Disadvantaged learners’ : who are we targeting? Understanding the targeting of widening participation activity using geo-demographic data from Southwest England

Neil Harrison (Senior Research Fellow, School of Education, University of the West of England)
Sue Hatt (Regional Manager, Aimhigher Southwest)

Abstract : This paper analyses the definition of the appropriate target group for widening participation activities advanced by the Higher Education Funding Council for England in their Targeting Disadvantaged Learners advice to Aimhigher and other practitioners. This definition includes components of area deprivation and higher education participation rates, which are apparently intended to act as a proxy to reach learners from lower socio-economic groups.

Through statistical analysis of geo-demographical data from the Southwest region of England, this paper questions whether the HEFCE targeting guidance is likely to meet the policy aims which underpin it and reach the ‘disadvantaged learners’ of its title. It is found that the geographical proxy tends to miss learners from lower socio-economic groups in areas of wider affluence and those in rural areas. The paper concludes by questioning whether the targeted areas are likely to be the most fruitful locations for outreach activities in the short-term given the ingrained, multi-faceted and multi-generational challenges which they face.
Neil Harrison is a Senior Research Fellow at the University of the West of England, Bristol. His research interests include widening participation, student retention, student financial support policies and the inter-cultural student experience.

Sue Hatt is Principal Lecturer in Economics at the University of the West of England, Bristol, but is currently seconded to manage the Aimhigher Southwest programme and is also a member of the Action on Access team. Her main research interests include feminist economics and student progression and achievement in higher education.

**Corresponding Author**: Neil Harrison, School of Education, University of the West of England, Coldharbour Lane, Frenchay, Bristol, BS16 1QY.

Telephone : 0117 328 4190

E-mail : Neil.Harrison@uwe.ac.uk
**1. Targeting learners to widening participation**

Targeting is central to widening participation. It determines those who do – and do not – take part in widening participation programmes, those whose awareness of higher education is – and is not – increased and those whose aspirations to progress to higher education are – and are not – supported by activity programmes. If the targeting criteria are not fit-for-purpose, scarce resources will be misdirected towards young people from groups that are already well-represented in higher education (Thomas, 2001) or alternatively spent on those for whom higher education is not at present a realistic option. Accurate targeting is therefore of prime importance in evaluating whether funding has been used effectively.

Although ‘good’ targeting is an essential prerequisite for widening participation, finding the ‘right’ participants is challenging. It requires clarity of definition, appropriate selection criteria and accurate processes for operationalising the criteria. Since 1997, the focus of widening participation policy has been on groups that are under-represented in the higher education student population. These have been defined as

“*those from socio-economic groups III to V, people with disabilities and specific ethnic minority groups.*” (NCIHE, 1997, p. 14, para 29)

By far the most numerous group is those from lower socio-economic groups, now defined by the National Statistics Socio-Economic Classification (NS-SEC) as groups 4 to 8\(^1\), who have remained the priority for widening participation for the last ten years (DfES, 2003; Greenbank, 2006; HEFCE, 2007a). The social class gap has, however, proved difficult to close (NAO, 2008).

\(^{1}\) This is sometimes rendered as groups 4 to 7. Group 8 comprises household which have been long-term unemployed.
Clarity of definition has thus been available for over a decade, but choosing appropriate selection criteria and putting them into practice has proved more challenging. Aimhigher\textsuperscript{2} partnerships and widening participation managers are not able to use data on the occupational background of young people’s parents as the basis for selection. In response they have turned towards proxies, which pick up on particular dimensions of social class. For example, the link between class and parental education has been a popular criterion. This approach focused on the families’ cultural capital (Bourdieu, 1997) gave selected participants a sense of pride and resonated with policy documents that made explicit references to young people “who come from families with no HE in their backgrounds” (DfEE, 2000, p. 3). Although partnerships have shown remarkable accuracy in reaching those with no parental experience of higher education (Hatt, Baxter & Tate, 2005), it has not proved a good proxy for social class; nor has parental income (Hatt, Baxter & Harrison, 2003), nor school type (Universities UK, 2005).

An alternative approach has been to focus on the geography of under-representation, an approach that was validated by recommendations to target those “living in deprived geographical areas, including deprived rural and coastal areas” (HEFCE, 2004, p. 10) or “from inner city comprehensives” (DfEE, 2000, p. 2). Although postcode data makes geographical criteria relatively easy to operationalise, its reliability as a proxy for social class has been questioned in both rural and urban areas and it can lead to:

“the wrong side of the street’ issue where a young person on one side of the street qualifies for help but someone on the other side does not.” (Blicharski, 2000, p. 179)

Serious practitioner concerns about the use of postcode data for targeting learners were expressed in the Universities UK’s (2002) \textit{Social Class and Participation} report, while Osborne & Shuttleworth (2004) found that although postcode proxies could be effective at the extremes, they had much

---

\textsuperscript{2} \textit{Aimhigher} is an outreach programme operating across England to raise awareness of and aspirations for higher education amongst young people from groups that are currently under-represented within the student population. It was formed in 2004 through the merger of two closely related programmes, \textit{Excellence Challenge} and \textit{Partnerships for Progression} and is delivered through area partnerships spanning schools, colleges and universities.
lower validity in the middle ranges. Given that nationally around half of young people come from households where the main wage earner is employed in an occupation classified within NS-SEC groups 4 to 8, focusing on extreme deprivation is unlikely to be sufficient. As UK Secretary of State John Denham said in his speech to HEFCE’s 2008 conference:

“Widening participation is an issue for the majority, not the minority” (Denham, 2008)

By the mid 2000s, there was growing concern that inaccurate targeting was dissipating the widening participation resource which could be more effective were it used more selectively. Consequently, HEFCE (2007a) published guidance to reiterate the emphasis on lower social groups and to provide a methodology for effective targeting.

