these issues we draw upon the first four sweeps of the Millennium Cohort Study (MCS) – a large-scale birth cohort study of approximately 19,000 children born across the UK over 12 months in 2000–2001. The first four sweeps of data collection were conducted when the children were around nine months, three years, five years and seven years old.<sup>2</sup>

The MCS has a number of qualities that make geographical comparisons of the different education systems of the UK possible. First, it collects very detailed information relating to the children in the cohort, including information about their families and, more recently, about their education and schooling. Crucially, it collected information about the children from birth and as they have grown up, helping to observe changes over their life-course. This makes the analysis that follows very distinct from previous ‘home international’ comparisons that rely on cross-sectional data. As Goldstein (2004) argues ‘With only cross sectional data it is very difficult, if not impossible, to draw satisfactory inferences about the effects of different educational systems’ (328).

Second, despite including children in different education systems of the UK it undertook the same assessments (physical, cognitive and behavioural) of the children that allow results to be compared, irrespective of where a child lives and grows up. Third, the sample design for the MCS is very different from previous national longitudinal cohort studies in that children were selected to participate in the MCS based on where they lived. Importantly, two geographical criteria were used in the selection of the sample at: (i) country-level and (ii) ward-level. At the country-level the MCS includes an over-representation of children from Wales, Scotland and Northern Ireland. This means that the MCS sampling design ensures there are enough children in the cohort for meaningful comparisons between England and these other countries, despite their relatively smaller sizes.<sup>3</sup> Furthermore, the MCS cohort was also sampled on the basis of the neighbourhood where they were born; children born between September 2000 and August 2001 were included in the MCS sample if they lived in selected electoral wards across the UK. These wards were selected on the basis of a local measure of child poverty,<sup>4</sup> which meant that the MCS was able to include an over-representation of children born in the most deprived areas of the UK.<sup>5</sup> The natural geographical clustering of children in the MCS cohort that arises from this sampling design has also meant that many children in the MCS share similar ‘geographical’ characteristics that provide a basis of further geographical analyses,

particularly in relation to neighbourhood quality (Ketende, McDonald, and Joshi 2010), but also means that the cohort can be compared at a regional level (within England) and that urban–rural differences can be observed within the cohort (Joshi, Dodgeon, and Hughes 2008).

One of the best examples of the comparative use of the MCS is by Jones, Blackaby, and Murphy (2010) who compare the health and cognitive outcomes of children up to age five years in Wales with other parts of the UK. They also consider intra-regional (within-Wales) differences, comparing children from advantaged wards with children from disadvantaged wards within Wales. The analysis we present here extends the work of Jones et al. by focusing on education-related outcomes when the children were aged seven years. The analysis is divided into two parts. First, we are interested in how children from different countries of the UK compare in terms of their educational development. In this analysis we are also interested in how inequalities in child development at three years of age in each country continue to develop or reduce as the children grow up. In the second part of the analysis we compare the same outcomes at the regional level. Not only are we interested in identifying the presence of any regional ‘effects’ on the outcomes, such as a London ‘effect’, we are also interested in how different regions of England compare with other countries of the UK. Throughout both sets of analyses we identify a number of exogenous factors that are used to control for the background characteristics of the children when they were born. We also incorporate in to both sets of analyses two additional geographical factors: whether the children live in urban, rural or mixed neighbourhoods and the level of neighbourhood deprivation they live in. In addition we also examine the mediating influence of the home learning environment on outcomes. Table 1 outlines the exogenous factors, the mediating geographical factors and outcome measures used in the subsequent analyses.

Although there is a great deal of commonality in the two sets of analyses (we tend to compare the same outcome measures and use the same exogenous factors to control for the children’s background characteristics) we use different approaches and analytical techniques. For the comparative analysis by country of the UK we use the technique of propensity score matching in an attempt to simplify the way in which we present differences in the outcomes of children by country whilst ensuring we are comparing ‘equivalent’ groups of children. Propensity score matching works by matching children from one country (the ‘control’ group) to children from the country of interest (the ‘intervention’ group) based on sharing similar characteristics (Rosenbaum and Rubin 1983). In practice we are interested in identifying a sub-group of children in the MCS living in England who share similar characteristics with children in the MCS living in Wales. We then repeat this to identify a sub-group of children in the MCS living in England who share similar characteristics with children in the MCS living in Scotland.<sup>6</sup>

Matching of children is derived from using a logistic model that estimates the probability of being in the intervention group based upon observed characteristics (the exogenous factors identified in Table 1). These probabilities are calculated for members of both the ‘control’ and ‘intervention’ groups. There are a number of ways of determining which members of the ‘control’ and ‘intervention’ groups are included in the final matched samples.<sup>7</sup> In the results that follow we present the results from two matching estimators that could be considered to reflect examples of a more liberal ‘matching’ approach and a more ‘conservative’ approach to matching,<sup>8</sup> but generally results do not vary significantly in the way the matching is undertaken.

Table 1. Variables used in comparative analyses.

Exogenous variables*	Mediating variables	Outcome variables
Gender of child	Country at age seven	Naming vocabulary at age three <sup>d</sup>
Household structure (e.g. single-parent families) at nine months	Government Office Region (GOR) at age seven	Naming vocabulary at age five <sup>d</sup>
Highest social class (of one or both parents) at nine months	Urban–rural combined UK classification at age three years <sup>a*</sup>	Word reading ability at age seven <sup>d</sup>
Ethnicity of child	Neighbourhood disadvantage (Index of Multiple Deprivation in deciles) at age seven <sup>b*</sup>	Maths ability at age seven <sup>d</sup>
Highest NVQ level (of one or both parents) at nine months		Pattern construction at age seven <sup>d</sup>
Mothers age at birth	Home learning environment (HLE) at age three <sup>c</sup>	Personal wellbeing at age seven <sup>e</sup>
Number of older siblings of child at birth		Wellbeing at school at age seven <sup>e</sup>
Whether mother had depression at child's birth		Attitudes to learning at age seven <sup>e</sup>
Whether mother had long-term illness at child's birth		Enjoyment of hobbies at age seven <sup>e</sup>
Season of birth of child		
Birth weight of child		
How old child was when they were last breastfed (if applicable)		

\*These variables are used in the propensity score matching.

<sup>a</sup>The urban–rural classifications available in the MCS for cohort members in each country of the UK are derived slightly differently to one another. However, for the purpose of this analysis these are combined into a common classification that distinguishes between wards as urban, rural or mixed. However, in Scotland there is no equivalent category for mixed wards so this should be noted when interpreting the results.

<sup>b</sup>The Index of Multiple Deprivation is calculated for each country.

<sup>c</sup>More information about the HLE and how it is calculated from the MCS can be found in de la Rochebrochard (2012), and is based on the work of Melhuish et al. (2008).

<sup>d</sup>More information about these assessments can be found in Hansen et al. (2010). All measures are standardised and age-adjusted.

<sup>e</sup>These measures of subjective wellbeing are based on the composite responses to a series of questions in a self-completion questionnaire for cohort members of the MCS when they were aged approximately seven years. NB: These measures are not age-adjusted.

Table 2 attempts to illustrate this process based on children in the MCS for whom the necessary data were available, including valid outcome measures, in this case word reading ability scores at age seven. The results using matching estimator 1 are based on matching children with the most similar probability of being from Wales (or from Scotland), i.e. their nearest neighbour. A calliper is used to limit how similar, or not,

Table 2. Matched samples (word reading ability at age seven).