The HEFCE guidance proposed a three stage targeting process by which partnerships: (a) firstly identified the schools with which they would work most intensively, (b) secondly worked with teachers to locate individuals who met the targeting definition and, (c) thirdly, monitored the socio-economic status of those who took part in their activities to feedback into the cycle and improve the process. This article will focus on the second step in that process, that is the targeting of individuals. While it is important to acknowledge that “there is no correct way to target” (Blicharski, 2000, p. 179), and that no methodology will be 100% accurate in locating the target group, this article will explore the proposed methodology to consider the extent to which it is likely to be accurate in locating individuals from lower socio-economic groups who have the potential to benefit from higher education.

2. Using geographical proxies to identify individuals

A policy feature of recent years has been the increasing governmental emphasis on so-called ‘area-based initiatives’, where resources and other forms of intervention have been employed on a geographically targeted basis in order to reach individuals and communities considered to be
deprived or disadvantaged under some form of criteria. Examples from the field of education have included Education Action Zones (to raise school attainment) and the Sure Start initiative (to improve pre-school provision).

Area-based initiatives have been justified on the basis that disadvantage tends to be geographically concentrated and that there are specific effects in deprived areas which heighten levels of disadvantage and which need focused solutions. A review of the debate on the theoretical foundation of area-based initiatives is beyond the scope of this paper (see Dorling et al, 2001 for a useful overview). However, doubts have recently been expressed about the success of these initiatives in meeting their objective of reducing educational inequalities (e.g. Gewirtz et al, 2005; Batey & Brown, 2007; Rees, Power & Taylor, 2007).

With any form of area-based targeting, there is the potential for two types of ‘error’. As Batey & Brown (2007) explain,

“a degree of inefficiency is built into targeting by area, because people who are not the intended beneficiaries will be included. At the same time, targeting will be incomplete, because intended beneficiaries living outside the targeted area will be excluded.” (p. 2775)

In terms of targeting ‘disadvantaged learners’ from lower socio-economic households, these errors would be:

I. engaging with those from NS-SEC groups 1 to 3 who live within the targeted areas, and

II. failing to engage with those young people from NS-SEC groups 4 to 8 who live outside the targeted areas.

The Type II Error describes the stereotypical postcode lottery where ‘worthy’ young people living in the ‘wrong’ area are excluded due to where they happen to live (i.e. incompleteness). The Type I Error reflects wasted effort spent on the ‘unworthy’ young people who serendipitously live in a
targeted area (i.e. inefficiency); if the objective is to reduce disparities in deprivation or behaviour, this effort may even be counter-productive.

Against this background and the government policy objective to widen participation in higher education and increase equity of opportunity (Labour Party, 2005), the key passage from the HEFCE targeting statement to which this paper will refer explains that,

“resources should be targeted at learners with the potential to benefit from higher education who come from under-represented communities. Overwhelmingly these learners are from lower socio-economic groups … and those from disadvantaged backgrounds who live in areas of relative deprivation where participation in HE is low.” (HEFCE, 2007, p. 8, para 21)

The first aspect of the statement is that the focus is on individuals with “the potential to benefit from higher education”. This was first established as a key principle within the sector’s thinking as early as the 1960s (Robbins, 1963). Widening participation is about extending opportunity to enter higher education to those who have the potential but, due to their educational or social background, might not consider it as a realistic option. These young people can be extremely able and studies (Hatt, Baxter & Harrison, 2003; Hatt, Hannan & Baxter, 2005; Crozier et al, 2008a) indicate that, once at university, students from under-represented groups can demonstrate a high level of commitment to their studies that tends to reinforce success. Widening participation is not about ‘more means worse’ and the targeting guidance reinforces this by recommending a focus on those with potential to benefit from higher education.

The second aspect is about increasing equity and has two components:

1. To target learners from households which fall into NS-SEC groups 4 to 8.
2. To target learners who are resident in geographical areas where deprivation is high and youth progression to higher education is low.
The first component is self-explanatory and refers directly to the target audience established over a number of years due to the historic under-representation of these groups (NCIHE, 1997; Thomas 2001; Hatt et al 2005; Gorard et al, 2006; NAO, 2008). However, the identification of these individuals is often problematic as schools, colleges and local authorities do not collect information on socio-economic classification and so individuals from these groups are usually located through various proxies.

This is precisely what the second component seeks to establish. The theory underpinning this geographical proxy is not discussed, but it would appear that there is a working assumption that individuals in NS-SEC groups 4 to 8 are most likely to be found in these sorts of areas. There is also an implicit assumption that these two types of area (high deprivation and low participation) are correlated to a significant degree.

3. Research question

This paper will therefore seek to examine the inter-relationship between the three pieces of data which are used by the HEFCE targeting statement (i.e. NS-SEC group, area deprivation and participation rates). It will use spatial data from the Southwest region of England to explore to what extent area measures of deprivation and participation in higher education provide a useful proxy for identifying individual ‘disadvantaged learners’ with ‘the potential to benefit’ from higher education.

The article will conclude with a discussion about the policy implications of the findings and about the extent to which widening participation activity needs to work differently with contrasting communities and groups of learners. For example, some learners from under-represented groups already have Level 3 qualifications, but have never considered accessing higher education. For them, attainment is not the issue and, if higher education can be presented as a meaningful opportunity, a significant difference can be made in a relatively short period of time. For others, on
the other hand, multi-generational disengagement with education coupled with limited access to post-16 provision presents a substantial challenge that can only be addressed in the longer term. As a result, outreach activity may need to work to different aims and be evaluated against different criteria depending upon the communities and individuals at which it is targeted.

4. Methodology

A dataset was assembled in the SPSS v13 statistics software package containing the following pieces of information, indexed by Output Area. Each Output Area represents around 100 to 150 contiguous households (for a full explanation of the spatial geography used, see Appendix A):

1. **Proportion of NS-SEC 4 to 8 households (‘SEC2001’) -** drawn from the 2001 Census (ONS, 2008a) and using the NS-SEC classification of the Household Reference Person. The metric analysed in this paper is the percentage of households which are classified within NS-SEC groups 4 to 8 relative to the total, excluding households which were defined as ‘unclassifiable’. It therefore ranges from 0% to 100%.

   The SEC2001 dataset followed a normal distribution with a mean of 51.3% households being classified within the lower socio-economic groups. The English mean (from the 2001 Census) is 49.9%; the Southwest region thus has slightly more households from lower socio-economic groups than the national average.

2. **Index of Multiple Deprivation score (‘IMD2007’) -** drawn from data published by the Department for Communities and Local Government in late 2007 (DCLG, 2007). It provides a single scalar ‘score’ which represents the relative deprivation of an area across a range of factors including income, employment, education, crime and health. Scores range from 0 (low deprivation) to up to around 100 (high deprivation). IMD scores are
available at the lower level Super Output Area (SOA) level and so the same figure is shared by all the Output Areas in the same SOA.

The IMD dataset showed a distribution which is heavily skewed, with the majority of Output Areas clustered at the lower end of the deprivation scale and with a long tail of increasingly deprived areas. The mean IMD score was 17.88 for the Southwest; when compared with the English mean of 21.67, this shows the region overall to be less deprived than average.

3. **Youth rate of progression to higher education (‘POLAR2’) -** drawn from the POLAR2 dataset published by HEFCE (HEFCE, 2007b). It relates to the proportion of the estimated 18 year old cohort progressing to full-time higher education aged 18 or 19 and remaining until the 1st December census date. The metric provided is a five-point ordinal ranking between 1 (low participation) and 5 (high participation) which broadly represents the quintiles of the continuous distribution for the entry cohorts between 2000 and 2004. POLAR2 categories are available at the Ward level and so the same figure is used for all the Output Areas in the same SOA.

[Table 1 here]

The POLAR2 scores for the Southwest region are somewhat skewed towards the mid-range (POLAR categories 3 and 4), being less well represented at the extremes than the picture for England as a whole.

In addition to the three main components of the dataset, information was also collected from the National Statistics Postcode Directory relating to the relative rurality of the each Output Area (ONS, 2008b). This is expressed on a four point scale: 1 = urban, 2 = small towns and fringes, 3 = villages and 4 = hamlets and isolated dwellings.

[Table 2 here]
As can be seen from the table above, the Southwest region is notably less urban than the English average. It can be typified as being comprised of a small number of cities and large towns, with large rural expanses in between, dominated by villages and hamlets. Some of these rural areas (e.g. Dartmoor and Cornish coastal areas) are very remote and there is concern (Shucksmith, 2000; Commission for Rural Communities, 2004) about access to higher education from locations like these.

5. Analysis of the interaction between the measures

A. POLAR2 and SEC2001

As predicted by the HEFCE targeting statement, households classified within NS-SEC groups 4 to 8 are more likely to be found in areas where youth participation in higher education is low. However, the variation within the POLAR2 categories is such that the relationship, while statistically significant, is not strongly predictive.

This can be seen in more detail in Table 3 and Figure 1 in Appendix B. There is a clear distinction in the distribution of SEC2001 percentages between each of the POLAR2 categories and a one-way ANOVA test (F = 1289.993, d.f. = 4 & 17009, p < 0.001) demonstrates there to be a significant difference between means. However, interquartile ranges overlap for each of the POLAR2 categories except 1 and 5. Whilst the relationship between POLAR2 and SEC2001 is significant, the variability within the distributions means that neither measure is a good predictor for the other, except at the extremes. In other words, neighbourhoods with concentrations of lower socio-economic households can be found in areas across the whole range of POLAR2 categories.

[Table 3 here]
B. SEC2001 and IMD2007

Similarly, there is a significant positive correlation ($r = 0.556, p < 0.001$) between the SEC2001 and IMD2007 measures, such that greater deprivation within Output Areas is generally associated with higher proportions of households from lower socio-economic groups. Once again, however, despite the statistical significance of the correlation between the two measures, there is only limited predictive power in the relationship.

Figure 2 in Appendix B shows significant variability at the lower end of the IMD2007 scale, with less deprived areas being occupied by a wide range of households. Conversely, Output Areas with higher levels of deprivation are more uniformly to be associated with the prevalence of lower socio-economic groups. In other words, a neighbourhood with few households from lower socio-economic groups is very likely to be found within an area of low deprivation. However, the picture is much less certain for areas with higher proportions of households from NS-SEC groups 4 to 8, which are found across the whole geographical gamut of deprivation. Specifically, it is relatively common for there to be concentrated enclaves of households from lower socio-economic groups in wider areas of affluence; the opposite is less common.

C. POLAR2 and IMD2007

Parallel to understanding how the two proxies operate in terms of locating households containing the 'disadvantaged learners' to be targeted, it is also important to consider how the two proxies themselves interact. In the Southwest region, there is a strong relationship at the affluent end of the spectrum, such that areas of low deprivation are almost always associated with high youth progression rates into higher education. However, the association is much less strong at higher levels of deprivation.
There is a statistically significant relationship between POLAR2 categories and IMD2007 scores ($F = 1707.558$, d.f. = 4 & 17009, $p < 0.001$)\(^3\). As can be seen in Table 4, mean IMD2007 scores fall rapidly as one moves from POLAR2 categories 1 to 5, indicating that higher youth progression rates are generally associated with areas of lower deprivation.