Country of sample	Matching estimator 1			Matching estimator 2	
	Unmatched	Matched	Final sample	Unmatched	Matched
'Control'					
England	0	7615	1601	0	7615
'Intervention'					
Wales	63	1601	1601	222	1442
Total	63	9216		222	9057
'Control'					
England	0	7074	1035	0	7074
'Intervention'					
Scotland	43	1035	1035	236	842
Total	43	8109		236	7916

Matching estimator 1: Nearest neighbour matching, calliper 0.001, no replacement.

Matching estimator 2: Radius matching, calliper 0.0001.

the probabilities can be before they can be matched. So, for example, Table 2 shows that 63 children in Wales were unable to be matched because they were not similar enough to a comparable child in England. However, it also shows that only 1601 children from England were used in the final matched comparison, since this is the maximum number of children in Wales they could be matched to. In other words, those children in England who were the least like any of the children in Wales are also not included in the comparison.

The second matching estimator is slightly different. This has a smaller calliper, which means that the threshold for being matched is much stricter, hence there are more children from the two 'control' groups that are unmatched; 222 in Wales and 236 in Scotland. However, in contrast to the nearest neighbour technique of the other matching estimator the final selection for comparison is based on radius matching. This means that every child in the intervention group could be matched to one or more children in the control group,<sup>9</sup> but within the constraints set by the calliper. Although this means that more children in England are included in the final matched comparison, there are still some who were not included because they were too different to other children in Wales (or Scotland).

For the second set of analyses, which focus more on regional comparisons, we revert to traditional general regression models to derive standardised estimates that indicate the strength of association with the outcomes measures. The results from a series of models are presented to help identify what mediating influence, if any, these additional geographical characteristics have on the country 'effects'.

A final important methodological remark relates to the sampling design of the MCS. As previously noted, the way in which children were selected to be part of the MCS was slightly more complicated than other well-known birth cohort studies. Furthermore, like most other longitudinal studies the MCS suffers from attrition over time. This means that most analyses of the MCS have to take in to account the resulting selection biases that exist in the MCS. Fortunately the effects of these biases can be significantly reduced by the use of sampling and attrition weights

especially prepared for the MCS (Plewis 2007). For the second set of analyses all the regression models use these weighted estimates. This ensures that the results are based on a nationally representative sample. However, in the first set of analyses, using propensity score matching, it is not possible to include these sampling and attrition weights.<sup>10</sup> Indeed, we have already seen that matching can only occur where children from one country have similar characteristics to children in another country, leaving a number of children in each country unmatched. Clearly this means that these comparisons are not based on nationally representative samples, and any interpretation of the results must recognise this; but in presenting results from two methodological approaches we are also able to consider what impact this sampling issue has on the results.

### **Country analysis**

Table 3 presents the results of using matched samples of children between Wales and England and between Scotland and England. Here we focus on three assessments of literacy skills as children grow up. The average scores without matching are also provided to demonstrate the impact of using matched samples of children in the comparison. For example, without matching there are significant differences in the vocabulary skills of children in the three countries at age three. On average, children in England have the lowest scores and children in Scotland have the highest scores. But using matched comparisons the difference between the samples of children from Wales and England disappears, i.e. children sharing comparable characteristics in England and Wales are likely to get similar scores. Similarly the ‘gap’ between Scotland and England narrows but remains significant.

However, as the children grow up this pattern begins to change. By age five children in England appear to have improved, on average, their literacy skills at a faster rate than comparable children in Wales and Scotland. So by the time the children have entered compulsory education the ‘gap’ between Scotland and England has nearly disappeared and a significant but relatively small ‘gap’ between Wales and England has appeared.

By age seven this trend appears to continue; children in England appear to continue to make greater progress in their literacy skills than comparable children in Wales and Scotland. Of particular significance is the growing ‘gap’ in literacy skills between Wales and England; however, a difference of 2.5 in word reading ability at age seven (the difference in average scores using matching estimator 1) is only the equivalent of around a month in vocabulary development. Although we do not directly compare the cognitive development of children in Wales with comparable children in Scotland the results presented in Table 3 suggest any differences already exist by age three years and do not appear to worsen (or improve) as children enter the primary phase of their education.

Crucially, these country-by-country trends and patterns in assessments of children’s literacy and language development are not repeated in other, equally important, areas of cognitive development. For example, whilst children in Scotland are, on average, slightly behind comparable children in England in maths ability at age seven, the average score of children in Wales is similar to the average score of equivalent children in England; and in pattern construction, Wales does significantly better than England and Scotland. Here a difference of 1.5 in pattern construction is the equivalent to around three months in the development of spatial visualisation and non-verbal reasoning.

Table 3. Matched comparisons in cognitive development to age seven.

	Wales–England comparison			Scotland–England comparison		
	Wales	England	T-stat	Scotland	England	T-stat
<i>Naming vocabulary at age three</i>						
Without matching	49.98	48.48	5.5	52.69	48.17	12.9
Matched estimator 1	50.04	50.02	0.1	52.80	50.87	4.5
Matched estimator 2	50.25	50.01	0.8	52.67	50.91	4.0
<i>Naming vocabulary at age five</i>						
Without matching	53.55	53.65	-0.4	56.62	53.35	9.1
Matched estimator 1	53.70	55.17	-4.6	56.52	55.99	1.2
Matched estimator 2	53.91	55.16	-4.0	56.39	55.90	1.1
<i>Word reading ability at age seven</i>						
Without matching	47.67	51.11	-13.0	50.85	51.15	-1.0
Matched estimator 1	47.78	50.28	-6.9	50.77	50.82	-0.1
Matched estimator 2	48.22	50.33	-6.2	50.83	51.03	-0.5
<i>Maths ability at age seven</i>						
Without matching	49.87	50.23	-1.4	49.01	50.23	-3.8
Matched estimator 1	49.92	50.65	-2.2	49.03	51.45	-5.8
Matched estimator 2	49.97	50.45	-1.5	48.71	51.75	-6.8
<i>Pattern construction at age seven</i>						
Without matching	54.29	52.73	5.5	53.94	52.63	3.6
Matched estimator 1	54.30	53.02	3.5	53.97	54.28	-0.6
Matched estimator 2	54.35	53.00	3.8	53.76	54.00	-0.5

Matching estimator 1: Nearest neighbour matching, calliper 0.001, no replacement.

Matching estimator 2: Radius matching, calliper 0.0001.

These results suggest that although we observe some significant differences in the cognitive development of children according to which country they grow up in they are often very modest differences and suggest the ‘ranking’ of countries depends on which area of cognitive development is being contrasted.

But following Feinstein (2003) and Blanden and Machin (2010), how do differences in the education systems and policies of different countries within the UK influence different groups of children, particularly those living in poverty? The method of propensity score matching allows us to examine the progress of two equivalently matched groups of children over time. Here we match the children in each country based on their background characteristics (as before) but compare the educational outcomes of particular sub-groups of those matched children based on levels of family income. Specifically we compare the results of children from the ‘richest’ and ‘poorest’ 25% of households in each matched sample.<sup>11</sup> We then compare the assessment scores of these sub-groups of children as they grow up against the same three measures of literacy skills at ages three, five and seven years as we used above. Although these sub-groups of children are not an accurate reflection of the poorest and richest quartiles of the population in each country, they provide an indication of the relative differences in cognitive development of children at contrasting ends of the income spectrum.