[Table 4 here]

In addition, the IMD2007 scores showed a strongly significant difference between the variances in each of the POLAR2 categories ($L = 578.345$, d.f. = 4 & 17009, $p < 0.001$). This is reflected in the relative heights of the ‘boxes’ in Figure 3 in Appendix B. There was significantly more variability in POLAR2 category 1 than in the other categories, suggesting that areas with a low youth progression rate could include neighbourhoods with a very wide range of deprivation levels. Conversely, Output Areas in the higher POLAR2 categories were uniformly affluent. This limits the power of IMD2007 scores to predict progression rates, or vice versa, especially in areas with the lower participation in higher education.

D. Effect of rurality

Since its inception in 2000, there have been concerns about whether the statistical construction of the Index of Multiple Deprivation has led it to be biased towards or against rural areas (Deas et al, 2003; Tunstall & Lupton, 2003; Countryside Agency, 2004; Knowles, 2006; OCSI & JH Research, 2008). Similarly, previous studies tend to suggest that young people in urban areas might be advantaged in respect to demanding higher education due to, for example, better access to information (UCAS, 2002) and their ability to live at home while studying (Commission for Rural Communities, 2006).

\(^3\) As Levene’s test revealed heterogeneity of variances, a Brown-Forsythe test was also undertaken confirming the findings of the standard one-way ANOVA ($B-F = 1610.207$, d.f. = 4 & 10616.415, $p < 0.001$). ANOVA results are reported in the main text throughout for ease of comparison and understanding for the general reader. In all instances, the relationships reported were also significant under the Brown-Forsythe test.
Through the use of the four-part measure of rurality, it was found that households from the lower socio-economic groups were relatively evenly spread across different densities of population within the Southwest. However, deprivation was significantly higher in urban and the most remote areas, while low participation rates were overwhelming more likely to be found in urban areas.

There is a very clear and significant relationship between rurality and the POLAR2 categories ($X^2 = 2494.663$, d.f. = 12, $p < 0.001$), with rural areas, in general, being associated with higher youth progression rates, as can be seen in Table 5. Thus, over 90% of POLAR2 low participation areas were to be found in urban localities, while areas of high participation were particularly over-represented in villages and hamlets.

Table 6 shows the relationship between rurality, deprivation and concentration of low socio-economically grouped households. There is a significant difference between the IMD2007 mean scores ($F = 222.025$, d.f. = 3 & 17010, $p < 0.001$), with Output Areas in urban areas and hamlets showing a higher level of deprivation than towns and villages. With the SEC2001 dataset, the distribution of households from lower socio-economic groups is relatively evenly spread across the four rural/urban categories. While there are statistically significant differences in the means ($F = 31.431$, d.f. = 3 & 17010, $p < 0.001$), these are very small, with mean proportions varying between 50% and 54% across the range of population densities.

In summary, while areas in the lowest POLAR2 categories are nearly exclusive confined to urban areas, the neighbourhoods which are defined as relating to the more remote rural areas have, on average, only slightly less deprivation and, in fact, a higher proportion of households from NS-SEC groups 4 to 8. Looking at this from a different perspective, Output Areas in hamlets and villages
which look similar to those in cities in terms of deprivation and percentage of households from the lower socio-economic groups will typically have higher youth participation rates.

E. Which areas does the HEFCE definition effectively identify?

HEFCE (2007a) explains that ‘relative deprivation’ should be interpreted to mean the highest 13,000 IMD2007 scores, effectively describing the 40% most deprived Super Output Areas. Although the document provides no definition of ‘low participation’, this has generally been interpreted to mean POLAR2 categories 1 and 2. Remembering that the POLAR2 categories are quintiles, this working definition therefore includes 40% of wards. The geographical proxy therefore targets areas which simultaneously are both in the top 40% for deprivation and bottom 40% for participation rates.

Looking at this from the perspective of the Southwest region, we find that only 29% of Output Areas have IMD2007 scores which fall in the highest 13,000 nationally, while 36% fall within POLAR2 categories 1 or 2. The conjunction of these two measures identifies only 19% of the total Output Areas in the region, which are very heavily grouped within the region’s urban areas.

At this point, it is possible to estimate the error rates outlined using the methodology and typology for area-based studies proposed in Batey & Brown (2007). Based on an assumption of equally-sized Output Areas each with 100 households, Table 7 below demonstrates that the HEFCE statement captures only one-quarter of the households in the Southwest which were classified as being in NS-SEC Groups 4 to 8 at the time of the 2001 Census.

[Table 7 here]

The exact numbers of households in each Output Area was not included in the dataset analysed, although the data is available and may be incorporated in future work. Nevertheless, using a mean of 100 households provides a reasonable estimate for the purposes of this exploratory paper.
Correctly targeted  = \( \frac{213,700 + 713,300}{1,701,400} \times 100 = 55\% \)

Type I Error (inefficiency)  = \( \frac{116,000}{1,701,400} \times 100 = 7\% \)

Type II Error (incompleteness)  = \( \frac{658,400}{1,701,400} \times 100 = 39\% \)

We find that 55% of households are correctly allocated to targeted or non-targeted areas depending on their socio-economic status. Of those incorrectly targeted, incompleteness (i.e. failing to include ‘right’ households in the target area) is a bigger source of error than inefficiency (i.e. targeting ‘wrong’ households within the target area). This would appear to be a high error rate for the geographical proxy, especially in terms of the number of households from lower socio-economic groups which have been missed because they are in areas which are too affluent or too successful at sending young people on to university.

F. Summary of main findings

From the statistical analysis above, the following findings can be deduced:

1. There are statistically significant and positive relationships between the proportion of households in NS-SEC groups 4 to 8 and both (a) high levels of deprivation, and (b) low higher education participation rates.

2. However, these relationships do not have a strong predictive value. In particular, areas of high deprivation and/or low participation contain highly variable proportions of households from the lower socio-economic groups.