Variations in the scores between assessments and the norms used to standardise assessment scores in the MCS make the direct comparison of average scores difficult to interpret. Therefore, Figure 1 presents proportionate differences in the corresponding mean scores.<sup>12</sup>

In the main, the same trends and patterns in literacy development identified above can be seen amongst both the richest and poorest groups of the matched samples of children (Figure 1). Despite different levels of vocabulary development at age three, children in England tend to make the greatest improvement in literacy as they grow up, such that by the age of seven the word reading ability of children in Wales is behind that of England and Scotland, irrespective of whether they are from families with relatively low or high incomes. However, Figure 1 does illustrate some interesting income-related differences in the cognitive development of children in different countries. For example, at age five the poorest children in the matched sample from Wales were not too far behind in their vocabulary skills as the poorest children in England. But by age seven a significant gap between the poorest children in Wales and England has emerged. Similarly in Scotland it is the poorest children that perform less well than their low-income counterparts in England; high-income children in Scotland continue to out-perform high-income children in England throughout the first seven years.

These results seem to make the comparisons between countries even more complex, since the differences in literacy skills vary according to whether we are comparing 'rich' or 'poor' children. As Dex et al. (2008) suggest, such variances could be due to 'differences in early years education provision, in pre-school education, activities in the home or in grandparent influence' (8).<sup>13</sup> But importantly, differences in the literacy development of low-income children in England, Wales and Scotland by age seven suggest that the more 'comprehensive' and perhaps less target-driven systems of Wales and Scotland appear to be associated with greater inequalities in child

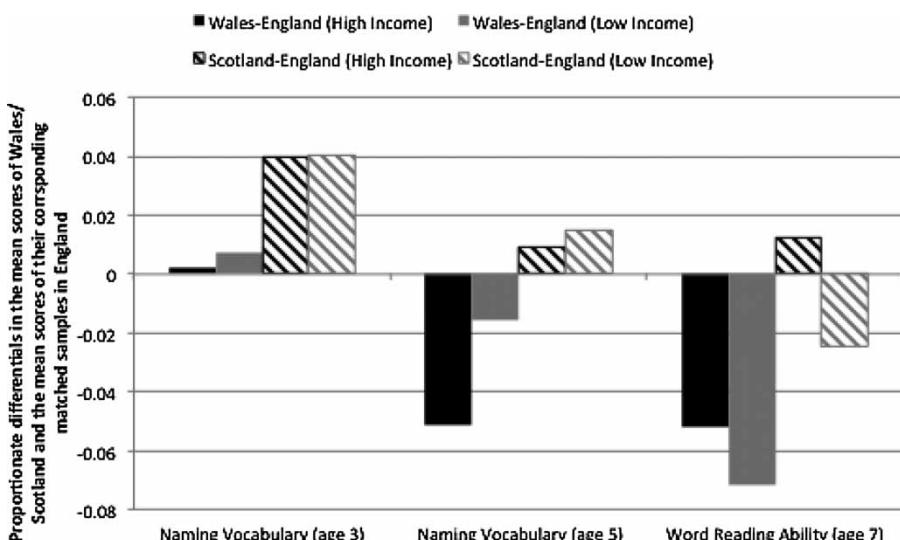


Figure 1. Inequalities in literacy development of children from 'rich' and 'poor' families by country.

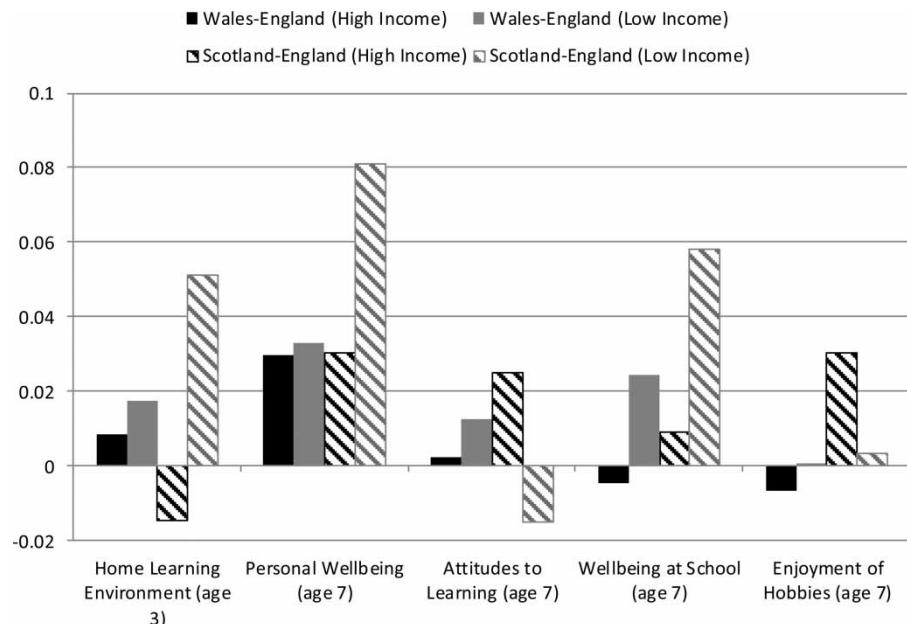


Figure 2. Inequalities in home learning environment and subjective wellbeing of children from 'rich' and 'poor' families by country.

development. This would seem to corroborate similar conclusions by Croxford (2010) for Scotland based on international comparisons made by the OECD (2007):

The Scottish system serves pupils from higher social class backgrounds well and produces relatively high levels of academic attainment and entry to higher education. However, the review [by the OECD] also indicates that the system does not serve pupils from less advantaged backgrounds well. (Croxford 2010, 17)

Despite the relatively low achievement of comparable 'poor' children in Wales and Scotland compared to England such children generally report greater levels of wellbeing than comparable children in England (Figure 2). This is particularly striking given these groups of children generally have, on average, a lower word reading ability than similar children in England. This might suggest that the possible attention on developing literacy skills in England could come at the expense of children's subjective wellbeing. It is also interesting to note that the home learning environment of children from relatively low-income families is greater in Wales and Scotland than it is for their counterparts in England.<sup>14</sup> Perhaps contrary to expectations these results may be worrying to policy-makers in Wales and Scotland. This is because differences in the cognitive development of children, particularly from the 'poorest' families in the sample, appear to exist *despite* 'better' home learning environments and more positive dispositions to school and learning in these countries.

### Regional analysis

In this second part of the analysis we focus our attention on regional comparisons of cognitive development amongst children by the age of seven years. In particular, we

are interested in identifying geographical patterns to these outcomes within England, after attempting to control for differences in the background characteristics of children. In doing this we are also able to see what impact, if any, taking a regional approach to the analysis of outcomes in England has on our interpretation of the educational outcomes in other countries of the UK. This may be particularly important if, for example, we observe a London 'effect' on child development.

As discussed earlier we take a different analytical approach to this analysis, using multivariate regression modelling to compare outcomes. Although the results of this approach may be slightly more complicated to many readers this does have the benefit of allowing us to consider the way in which geographical factors may mediate the influence of other background factors. It is also important to restate that this part of the analysis uses all children in the MCS and utilises sampling and attrition weights. Consequently, although these results may be more difficult to interpret than those presented above, this part of the analysis is not susceptible to concerns about the representative nature of the results for the rest of the UK. We are also able to consider the influence of growing up in Northern Ireland on children's cognitive development.

The statistical models presented in Table 4 all use a number of control variables (the exogenous variables outlined in Table 1) to predict three measures of children's cognitive development at ages three, five and seven. For each measure of cognitive development we present three statistical models: the first only includes an indicator of the country that a child lives in (Models 1, 4 and 7); the second includes an indicator of the region that a child lives in, specifically to help distinguish between the outcomes and living in different regions of England (Models 2, 5 and 8); and the third model introduces two additional geographical variables to the models – an urban–rural indicator and a measure of the level of multiple deprivation for the ward that a child lives in – to see what influence, if any, they have on other predictors of cognitive development.