3. Smaller areas with concentrations of households from NS-SEC groups 4 to 8 are often to be found in larger areas of relative affluence; the opposite is less common, but not unknown.
4. There is a strong urban bias both in terms of areas of high deprivation and low progression rates, but relatively little difference between locales in terms of proportions of households from the lower socio-economic groups.

5. Due to the relatively limited overlap between deprivation and participation, the HEFCE geographical proxy identifies just under 20% of areas in the Southwest as likely locations for concentrations of the target groups.

6. Only just over half of households are appropriately allocated to targeted or non-targeted areas by the proxy.

6. Discussion

A. Is POLAR2 a good basis for targeting disadvantaged learners?

The implicit assumption in HEFCE (2007a) is that the areas which have low POLAR2 scores are those where the greatest gains in equity can be made – i.e. where the most young people from lower socio-economic backgrounds can be encouraged to aspire to higher education through initiatives and activities such as those organised by Aimhigher or individual higher education providers. However, the findings of this paper cast doubt on that assumption.

Firstly, there is not a strong correlation between progression rates and socio-economic groups when looking at the Output Area level of definition. POLAR2 categories 1 and 2 do contain 51% of Output Areas with above-mean proportions of households from the lower socio-economic groups, but they therefore exclude 49%. They also capture 21% of Output Areas which are predominantly from NS-SEC groups 1 to 3. Areas with low youth progression rates do have a distinct flavour by socio-economic grouping, but they are far from monocultural.
Because of the low geographical definition of the POLAR2 dataset, there is a smoothing effect. Where two communities live adjacently within the same ward, the overall progression rate is averaged. One of the failings of the POLAR2 calculations is, therefore, that it lacks the geographical definition to identify small communities of lower socio-economic groups with low progression rates. This is likely to occur particularly in rural areas, where the physical area covered by a ward is often very large and may include numerous settlements.

Secondly, we have noted that there is a strong bias towards low participation wards within the urban concentrations in the southwest. In some ways this is counterintuitive, as one might imagine that proximity to higher education providers might work to boost aspirations through access to information, community activities and localism (UCAS, 2002; Reay, David & Ball, 2005). Output Areas with similar profiles by socio-economic mix and deprivation tended to fall into higher POLAR2 categories if they were in rural areas than if they were in urban areas. It is likely that this is, in part, related to the environment in rural secondary schools, which, due to their wider geographical catchment areas, are likely to draw in a more mixed demographic profile with positive influences through peer pressure, role modelling and teacher expectations.

B. What does the IMD2007 measure actually measure and how does this impact on targeting?

The second component of the HEFCE targeting proxy is that interventions should be focused on those living in the 40% of Super Output Areas nationally with the highest levels of deprivation. While the very highest levels of deprivation are reliably related to low progression rates, areas with mid and low levels of deprivation can be found across the whole range of POLAR2 categories. There is a statistically significant link between deprivation and participation, but its predictive power is not strong. Similarly, there is a significant correlation between the IMD2007 score and the SEC2001 percentage in a given Output Area, but there is considerable variability, in areas with lower levels of deprivation.
Pulling these two findings together, we find that the proxy value of the IMD2007 is rather compromised. Similarly to the POLAR2 categories, it is effectively blind to small pockets of lower socio-economically grouped households in otherwise less deprived areas, while it is also not a good predictor for university progression where this is low.

[Table 8 here]

It is useful to understand the Output Areas which are excluded from the HEFCE target definition. 34% of Output Areas (1,671 out of 4,968) showing high levels of deprivation are excluded as they have relatively high youth progression rates, while 46% of Output Areas (2,761 out of 6,058) with low progression rates are excluded because they are not sufficiently deprived. Perhaps most significantly, the net result is that the HEFCE targeting statement effectively excludes two-thirds of Output Areas in the Southwest with above-mean percentages of households from lower socio-economic groups.

Table 8 demonstrates this relationship. 3,297 Output Areas in the Southwest meet both components of the HEFCE statement. However, a further 4,432 (2,761 + 1,671) meet one part, but not the other. A companion paper to this one (Harrison & Hatt, in preparation) investigates these areas in detail through the use of case studies. It is found that such areas appear, *prima facie*, to provide good opportunities for widening participation, with above average proportions of households from the lower socio-economic groups, schools which have levels of attainment which can support mainstreamed progress into higher education and economies which provide some opportunity for employment types which require higher levels of qualification.

C. Is the bottom 19% the best place to increase participation from ‘disadvantaged learners’?

The calculations above suggest that in the Southwest of England, 3,297 Output Areas would be included in the HEFCE proxy to identify ‘disadvantaged learners’, comprising just 19% of the total.
Based on the assumption that Output Areas are of broadly equivalent population, 65% of the households under this definition would be classified as being from lower socio-economic groups. The comparable figure for the areas falling outside the targeting statement would be 45%. At first glance, it would therefore appear that this bottom 19% is a relatively strong starting point for finding disadvantaged learners.

Progression to university is, in essence, the result of the interaction between the individual and the education sector. In order to demand higher education, the young person must both have the post-16 results which permit entry and the aspiration to do so. Both of these are related to the supply of post-16 education in the area. Similarly, the higher education sector itself controls supply through the availability of student numbers, their entry requirements and their geographical distribution.

Studies of the areas which meet the HEFCE targeting criteria have found that they often contain schools with poor educational outcomes (Raphael Reed *et al*, 2007; Raphael Reed, Gates & Last, 2007; Harrison & Hatt, in preparation), to the point that progression to post-16 study is very much a minority pursuit and higher education is simply not a viable option for the vast majority. More generally, Warrington (2005) argues for the importance of ‘place’ in regard to educational opportunity and social exclusion, while Lupton (2005) found that the ‘quality’ of schools in heavily deprived urban areas was subject to a downward pressure from high staff turnover, administrative pressures, lack of parental support, low teacher expectations and inadequate resourcing, producing systemic fragility and severely limiting their ability to achieve better results.

In their HEFCE-sponsored study of south Bristol, Raphael Reed *et al* (2007) found strong multi-generational and multi-faceted cultures working against educational engagement and university progression. Parents and communities had memories of how the school system had failed them, coupled with a sense of disempowerment for many (Gewirtz *et al*, 2005; Brine, 2006). It has been known for nearly fifty years (Jackson & Marsden, 1962) that ‘middle class links’ are important in framing educational aspirations for the working class. However, in their analysis of the four urban
areas in England with the lowest higher education participation rates, Raphael Reed, Gates & Last (2007) found them to have very high levels of deprivation and to contain strong concentrations of households from the lower socio-economic groups, assembled into communities that had lower levels of the ‘linking’ and ‘bridging’ social capital (Woolcock, 2001) which would enable them to access middle class pathways and without the ‘hot’ information about higher education which would support participation (Archer, Hutchings & Ross, 2003; Reay, David & Ball, 2005). They found schools which were struggling to cope with the demands placed on them, similar to those described by Lupton (2005). This is not to infer a deficit on the part of the community (Mills & Gale, 2004; Gerwirtz et al, 2005), but of the social and educational system which predominates there and which often has done for some time. This is often aggravated by the ability of wealthier families to school their children outside the area.

This backdrop is unlikely to be challenged in the short-term by widening participation initiatives. The tension between the longer term objective of raising attainment and educational outcomes for those in the lowest achieving communities and increasing higher education participation has long been present in the Aimhigher programme and is reflected in the objectives of the two funding bodies. The central government is primarily interested in schools and educational attainment, while HEFCE’s focus has been on increasing and widening participation to higher education. Although these objectives can be reconciled, particularly when low attainment is preventing young people from reaching their aspiration, there are many communities for whom one or other of the objectives predominates. For example, some young people have attained the higher education entry requirements, but never used these to progress to higher education, either because of the lack of local provision or ‘pull factors’ from the labour market. For these young people, raising attainment is not an issue, but higher education participation is. On the other hand, in schools where attainment is an issue, Aimhigher and other widening participation practitioners can work with the school to help those on the borderline at GCSE but, against a background of challenging circumstances in the school and widespread educational disengagement within the community, the impact of this work on progression will be limited in the short-term.
7. Conclusion

Rees, Power & Taylor, (2007) conclude from their analysis of evaluations across a range of recent area-based initiatives that the evidence for the successful achievement of the objectives is limited. Possible explanations include the foibles of demographic distribution or the competitive bidding process underpinning the allocation of resources. They find that the timescales employed are unrealistic to demonstrate meaningful success in challenging areas, with cultures of ‘short-termism’ and ‘initiative overload’ being commonplace. They suggest that initiatives

“are underpinned by overly simplistic and under-theorised conceptions of inequality and geography, which fail to recognise the complexity of processes by which social disadvantage are generated” (p. 267).

Returning to the error typology explored above, it was proposed that targeting could fail by either wasting effort on the ‘wrong’ individuals in the ‘right’ areas (Type I Error - inefficiency) or missing the ‘right’ individuals in the ‘wrong’ areas (Type II Error - incompleteness).

In terms of the Type I Error, even the areas targeted by proxy do not have particularly high densities of households from NS-SEC groups 4 to 8. This is partly due to the foibles of the geographical boundaries which are used and the level of definition which is available in the participation and deprivation datasets. In fact, a little over a third of the households in the targeted areas in the Southwest are actually from higher socio-economic groups. The young people within these households will be offered the opportunity to benefit from activities and initiatives which may not be needed to encourage them to demand higher education (Osborne & Shuttleworth, 2004). As such, working with these individuals will not fulfil the government’s objective of increasing equity. It may even decrease equity as the ‘sharp elbows of the rich’ take up a disproportionate share of resources, possibly with tacit complicity from the school (Thomas, 2001; Universities UK, 2002; Hatt, Baxter & Tate, 2005; Crozier et al, 2008b).
In terms of the Type II Error, we have established that many young people from lower socio-economic groups will be missed by the HEFCE targeting statement. Historical participation rates and deprivation scores are not very good proxies for finding young people from NS-SEC groups 4 to 8. There is a relatively strong correlation in affluent areas, but the distribution becomes much more variable in more deprived areas. By limiting the proxy to “areas of relative deprivation where participation in HE is low” the focus becomes on the very most deprived communities where progression to university is very uncommon. It is questionable whether these are the types of areas where the most significant advances might be made, in the short-term, at least.

Even outside the proxy target areas, just under half of households are actually from the lower socio-economic groups. Young people from these households could be denied access to positive activities and initiatives, especially if their school is not one of those targeted. These young people are, almost by definition, living in areas with a more affluent and educationally-focused outlook. They are likely to have access to more ‘linking’ and ‘bridging’ social capital and educational provision which is supportive of university entrance, and, as such, more likely in the short-term to have their motivation for higher education increased than their peers in the most educationally and socially deprived areas. Even if they live in more affluent areas, they remain members of precisely the groups which the government wishes to influence.

In addition, we have established that there is a strong bias towards urban areas inherent in the proxies, due to an over-representation of deprivation and low participation in higher education in these locales. A number of explanatory factors have been suggested in this paper, including the presence of low performing schools and the typically more mixed communities in rural areas. OCSI & JH Research (2008) found that “rural areas are substantially more deprived based on the location of deprived people than based on the location of deprived areas” (original emphasis, p. 7). There is, therefore, a serious issue of equity here, especially given the particular economic challenges of many of these areas where traditional forms of employment (e.g. farming, fishing and
tourism) are faltering. Deas et al. (2003) argue that it may not be possible to construct a single measure of deprivation which is equally applicable to rural and urban areas.