As Table 4 highlights, there are many significant and important relationships between the socio-economic background of the children – such as their social class background, the educational levels of their parent(s) and levels of household income – and their literacy skills throughout the first seven years of their lives. The analysis also shows how the influence of some of these factors change as the children grow up and enter school. For example, one of the most striking points to note is the influence of ethnicity on literacy skills. Up to the age of seven, children from ethnic minority backgrounds have, on average, significantly lower levels of literacy, but at age seven not only has the 'gap' between White and ethnic minority children closed, we find that children from ethnic minority backgrounds have on average a *higher* word reading ability score than White children all other things being equal. This is particularly the case for Indian, Pakistani and Bangladeshi children. We also see the emerging importance of season of birth on literacy skills once children reach the age of seven and after entering school. However, it is not the purpose nor is there the scope to discuss these results in detail. Indeed, many of the findings have been discussed elsewhere. Instead we are concerned with whether the country or region a child lives in is associated with their literacy development after controlling for these background factors.

The first thing to note from the regression analyses (Models 1, 4 and 7) is that we see exactly the same patterns and trends for the country 'effects' that we saw in the propensity score matching analysis above. Children in Scotland are associated with high vocabulary scores at age three, whereas children in England and Wales share similar scores.

Table 4. Regression of geographical factors on children's literacy development at ages three, five and seven.

	Naming vocabulary at age three			Naming vocabulary at age five			Word reading ability at age seven		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Sample size (unweighted)	11,512	11,512	11,512	11,590	11,590	11,590	11,903	11,903	11,903
Sampling design degrees of freedom	389	389	389	389	389	389	389	389	389
R square	0.238	0.243	0.244	0.251	0.254	0.257	0.166	0.170	0.171
Constant	30.73	29.60	30.02	37.36	39.34	42.65	38.58	37.55	38.03
Sex of child									
Female	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Male	-2.54***	-2.54***	-2.56***	-0.54*	-0.54*	-0.56*	-1.52***	-1.51***	-1.52***
Household structure at birth									
Two parents/carers	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
One parent/carer	-0.35	-0.34	-0.30	0.05	0.10	0.18	-0.21	-0.30	-0.29
Highest social class at birth									
Professional or managerial	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Intermediate	-0.62	-0.58	-0.55	-1.16**	-1.12**	-1.02**	-1.02**	-1.00***	-1.01**
Small employer or self-employed	-0.88	-0.96*	-0.87	-1.43**	-1.51**	-1.46**	-2.03***	-1.99***	-1.94***
Low supervisory or technical	-2.32***	-2.33***	-2.20***	-2.53***	-2.56***	-2.39***	-2.74***	-2.71***	-2.65**
Semi-routine or routine	-2.40***	-2.42***	-2.26***	-2.26***	-2.25***	-2.00***	-2.68***	-2.63***	-2.58***
Not applicable or unemployed	-3.01***	-3.03***	-2.86***	-2.35**	-2.30**	-2.05**	-3.32***	-3.34***	-3.27***
Ethnicity of child									
White	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mixed	-1.58*	-1.20	-1.08	-1.40*	-1.11	-0.83	1.36	1.07	1.08
Indian	-7.75***	-7.15***	-7.04***	-5.39***	-4.89***	-4.62***	4.21***	3.78***	3.76***
Pakistani or Bangladeshi	-11.01***	-10.66***	-10.51***	-10.09***	-9.79***	-9.44***	4.22***	4.12***	4.14***
Black or Black British	-5.48***	-4.60***	-4.39***	-6.12***	-5.24***	-4.85***	2.79**	1.65	1.73

	-9.37***	-8.62***	-8.60***	-9.20***	-8.36***	-8.21***	2.87**	1.89	1.91
Other ethnic group									
Not known	1.74	1.52	1.60	0.27	0.42	0.49	1.01	1.18	1.17
Highest NVQ level at birth									
None or overseas	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NVQ Level 1–3	2.21***	2.13***	2.08***	2.79***	2.74***	2.58***	0.90*	0.93*	0.94*
NVQ Level 4–5	3.77***	3.59***	3.48***	4.79***	4.73***	4.40***	2.35***	2.35***	2.32***
Age of mother at birth									
13–20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21–25	1.23**	1.16*	1.14*	1.02*	1.04*	0.99*	0.99*	0.97*	0.97*
26–30	2.61***	2.57***	2.49***	2.83***	2.86***	2.71***	2.04***	2.03***	2.00***
31–35	2.89***	2.84***	2.72***	3.05***	3.10***	2.86***	2.23***	2.16***	2.14***
36+	3.05***	3.01***	2.90***	3.31***	3.38***	3.11***	2.69***	2.58***	2.53***
Number of older siblings at birth									
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	-2.65***	-2.68***	-2.67***	-2.12***	-2.16***	-2.16***	-1.19***	-1.16***	-1.18***
2	-3.58***	-3.56***	-3.51***	-3.40***	-3.42***	-3.35***	-2.87***	-2.86***	-2.87***
3+	-5.22***	-5.28***	-5.20***	-5.20***	-5.25***	-5.10***	-4.30***	-4.29***	-4.27***
Season of birth									
Autumn 2000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Winter 2000/2001	0.42	0.43	0.45	-0.63*	-0.62*	-0.59*	-1.20***	-1.18***	-1.18***
Spring 2001	0.77*	0.79*	0.78*	-0.12	-0.09	-0.13	-1.07**	-1.07**	-1.07**
Summer 2001	0.50	0.48	0.48	0.71*	0.71*	0.73*	-2.46***	-2.46***	-2.46***
Autumn/Winter 2001/2002	1.23*	1.24*	1.25*	0.17	0.19	0.20	-3.20***	-3.19***	-3.20***
Birth weight (kilos)	1.53***	1.54***	1.54***	0.75***	0.77***	0.76***	1.11***	1.09***	1.10***
Weekly predicted income (£) (divided by 100)	0.12***	0.14***	0.12***	0.20***	0.20***	0.18***	0.27***	0.26***	0.25***

(Continued)

Table 4. (Continued).

	Naming vocabulary at age three			Naming vocabulary at age five			Word reading ability at age seven		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Home learning environment (composite index)	0.19***	0.20***	0.19***	0.18***	0.18***	0.18***	0.14***	0.14***	0.13***
Country at age seven									
England	0.00			0.00			0.00		
Wales	0.17			-1.18***			-2.22***		
Scotland	1.95***			0.68			0.29		
Northern Ireland	1.49**			1.10*			-1.62***		
Government Office Region at age seven									
London	0.00	0.00		0.00	0.00		0.00	0.00	
North East	1.54	1.69		0.81	0.89		-1.36	-1.24	
North West	0.69	0.70		1.70*	1.73*		-1.85**	-1.68**	
Yorkshire & Humber	0.62	0.70		0.80	0.84		-2.20**	-2.13**	
East Midlands	0.56	0.53		1.87*	1.49		-2.47**	-2.36**	
West Midlands	1.54*	1.56*		1.00	0.95		-1.73*	-1.71*	
East of England	1.40*	1.26*		2.22*	1.79		-1.84*	-1.76*	
South East	0.64	0.49		1.03	0.73		-1.96**	-1.94**	
South West	3.25***	3.06**		1.82*	1.35		-0.86	-0.65	
Wales	1.28*	1.24*		0.14	-0.21		-3.88***	-3.76***	
Scotland	3.05***	3.08***		1.98*	1.67*		-1.37*	-1.37*	
Northern Ireland	2.61***	2.71***		2.41**	1.78*		-3.29***	-3.09***	

Urban–rural indicator (at nine months)			
Urban	0.00	0.00	0.00
Mixed	0.40	0.68	-1.13*
Rural	-0.48	1.37**	-0.31
Index of Multiple Deprivation (deciles by country) (at nine months)	0–<20 (most disadvantaged)	0.00	0.00
20–<40	0.06	0.62	0.20
40–<60	0.81*	0.91*	0.20
60–<80	0.75	1.35**	0.31
80–<100 (least disadvantaged)	0.95	1.63**	0.73

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  All models are based on weighted estimates.

As they grow up children in Wales fall significantly behind children in England in their literacy abilities, whilst children in England appear to ‘catch up’ with children in Scotland. Furthermore, the estimates presented in Table 4 are very similar to the ‘gaps’ in average scores of the matched comparison groups in Table 3.

However, in contrast to the earlier analysis, Table 4 is able to show what influence growing up in Northern Ireland seems to have on cognitive development. These results are particularly noteworthy; despite being associated with significantly higher vocabulary scores at ages three and five, children in Northern Ireland have, on average, a significantly low word reading ability by age seven compared to children in England. Again, this illustrates the relatively greater improvement of children in England as they grow up. Nevertheless, policy-makers in Northern Ireland may be particularly concerned about what happens to children between the ages of five and seven.

Another key benefit of these regression models is that we are able to see the relative influence of which country the children are from compared to other key predictors of cognitive development. It is notable that the significant estimates for the country indicators are generally smaller than the main socio-economic predictors, such as social class, ethnicity and educational levels of parent(s). Indeed, they are generally lower than other predictors of cognitive development, such as the number of older siblings and the season of birth.

We also see from Table 4 that, despite previously highlighting country-level differences in the home learning environment, this remains a significant predictor of cognitive development at all stages of a child’s early development at the level of individual households. This apparent contradiction could be due to biases in the matched samples used earlier or it could suggest there are differences in the degree of influence that the home learning environment has on children’s cognitive development in different countries. Only with more detailed analysis, perhaps using multi-level statistical models, can this be explored further.

We now move on to consider whether other geographical factors mediate the influence of these main predictors of cognitive development, including the country the children are from (Models 2, 3, 5, 6, 8 and 9 in Table 4).

The first thing to note is that there appears to be significant variation in the literacy development of children by region of England. Most notably, children in London at age three are associated with low vocabulary scores, particularly when compared to children in Wales, Scotland and Northern Ireland (Model 2). We also see a significant positive association in these scores for children from the south west of England, who have, on average, the highest vocabulary scores at age three of all children in the UK. At age five these patterns shift slightly (Model 5). Although children in London appear to make greater improvement in their vocabulary ability than children in Wales and Scotland there are other regions of England where children would seem to make even greater progress (the North West, the East Midlands and the East of England). However, by age seven the picture is much clearer; children in London are associated with significantly higher word reading ability scores, all other things being equal (Model 8) compared with children from most other regions/countries of the UK. Importantly, if we compare the estimates for Wales and Scotland with the estimates for other regions in England we actually see that much of the ‘improvement’ in literacy skills in England identified above, could be attributed to children in London.

Next we consider the mediating influence of whether the children live in urban or rural areas and the level of multiple deprivation of their local neighbourhood (Models 3, 6 and 9). Here we see a negative association between neighbourhood disadvantage and

cognitive ability, particularly at age five. The relationship between urban–rural geography and cognitive development are less clear, with urban areas associated with low vocabulary scores at age five but high word reading ability scores at age seven. However, we do see that, in general, these two geographical factors have a small but important mediating effect on the influence of the region and country that the children live in. These geographical factors also appear to mediate, to some extent, the influence of other family-level socio-economic factors, most notably the social class background of the children; but equally it is also worth noting that some factors are unaffected by the inclusion of geographical variables, such as the sex of the child, a child's birth weight, the number of older siblings, the child's season of birth and their home learning environment.

Finally, it is useful to return to the influence of ethnicity on this discussion, particularly because of the strong connection between some of the geographical indicators and non-White ethnic groups. As noted above ‘improvement’ in literacy skills of ethnic minority children, relative to White children, is quite significant. Importantly, here too we see the mediating influence of the geographical factors on ethnicity. However, what is not clear is whether the ‘cause’ of literacy improvement is due to the presence of non-White children in an area or the influence of the education system (and its associated policies and initiatives) in those same areas. This would seem to be of particular relevance to the apparent improvement in literacy development for children in London. However, as far as it is possible to examine such issues in this paper it does appear that both ethnicity and geography have some role in determining cognitive outcomes, and that in combination these probably go some way in helping to understand the relative under-performance of children in other areas with limited ethnic minority populations by age seven, such as in Wales and Northern Ireland.

## Conclusion

In this paper we have attempted to demonstrate the potential of ‘home international’ comparisons as an important tool for policy analysis, embodying some of the necessary features of ‘natural experiments’. However, we have also shown that simple, and often crude, ‘home international’ comparisons, often undertaken within the media and by politicians, can lead to very crude, and potentially misleading, ‘evaluations’ of alternative policy approaches in different countries of the UK. It is very clear from this discussion that careful and detailed ‘home international’ comparisons are warranted. But these are often dependent on the quality of available data.

This analysis has been based entirely on the Millennium Cohort Study, which has helped us to illustrate some of the complexities of ‘home international’ analyses, especially in relation to different geographical scales. Although we have only been able to consider the cognitive development of children up to the age of seven years, we have still been able to provide new and important insights into the relationships between divergent educational policies and cognitive development.

There would seem to be three main conclusions from this analysis. The first is that there is no single national ‘success story’, suggesting one education system in the UK is ‘better’ than another. For a start, any ranking of countries depends on which measure of cognitive development is being considered. Also, where there are significant differences in cognitive abilities by country these are relatively small compared to the influence of other conditions that children are born in to. Furthermore, it is not entirely clear whether the apparent ‘benefit’ of living in England on literacy at age seven can be

attributed to its distinctive national education policies or simply a reflection of processes and influences on cognitive development that exist at a regional or local level. For example, differences found between England and the rest of the UK may in part be due to the significant improvement in cognitive development of children living in London. In turn, however, the differences between London and the rest of England may be due to the significant improvement in cognitive development of ethnic minority children who are concentrated in the capital.

This leads on to the second main conclusion, that there is still a considerable need for further ‘home international’ analyses that utilise genuinely comparative and longitudinal data. The MCS provides an excellent example of this, and as the MCS cohort ages and further sweeps of data are collected this is going to become increasingly more valuable in undertaking ‘home international’ comparisons and for conducting natural experiments of particular policy initiatives in the UK.

Despite detailed and complex datasets, such as the MCS, often requiring detailed and complex techniques for analysis we have also attempted to demonstrate, through the use of propensity score matching, that these comparisons can often be made very clearly and efficiently. This is important, as it can be the same reason why policy-makers and the media often rely on simple and crude comparisons, an issue that Goldthorpe (2012) also acknowledges when referring to the ‘media hysteresis’ surrounding social mobility.

This in turn leads on to our final conclusion, that much greater consideration should be given to the desire to use simple comparisons in education for the immediate purposes of policy evaluation, formation and borrowing. Comparative education researchers have, for a long time, raised concerns about this (see Crossley and Watson 2009) and queried the basis of recent international comparisons (see Sturman 2012). But despite this, policy-makers appear to be increasingly influenced by simple comparisons of a small number of, often narrowly defined, educational outcomes. Nóvoa and Yariv-Mashal (2003) and Ozga (2012) have gone further, suggesting that comparative studies of education are now essentially political tools and a form of policy technology that undermines the intellectual scholarship that they once had.

We would suggest that in order to counter these policy moves education policy analysis, and comparative studies in particular, must (a) make more use of ‘home international’ analyses within the UK, since they can demonstrate the complexities of making such comparisons despite comparing national education systems with so much in common, and (b) take issues of geography, scale and context more seriously when interpreting the influence of ‘national’ education systems and policies on educational outcomes.