In urban areas where deep-seated, multi-generational issues affect educational performance, it is debatable whether initiatives with short-term funding can make a significant difference. The focus on the ‘most deprived’ communities might be misplaced for initiatives that need to show impact within relatively short funding periods. Short-term initiatives might best be focused on those from NS-SEC groups 4 to 8 in mixed communities where a supportive infrastructure is already in place. This would effectively result in a two track widening participation policy, with one track focusing on attainment and severe deprivation whilst the other looks to widening participation for those from lower socio-economic groups more generally. The former is what the Education Action Zones were intended to address, although Gerwirtz et al. (2005) cast doubt over their efficacy. The new Extra Mile programme (DCSF, 2008) might be a more successful approach.

With the relatively recent publication of the HEFCE targeting guidance (2007a), it is too early to judge whether or not it will be successful in guiding widening participation practitioners. However, there is good evidence from the Southwest region to cast doubt on its likelihood of success and we have argued along a parallel line to Rees, Power & Taylor (2007), who concluded that area-based initiatives

“downplay the heterogeneity within areas and the frequent coincidence of extreme wealth and poverty [such that] they often miss more poor people than they include, as well as providing resources for those who do not need them (p. 271).”

Our argument is not against the use of area-based approaches to widen participation, *per se*, nor against the social justice of addressing deprivation in its worst forms, but that, in this instance, the definition of the target group is not well-aligned with the targeting methodology and that this may undermine the scope for success against the government’s stated policy aim of narrowing the social class gap.
Tables

Table 1: the distribution of Output Areas by youth progression rates for England and the Southwest

<table>
<thead>
<tr>
<th>POLAR2 category</th>
<th>Southwest</th>
<th>England</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (progression rate &lt; 16%)</td>
<td>15.8%</td>
<td>18.9%</td>
<td>-3.1%</td>
</tr>
<tr>
<td>2 (progression rate 16 to 24%)</td>
<td>19.8%</td>
<td>19.9%</td>
<td>-0.1%</td>
</tr>
<tr>
<td>3 (progression rate 24 to 32%)</td>
<td>24.0%</td>
<td>20.7%</td>
<td>+3.3%</td>
</tr>
<tr>
<td>4 (progression rate 32 to 40%)</td>
<td>25.0%</td>
<td>20.6%</td>
<td>+4.4%</td>
</tr>
<tr>
<td>5 (progression rate &gt; 40%)</td>
<td>15.3%</td>
<td>20.3%</td>
<td>-5.0%</td>
</tr>
</tbody>
</table>

Table 2: proportions of Output Areas in the Southwest region and England by a measure of rurality

<table>
<thead>
<tr>
<th>Rural/urban category</th>
<th>Southwest</th>
<th>England</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (urban)</td>
<td>65.4%</td>
<td>80.6%</td>
</tr>
<tr>
<td>2 (small towns &amp; fringes)</td>
<td>14.0%</td>
<td>9.0%</td>
</tr>
<tr>
<td>3 (villages)</td>
<td>13.6%</td>
<td>7.2%</td>
</tr>
<tr>
<td>4 (hamlets &amp; isolated dwellings)</td>
<td>6.9%</td>
<td>3.1%</td>
</tr>
</tbody>
</table>

Table 3: mean proportion of households from NS-SEC groups 4 to 8 within Southwest Output Areas by POLAR2 categories

<table>
<thead>
<tr>
<th>POLAR2 category</th>
<th>Mean proportion of NS-SEC 4 to 8 households</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (progression rate &lt; 16%)</td>
<td>63%</td>
</tr>
<tr>
<td>2 (progression rate 16 to 24%)</td>
<td>57%</td>
</tr>
<tr>
<td>3 (progression rate 24 to 32%)</td>
<td>52%</td>
</tr>
<tr>
<td>4 (progression rate 32 to 40%)</td>
<td>47%</td>
</tr>
<tr>
<td>5 (progression rate &gt; 40%)</td>
<td>38%</td>
</tr>
</tbody>
</table>

Table 4: IMD2007 within Southwest Output Areas by POLAR2 categories

<table>
<thead>
<tr>
<th>POLAR2 category</th>
<th>Mean IMD2007 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (progression rate &lt; 16%)</td>
<td>29.98</td>
</tr>
<tr>
<td>2 (progression rate 16 to 24%)</td>
<td>21.32</td>
</tr>
<tr>
<td>3 (progression rate 24 to 32%)</td>
<td>16.67</td>
</tr>
<tr>
<td>4 (progression rate 32 to 40%)</td>
<td>12.93</td>
</tr>
<tr>
<td>5 (progression rate &gt; 40%)</td>
<td>10.92</td>
</tr>
</tbody>
</table>
Table 5: Distribution of POLAR2 categories by rural/urban categories in the Southwest

<table>
<thead>
<tr>
<th>POLAR2 categories</th>
<th>Rural/urban categories (1 = urban)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1 (progression rate &lt; 16%)</td>
<td>91%</td>
</tr>
<tr>
<td>2 (progression rate 16 to 24%)</td>
<td>79%</td>
</tr>
<tr>
<td>3 (progression rate 24 to 32%)</td>
<td>66%</td>
</tr>
<tr>
<td>4 (progression rate 32 to 40%)</td>
<td>47%</td>
</tr>
<tr>
<td>5 (progression rate &gt; 40%)</td>
<td>52%</td>
</tr>
<tr>
<td>ALL</td>
<td>65%</td>
</tr>
</tbody>
</table>

Table 6: Distribution of mean IMD2007 scores and percentages of households from NS-SEC groups 4 to 8 within Output Areas by rural/urban categories in the Southwest