### Acknowledgements

The authors are grateful to The Centre for Longitudinal Studies (CLS), Institute of Education for the use of these data, to the UK Data Archive and Economic and Social Data Service for making them available and to Heather Joshi at the CLS for her very helpful comments and advice. However, they bear no responsibility for the analysis or interpretation of these data. This work was supported by the Economic and Social Research Council [grant number RES-070-27-0070] and by the Wales Institute of Social and Economic Research, Data and Methods (WISERD).

### Notes

1. Scotland’s use of a different qualifications system makes comparison with other parts of the UK complicated. The complexities of the political situation in Northern Ireland also militate against ready comparison.

2. The most recent fifth sweep is currently being undertaken at the time of writing, so data relating to when the children are aged 11 years old were not available for this current analysis.
3. Despite country-level boosts to the original sample size, the limited size of the cohort population in the three smaller countries of the UK is still a limitation on particular kinds of analyses, particularly those that are dependent on comparing particular sub-groups of the MCS cohort as they get older due to attrition of the original sample. However, for national comparisons, as undertaken here, the cohort sizes are relatively large enough to still make meaningful and insightful comparisons.
4. The Child Poverty Index is defined as ‘the percentage of children under 16 in an electoral ward living in families that were, in 1998, receiving at least one of the following benefits: Income Support; Jobseekers Allowance; Family Credit; or Disability Working Allowance’ (Plewis 2007, 10). The measure and definition is common across all countries of the UK, although differences in the socio-economic demographics of each country meant that in some countries there were more wards and children to choose from [see Plewis (2007) for more details].
5. In England wards were also selected because they had a high concentration of ethnic minority families living there. This was to ensure that there was an over-representation of ethnic minority children in the England MCS sample. However, it is important to note that this criterion was not used in the selection of children in Scotland, Wales and Northern Ireland.
6. We do not match a sub-group of children in the MCS living in England with the MCS children living in Northern Ireland largely because of the smaller MCS cohort size in Northern Ireland and because of the smaller emphasis on the impact of devolution in Northern Ireland. However, MCS children in Northern Ireland are included in the second set of analyses.
7. Given the different number of MCS cohort children in each country it is inevitable that not all children from one country (typically in England) will be matched to a child in another country. Hence the selection of children to be included or excluded from the control group and subsequent comparison is important.
8. The two matching estimators we report results for are: (i) Nearest Neighbour Matching, calliper 0.001, no replacement; and (ii) Radius Matching, calliper 0.0001.
9. If a child in the intervention group is matched to more than one child in the control group the average outcome scores of those children are used for comparison.
10. This is particularly important when comparing children from Wales, since the over-representation of children from areas with high levels of child poverty was greater here than in other countries (see Plewis 2007).
11. This is a derived measure in the MCS that predicts weekly net family income that has been adjusted for family size. The use of household income here is independent to the process of matching children, as it was not used in the Propensity Score Matching. The quartiles have been defined on the unweighted results for the corresponding ‘intervention’ country – i.e. for the sample of children in Wales and Scotland – and then applied to the corresponding matched sample from England. Therefore, these sub-groups are not strictly the richest 25% and poorest 25% of each country’s population since the matched samples are not representative of their respective countries.
12. These differences are calculated proportionately – i.e.  $(a-b)/(a+b)$  – and are not the raw score differences.
13. Dex et al. (2008) also suggest that there may have also been selective attrition bias in Scotland compared to the rest of the MCS sample at age three years.
14. It is only compared to children from the ‘richest’ families in the Scotland matched sample that children in England appear to have a greater home learning environment.

### Notes on contributors

**Chris Taylor** is a Professor in the School of Social Science at Cardiff University and a research director in the Wales Institute for Social and Economic Research, Data and Methods (WISERD). He has been researching the geography of education for many years, particularly in relation to school admissions, education and neighbourhoods, education markets and participation in higher education. A central focus of all his research is the impact of education policy on issues of social justice. He was recently awarded an ESRC Mid-Career Fellowship to develop skills in the spatial analysis of the Millennium Cohort Study. He currently leads the official

three-year independent evaluation of the Foundation Phase in Wales, the early years flagship policy of the Welsh Government.

**Gareth Rees** is Professor and Director of the Wales Institute for Social and Economic Research, Data and Methods (WISERD). He has held a number of visiting positions, notably at the University of New South Wales in Sydney and at the University of British Columbia in Vancouver. Key research interests include the social impacts of higher education; lifelong learning and the Learning Society; vocational education and training and economic development; the governance of education policy; and research-capacity building in the social sciences. In addition to his academic research, he has been an adviser to a number of governmental bodies, including the OECD, the European Commission, UK government departments and the Welsh Government. He is an Academician of the UK Academy of Social Sciences and a Fellow of the Learned Society of Wales.

**Rhys Davies** is a senior researcher in the Wales Institute for Social and Economic Research, Data and Methods (WISERD) and has extensive experience of analysis of large-scale datasets related to employment and the labour market. This has been based on large-scale primary data collection, as well as conducting analysis on secondary and administrative data sources. Key research interests relate to occupational health and safety in relation to the economic environment and structural change; graduate labour market and destinations of leavers from HE and FE; the development and implementation of occupation and social classifications; gender differentials in labour market outcomes; and survey design and data linking.

## References

- Atkinson, R., and K. Kintrea. 2001. "Disentangling Area Effects: Evidence from Deprived and Non-deprived Neighbourhoods." *Urban Studies* 38 (12): 2277–2298.
- Ball, S. J. 2008. *The Education Debate: Policy and Politics in the Twenty-first Century*. Bristol: The Policy Press.
- Ball, S. J., M. Maguire, and A. Braun. 2012. *How Schools Do Policy: Policy Enactments in Secondary Schools*. Abingdon: Routledge.
- Blanden, J., and J. Machin. 2010. "Intergenerational Inequality in Early Years Assessments." *Children of the 21<sup>st</sup> Century: the First Five Years*, edited by K. Hansen, 153–168. Bristol: Policy Press.
- Burgess, S., D. Wilson, and J. Worth. 2010. *A natural experiment in school accountability: the impact of school performance information on pupil progress and sorting*, CMPO Working Paper No. 10/246. Bristol: Centre for Market and Public Organisation, University of Bristol.
- Butler, T., and C. Hamnett. 2007. "The Geography of Education: Introduction." *Urban Studies* 44 (7): 1161–1174.
- Byrne, D., B. Williamson, and E. Fletcher. 1975. *The Poverty of Education: A Study in the Politics of Opportunity*. London: Martin Robertson.
- Cook, C. 2012. London schools rise to the challenge. *Financial Times*, August 5<sup>th</sup> 2012.
- Crossley, M., P. Broadfoot, and M. Schweisfurth, eds. 2007. *Changing Educational Contexts, Issues and Identities: 40 Years of Comparative Education*. London: Routledge.
- Crossley, M., and K. Watson. 2009. "Comparative and International Education: Policy Transfer, Context Sensitivity and Professional Development." *Oxford Review of Education* 35 (5): 633–649.
- Croxford, L. 2010. "Tensions between the Equity and Efficiency of Schooling: The Case of Scotland." *Education Inquiry* 1 (1): 5–20.
- Dex, S., et al. 2008. *Millennium Cohort Study Exploration of Some Distinctive Results for Scotland*. Edinburgh: Scottish Government Social Research.
- Dunning, T. 2012. *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge: CUP.
- Feinstein, L. 2003. "Inequality in the Early Cognitive Development of British Children in the 1970 Cohort." *Economica* 70: 73–97.
- Goldstein, H. 2004. "International Comparisons of Student Attainment: Some Issues Arising from the PISA Study, Assessment in Education: Principles." *Policy & Practice* 11 (3): 319–330.

- Goldstein, H. 2008. "Comment peut-on utiliser les études comparatives internationales pour doter les politiques éducatives d'informations fiables?" *Revue Française de Pédagogie* 164: 69–76.
- Goldthorpe, J. 2012. *Understanding – and Misunderstanding – Social Mobility in Britain: The Entry of the Economists, the Confusion of the Politicians and the Limits of Educational Policy, 1/2012*. Oxford: Department of Social Policy and Intervention, University of Oxford.
- Gorard, S. 2000. "Underachievement is Still an Ugly Word: Reconsidering the Relative Effectiveness of Schools in England and Wales." *Journal of Education Policy* 15 (5): 559–573.
- Greer, S., ed. 2009. *Devolution and Social Citizenship in the UK*. Bristol: The Policy Press.
- Hansen, K., et al. 2010. *Millennium Cohort Study – First, Second, Third and Fourth Surveys: A Guide to the Datasets*. London: Centre for Longitudinal Studies, Institute of Education, University of London.
- Jones, M., D. Blackaby, and P. Murphy. 2010. *An Investigation into Regional Differences in Child Health and Cognitive Function, Report for the Welsh Assembly Government Economic Research Unit*. Cardiff: Welsh Assembly Government.
- Joshi, H., B. Dodgeon, and G. Hughes. 2008. "A Profile of Population Change in Rural England." *Centre for Longitudinal Studies Working Paper 2008/4*. London: Institute of Education, University of London.
- Ketende, S., J. McDonald, and H. Joshi. 2010. "Neighbourhoods and Residential Mobility." In *Children of the 21<sup>st</sup> Century: The First Five Years*, edited by K. Hansen, et al. 115–130. Bristol: Policy Press.
- Melhuish, E. C., K. Sylva, P. Sammons, I. Siraj-Blatchford, B. Taggart, and M. Phan. 2008. "Effect of Home Learning Environment and Preschool Centre Experience Upon Literacy and Numeracy Development in Early Primary School." *Journal of Social Issues* 64 (1): 157–174.
- Nóvoa, A., and T. Yariv-Mashall. 2003. "Comparative Research in Education: A Mode of Governance or a Historical Journey?" *Comparative Education* 39 (4): 423–438.
- Organisation for Economic Cooperation and Development (OECD). 2007. *Quality and Equity of Schooling in Scotland*. Paris: Organisation for Economic Cooperation and Development.
- Ozga, J. 2012. "Comparison as a Governing Technology: The Case of PISA." *Research Intelligence*, 119, 18-1, C. (2011) *Fabricating Quality in Education: Data and Governance in Europe*, London: Routledge.
- Ozga, J., P. Dahler-Larsen, H. Simola, and C. Segerholm. 2011. *Fabricating Quality in Education: Data and Governance in Europe*. London: Routledge.
- Plewis, I., ed. 2007. *The Millennium Cohort Study: Technical Report on Sampling 4th ed.* London: Centre for Longitudinal Studies, Institute of Education, University of London.
- Power, S., G. Rees, and C. Taylor. 2005. "New Labour and Educational Disadvantage: The Limits of Area-based Initiatives." *London Review of Education* 3 (2): 101–116.
- Raffe, D. 2006. "Devolution and Divergence in Education Policy." *Devolution in Practice 2006*, edited by J. Adams and K. Schmucker, 52–69. London: IPPR.
- Raffe, D., K. Brannen, L. Croxford, and C. Martin. 1999. "Comparing England, Scotland, Wales and Northern Ireland: The Case for 'Home Internationals' in Comparative Research." *Comparative Education* 35 (1): 9–25.
- Rees, G. 2007. "The Impacts of Parliamentary Devolution on Education Policy in Wales." *Welsh Journal of Education* 14 (1): 8–20.
- Rees, G. 2012. "A Crisis in Welsh Education? New Approaches in Harsh Times." *Education Review* 24 (2): 40–49.
- de la Rochebrochard, E. 2012. *Millennium Cohort Study Data Note 1: The Home Learning Environment as Measured at Age 3*. London: Centre for Longitudinal Studies, Institute of Education, University of London.
- Rosenbaum, P., and D. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41–55.
- Sturman, L. 2012. "Making Best Use of International Comparison Data." *Research Intelligence* 119: 16–17.
- Taylor, C. 2009. "Towards a Geography of Education." *Oxford Review of Education* 35 (5): 651–669.
- Taylor, C., and H. Joshi. forthcoming.

- Thiem, C. H. 2009. "Thinking through Education: The Geographies of Contemporary Educational Restructuring." *Progress in Human Geography* 33 (2): 154–173.
- Warf, B., and S. Arias, eds. 2008. *The Spatial Turn: Interdisciplinary Perspectives*. London: Routledge.
- Whitty, G., and J. Anders. forthcoming. "Narrowing the Achievement Gap? Policy and Practice in England 1997–2010." In *Closing the Achievement Gap: International Perspectives*, edited by J. V. Clark. NSF/Springer.

## Appendix 1

Table A1. Balancing for nearest neighbour matching using calliper of 0.0001, no replacement – England matched to Wales (word reading ability at age seven).

	Sample	Mean			t-test		
		Wales	England	% bias	% reduction in bias	t	p > t
Word reading ability (age seven)	Unmatched	47.67	51.11	-33.9		-13.0	0.00
	Matched	47.78	50.28	-24.6	27.4	-6.9	0.00
Gender							
Male	Unmatched	0.51	0.50				
	Matched	0.51	0.50				
Female	Unmatched	0.49	0.50	-2.8		-1.0	0.31
	Matched	0.49	0.50	-1.4	50.2	-0.4	0.70
Household structure							
One parent/carer	Unmatched	0.82	0.88				
	Matched	0.83	0.83				
Two parents/carers	Unmatched	0.18	0.12	17.7		6.9	0.00
	Matched	0.17	0.17	0.2	99	0.1	0.96
Highest NSSEC in household							
Managerial and professional	Unmatched	0.39	0.45				
	Matched	0.41	0.40				
Intermediate	Unmatched	0.13	0.13	0		0.0	1.00
	Matched	0.13	0.13	-0.4	-2302	-0.1	0.92
Small employer and self-employed	Unmatched	0.06	0.07	-2.6		-0.9	0.35
	Matched	0.06	0.07	-1.8	29.7	-0.5	0.61
Lower supervisory and technical	Unmatched	0.11	0.09	9.3		3.6	0.00
	Matched	0.11	0.11	0.2	97.8	0.1	0.96
Semi-routine and routine	Unmatched	0.25	0.22	7.8		2.9	0.00
	Matched	0.24	0.24	0.9	88.7	0.3	0.80
Not applicable	Unmatched	0.05	0.05	1.5		0.6	0.59
	Matched	0.05	0.05	0	100	0.0	1.00
Ethnicity of child							
White	Unmatched	0.97	0.78				
	Matched	0.97	0.97				
Mixed	Unmatched	0.02	0.04	-12.9		-4.3	0.00
	Matched	0.02	0.02	0.4	96.9	0.1	0.89

(Continued)

Table A1. (Continued).