<table>
<thead>
<tr>
<th>Rural/urban category</th>
<th>Mean IMD2007 score</th>
<th>Mean proportion of NS-SEC 4 to 8 households</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (urban)</td>
<td>19.52</td>
<td>51%</td>
</tr>
<tr>
<td>2 (small towns &amp; fringes)</td>
<td>14.59</td>
<td>54%</td>
</tr>
<tr>
<td>3 (villages)</td>
<td>14.47</td>
<td>50%</td>
</tr>
<tr>
<td>4 (hamlets &amp; isolated dwellings)</td>
<td>15.69</td>
<td>53%</td>
</tr>
</tbody>
</table>

Table 7: Distribution of Southwest households by NS-SEC category between Output Areas targeted and not targeted by the HEFCE statement

<table>
<thead>
<tr>
<th></th>
<th>Households in NS-SEC Groups 1 to 3</th>
<th>Households in NS-SEC Groups 4 to 8</th>
<th>All households</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target areas</td>
<td>116,000</td>
<td>213,700</td>
<td>329,700</td>
</tr>
<tr>
<td>Non-target areas</td>
<td>713,300</td>
<td>658,400</td>
<td>1,371,700</td>
</tr>
<tr>
<td>All areas</td>
<td>829,300</td>
<td>872,100</td>
<td>1,701,400</td>
</tr>
</tbody>
</table>

Table 8: Interaction between POLAR2 categories and IMD2007 scores in the Southwest

<table>
<thead>
<tr>
<th>Output Areas in the Southwest</th>
<th>In top 13,000 English SOAs for IMD2007</th>
<th>In other SOAs</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>POLAR2 categories 1 or 2</td>
<td>3,297 (19%)</td>
<td>2,761 (16%)</td>
<td>6,058 (36%)</td>
</tr>
<tr>
<td>POLAR2 categories 3 to 5</td>
<td>1,671 (10%)</td>
<td>9,285 (55%)</td>
<td>10,956 (64%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>ALL</td>
<td>4,968</td>
<td>12,046</td>
<td>17,014</td>
</tr>
<tr>
<td></td>
<td>(29%)</td>
<td>(71%)</td>
<td>(100%)</td>
</tr>
</tbody>
</table>

Appendix A : Geographical levels of definition - a methodological note

Due to ways in which they are calculated, the three individual datasets described above are only available at different levels of geographical definition. It is important to have a concept of how these levels nest together in order to understand the inter-relationship between the datasets.

This paper uses the hierarchy of geographical definition developed by the Office of National Statistics (ONS, 2008c) for population analysis.

- At the highest level of definition are ‘Output Areas’. These are of roughly equal size by population, comprising around 100 to 150 adjacent households with a similar demographic profile. Clearly the geographical area defined by Output Areas can vary considerably, especially between rural and urban locales. The SEC2001 dataset is based on Output Areas.

- Output Areas are then grouped into wider communities, termed ‘Lower Level Super Output Areas’ (SOAs). These are also of roughly equivalent size at around 800 to 1,000 households and generally between six and ten Output Areas. While there is clearly more variation, SOAs are defined so as to capture geographical areas which shared significant components of demography. The IMD2007 dataset is based on SOAs.

- Super Output Areas are then grouped to comprise electoral wards. Wards vary significantly in size and population both within and between local authority areas. While the overall intention is that they should demarcate wider communities, this is not always possible and wards are often typified by significant contrast between the SOAs within them. The POLAR2 dataset is based on electoral wards.

While postcodes were used to construct and index the dataset, no data is available at this level. Being defined for the purpose of mail delivery, they do not necessarily reflect demographic or
sociologically realities. As they are based on density of mailing addresses (including commercial ones) and not households, as well as ease of delivery, there is a highly variable number of postcodes per Output Area, especially in urban areas. As there is therefore no standardisation of population within a given postcode, their use as the basis for statistical analysis could unhelpfully skew results towards areas with few residents. As a result, Output Areas are used as the geographical unit of analysis in this paper, generally representing a small and discrete population in a close-knit area.

Within the dataset used in this study, there are 223,803 postcodes, 17,014 Output Areas, 3,226 Super Output Areas and 1,070 Wards. As can be seen in the figure below, while each Output Area has a unique piece of SEC2001 data attached to it, the data on POLAR2 and IMD2007 will be shared by each of the Output Areas in the ward or SOA. In other words, each line in the dataset has a unique value for the SEC2001 data, but each piece of IMD2007 data will be shared across an average of three lines and POLAR2 data across sixteen. This point is important in the context of identifying and understanding communities within geographical areas.

Figure 1: The hierarchical nesting of geographical units used in the dataset

---

5 Due to the timelags associated in changing boundaries to reflect new demographic realities, there is often some minor disparity between the various levels of definition. In these instances, simple arithmetic means were used to calculate the metrics within the dataset. Around a quarter of Output Areas required this treatment. The figures provided for the Southwest region differ very slightly from the actual figures due to a small number of exclusions where this form of estimation was not possible.
Appendix B : Graphical illustrations of interactions between measures

Figure 1 : boxplot of POLAR2 categories showing distribution of percentages of lower socio-economic groups by Output Area

Figure 2 : scatterplot showing distribution of IMD2007 scores against the proportion of households from NS-SEC groups 4 to 8 by Output Area

---

6 For readers unfamiliar with boxplots, the horizontal line marks the median, the box marks the quartiles above and below the median (i.e. the middle 50% of the distribution or ‘interquartile range’) and the vertical lines show the largest and smallest values, excluding identified outliers.

7 For clarity, the scatterplot shown is based on a 25% random sample of Output Areas.
Figure 3: Boxplot of POLAR2 categories showing distribution of IMD2007 scores by Output Area.
References


UCAS (2002) *Paving the Way project report: informing change in higher education and progression partnerships with the voice of the underrepresented*, UCAS: Cheltenham.


Woolcock, M (2001) The place of social capital in understanding social and economic outcomes, 


Total words (inc abstract, appendices and references) = 8,907