	Sample	Mean			t-test		
		Wales	England	% bias	% reduction in bias	t	p > t
Indian	Unmatched	0.00	0.04	-25.5		-7.5	0.00
	Matched	0.00	0.00	-0.5	98.2	-0.4	0.71
Pakistani or Bangladeshi	Unmatched	0.01	0.09	-39.2		-11.7	0.00
	Matched	0.01	0.01	0.6	98.5	0.5	0.65
Black or Black British	Unmatched	0.00	0.04	-26.8		-7.9	0.00
	Matched	0.00	0.00	-0.9	96.8	-0.6	0.53
Other ethnic group (incl Chinese)	Unmatched	0.00	0.02	-14		-4.3	0.00
	Matched	0.00	0.00	0	100	0.0	1.00
Highest NVQ in household							
None or overseas	Unmatched	0.09	0.11				
	Matched	0.08	0.09				
NVQ Level 1	Unmatched	0.07	0.06	6.3		2.4	0.02
	Matched	0.07	0.07	-0.5	91.9	-0.1	0.89
NVQ Level 2	Unmatched	0.26	0.25	3.9		1.5	0.15
	Matched	0.26	0.26	1.4	63.2	0.4	0.69
NVQ Level 3	Unmatched	0.17	0.15	4.7		1.8	0.08
	Matched	0.17	0.17	-2.4	49.6	-0.7	0.51
NVQ Level 4	Unmatched	0.36	0.37	-1.6		-0.6	0.56
	Matched	0.37	0.37	0.1	91.9	0.0	0.97
NVQ Level 5	Unmatched	0.05	0.07	-9.6		-3.4	0.00
	Matched	0.05	0.05	1.6	83.4	0.5	0.62
Mothers age at birth							
13–20	Unmatched	-0.17	0.04				
	Matched	-0.16	-0.16				
21–25	Unmatched	0.20	0.17	5.7		2.2	0.03
	Matched	0.19	0.19	-0.8	86	-0.2	0.82
26–30	Unmatched	0.30	0.31	-2.7		-1.0	0.32
	Matched	0.30	0.31	-1.2	55.1	-0.3	0.73
31–35	Unmatched	0.26	0.29	-7.6		-2.8	0.01
	Matched	0.26	0.25	2.9	61.4	0.9	0.40
36+	Unmatched	0.11	0.13	-8.3		-3.0	0.00
	Matched	0.11	0.12	-2.9	65.3	-0.8	0.41
Number of older siblings							
0	Unmatched	0.43	0.41				
	Matched	0.42	0.42				
1	Unmatched	0.36	0.36	1		0.4	0.71
	Matched	0.37	0.38	-2.1	-105	-0.6	0.56
2	Unmatched	0.15	0.15	-1.1		-0.4	0.69
	Matched	0.15	0.14	1.9	-77.4	0.6	0.58
3	Unmatched	0.06	0.08	-5.1		-1.8	0.07
	Matched	0.06	0.06	0.7	85.6	0.2	0.83
Urban/rural							
Urban	Unmatched	0.75	0.87				
	Matched	0.77	0.78				

(Continued)

Table A1. (Continued).

	Sample	Mean		% bias	% reduction in bias	t-test	
		Wales	England			t	p > t
Mixed	Unmatched	0.14	0.07	24.6	95.8	10.2	0.00
	Matched	0.13	0.12	1		0.3	0.79
Rural	Unmatched	0.11	0.06	17.3	93.5	7.0	0.00
	Matched	0.10	0.10	1.1		0.3	0.77
Mother's depression							
No	Unmatched	0.73	0.77				
	Matched	0.74	0.73				
Yes	Unmatched	0.27	0.23	9.5	95.4	3.6	0.00
	Matched	0.26	0.27	-0.4		-0.1	0.91
Mother's illness							
No	Unmatched	0.77	0.78				
	Matched	0.77	0.76				
Yes	Unmatched	0.23	0.22	2.1	-87.4	0.8	0.44
	Matched	0.23	0.24	-3.9		-1.1	0.28
Season of birth							
Autumn 2000	Unmatched	0.25	0.26				
	Matched	0.25	0.25				
Winter 2000/2001	Unmatched	0.27	0.24	6.3	65.9	2.4	0.02
	Matched	0.27	0.26	2.1		0.6	0.55
Spring 2001	Unmatched	0.23	0.25	-6	78.1	-2.2	0.03
	Matched	0.23	0.22	1.3		0.4	0.71
Summer 2001	Unmatched	0.25	0.24	1.3	-224	0.5	0.62
	Matched	0.25	0.26	-4.4		-1.2	0.22
Index of Multiple Deprivation							
Bottom decile (poorest areas)	Unmatched	0.11	0.17				
	Matched	0.11	0.11				
10-<20	Unmatched	0.15	0.13	4.3	32.7	1.6	0.11
	Matched	0.14	0.15	-2.9		-0.8	0.42
20-<30	Unmatched	0.17	0.11	15.4	94.2	6.0	0.00
	Matched	0.16	0.16	-0.9		-0.2	0.81
30-<40	Unmatched	0.12	0.10	6.7	70.3	2.6	0.01
	Matched	0.12	0.11	2		0.6	0.58
40-<50	Unmatched	0.07	0.10	-10	88.8	-3.5	0.00
	Matched	0.07	0.07	1.1		0.3	0.73
50-<60	Unmatched	0.06	0.09	-9.6	97.5	-3.4	0.00
	Matched	0.07	0.07	-0.2		-0.1	0.94
60-<70	Unmatched	0.06	0.08	-5.9	74.9	-2.1	0.04
	Matched	0.06	0.06	1.5		0.4	0.66
70-<80	Unmatched	0.08	0.08	0.6	-345	0.2	0.83
	Matched	0.08	0.08	-2.6		-0.7	0.48
80-<90	Unmatched	0.07	0.07	2	-130	0.8	0.45
	Matched	0.08	0.06	4.6		1.3	0.19

(Continued)

Table A1. (Continued).

	Sample	Mean		% bias	% reduction in bias	t-test	
		Wales	England			t	p > t
Highest decile	Unmatched	0.11	0.08	10.5	73.7	4.1	0.00
	Matched	0.12	0.12	-2.8		-0.7	0.48
Birth weight							
1	Unmatched	0.06	0.08				
	Matched	0.06	0.05				
2	Unmatched	0.15	0.18	-5.9	40.1	-2.1	0.03
	Matched	0.16	0.17	-3.5		-1.0	0.32
3	Unmatched	0.15	0.13	4	-7.5	1.5	0.13
	Matched	0.14	0.13	4.3		1.3	0.21
4	Unmatched	0.18	0.16	5.3	96.8	2.0	0.05
	Matched	0.17	0.17	0.2		0.1	0.96
5	Unmatched	0.13	0.14	-2.5	70.6	-0.9	0.36
	Matched	0.13	0.13	-0.7		-0.2	0.83
6	Unmatched	0.13	0.13	1.6	6.7	0.6	0.55
	Matched	0.13	0.14	-1.5		-0.4	0.68
7	Unmatched	0.09	0.09	0.8	-124	0.3	0.77
	Matched	0.09	0.10	-1.7		-0.5	0.63
8	Unmatched	0.12	0.10	5.1	96	1.9	0.06
	Matched	0.12	0.12	0.2		0.1	0.96
Breasted							
Not breastfed	Unmatched	0.79	0.70				
	Matched	0.78	0.79				
Up to 3 months	Unmatched	0.02	0.04	-10.8	78.9	-3.6	0.00
	Matched	0.02	0.02	2.3		0.8	0.42
3–6 months	Unmatched	0.10	0.15	-14.6	82.9	-5.1	0.00
	Matched	0.10	0.09	2.5		0.8	0.44
6 months or more	Unmatched	0.10	0.12	-7	68.4	-2.5	0.01
	Matched	0.10	0.10	-2.2		-0.7	0.52

This example of propensity score matching is based on 9279 valid cases (7615 in England and 1664 in Wales).