An analysis of detailed parental occupational differences and their effects on children's school attainment in Britain.

Vernon Gayle, University of Stirling

Paul Lambert, University of Stirling

12th April 2011

Paper presented to the ISA Research Committee 28 Spring meeting, 13-16 April 2011, University of Essex

Acknowledgements:
Work for this paper was supported by the ESRC research project 'Data Management through e-Social Science' (www.dames.org.uk, RES-149-25-1066). The data analysed were obtained from the UK Data Archive.

Abstract

It is often argued that detailed differences between occupational positions have important empirical effects upon socio-economic outcomes (e.g. Weeden and Grusky, 2005). In this paper we investigate the extent to which fine-grained measures of parental occupational positions, as are available in a major longitudinal UK social survey (Youth Cohort Study of England and Wales), add value to an analysis of children's educational attainment.

The focus of the presentation is school attainment measured by General Certificate of Secondary Education (GCSE) examination. GCSEs are public examinations and mark the first major educational branching point, and poor GCSE attainment is a considerable obstacle which precludes young people from pursuing more advanced educational courses. Young people with low levels of GCSE attainment are frequently disadvantaged in the labour market, and are also likely to have a less favourable longer term experiences in the adult labour market.

We analyse some alternative measures of GCSE attainment and consider a number of alternative classifications of parental occupations, which feature substantial variation in their level of detail using a range of GLMM models. We conclude that there is a relatively strong, and persistent, association between parental occupations and filial GCSE attainment. This is greater than the effect of both gender and ethnicity. This relationship is observed irrespective of however GCSE attainment and parental occupations are measured, however, the relationship is empirically stronger when the data on parental occupations recognises specific and detailed differences between jobs. For example the difference between school teachers and publicans, which in many UK socio-economic schemes are assigned to the same classification). We reflect upon the empirical and theoretical consequences for exploiting occupational data in studying educational attainment and in stratification research more generally.

1. Introduction

The Youth Cohort Study of England and Wales (YCS), like many large scale social surveys, features detailed data about occupations. In the YCS, which is a postal questionnaire sent to teenagers, the data available on occupations takes the form of fine-grained descriptions of the occupations of the respondent's parents. In the data analysed within this paper (Croxford et al., 2007), the occupational descriptions are coded into the UK's Standard Occupational Classification 1990 (OPCS, 1990), a taxonomy of up to 371 distinct occupational unit groups. Concise indicators of employment status are also collected in the YCS. These measures can be readily converted by means of published algorithms into various occupation-based social classifications (see section 3). They can also be converted with a reasonable degree of accuracy, into other occupational unit group taxonomies, such as ISCO-88 (OIU, 2001).

The principal objective of this paper is to evaluate the empirical features of different explanatory measures based upon occupational details, for the purposes of analysing the influence of parental background upon filial educational attainment. It is often argued that detailed differences between occupational positions have important empirical effects upon socio-economic outcomes (e.g. Weeden and Grusky, 2005). In this paper we investigate the extent to which fine-grained measures of parental occupational positions add value to an analysis of children's educational attainment. This is achieved by comparing a number of different possible occupation-based measures in what could be characterised as an extended sensitivity analysis of different occupation-based measures.

The empirical properties of occupational measures are intrinsically of interest since measures based on occupations are widely employed across the social sciences, and are extensively exploited within stratification research. Occupations are demonstrated to be central defining features of individual's lives, identities and material circumstances. We regard occupations as a preferable means to measure social circumstances due to the relative stability of occupational positions over the life-course when compared with income and wealth measures. Occupations are also relatively stable between birth cohorts when compared with the distribution of measured educational qualifications. Occupational data is also relatively easily collected through survey instruments.

In addition, the investigation of the empirical features of the transmission from parental occupational background to filial educational outcomes raises questions about the substantive interpretation of the mechanisms involved in how social background influences filial outcomes. A popular contemporary empirical strategy has been to link theoretical explanatory concepts with distinct occupation-based measures (e.g. Rose and Harrison, 2010). This line of argument states that since different occupation-based measures are conceptualised according to different theories, then the differences in patterns related to occupation-based measures can be linked to differences in the causal mechanisms associated with the distinct theory of social inequality.

Goldthorpe (2002; 2007), for instance, distinguishes between 'occupational' and 'class' based mechanisms of influence. In contrast, Chan and Goldthorpe (2007) distinguish between 'class' and 'status' effects. Whereas Jonsson et al. (2009), and Weeden and Grusky (2005), contrast the influence of 'big-class' processes (such as broad socio-economic influences) and 'microclass' effects (such as the particular expectations and cultural traditions of a specific occupational culture). By comparing the influences of a variety of more and less detailed occupation-based measures we may be able to adjudicate upon the most plausible theoretical mechanisms linked to the social origins influence upon educational attainment. Equally, if we were to find little systematic evidence of difference between measures, this may suggest that it is inappropriate to try to adjudicate between different theoretical explanations by using different occupation-based measures (cf. Lambert and Bihagen, 2007).

2. Data

The Youth Cohort Study of England and Wales (YCS) is a major longitudinal study that began in the mid -1980s. It is a large-scale nationally representative survey funded by the government. The YCS is designed to monitor behaviour of young people as they reach the minimum school leaving age and either remain in education or enter the labour market. The survey collects detailed information on the young person's experiences of education and their educational qualifications, as well as information on employment and training. A limited amount of information is collected on the young person's personal characteristics, their family and circumstances at home, and their aspirations.

The YCS sample is nationally representative of pupils in England and Wales (in school Year 11). A large sample from an academic year group (a cohort) is contacted in the spring following completion of compulsory education (Year 11). The young people are usually age

16-17 when they are first contacted. The main data collection instrument is a postal questionnaire. The cohorts are re-contacted on at least two occasions. The YCS is primarily a monitoring tool although it is sometimes analysed in social science research.

The YCS survey is organised by school leaving cohorts. Cross-cohort comparisons are potentially feasible, although in practice there are a number of practical barriers. Over the lifespan of the YCS there have been a number of major changes in the UK educational system. These changes include alterations to qualifications, the curriculum, and the structure, organisation, management and financing of schools. These changes add substantially to the complexity of comparing YCS cohorts. The survey has also been collected by different survey agencies and, at a more practical level, there have been changes to questions within the surveys and changes in measurements and coding. In addition, over the life-cycle of the YCS different government departments have been in charge of the survey and the structure and timings of data collection have varied between cohorts. A further obstacle is that the documentation of the curated YCS data is relatively poor, especially in the early cohorts.

Croxford et al. (2007) have recently deposited a harmonized dataset that comprises YCS cohorts from 1984-2002 (UK Data Archive Study Number 5765 dataset). The new data resource better facilitates cross-cohort comparative research. We analyse five YCS cohorts (cohorts 5, 7, 8, 9 and 10). These are pupils who reached the end of Year 11 (and therefore were eligible to leave education) in 1990, 1993, 1995 1997 and 1999 respectively. Young people in earlier YCS cohorts either did not undertake the same school examinations or did not have appropriate parental occupational information collected for comparable measures to be derived.

The General Certificate of Secondary Education (GCSE) was introduced in the late 1980s and is the standard qualification for pupils in England and Wales in year 11 (aged 15/16). GCSEs are usually a mixture of assessed coursework and examinations, and generally each subject is assessed separately and a subject specific GCSE awarded. It is usual for pupils to study for about nine subjects, which will include core subjects (e.g. English, Maths and Science) and non-core subjects. GCSEs are graded in discrete ordered categories, the highest being A*, followed by grades A through to G.

We argue that studying GCSE attainment is sociologically important because GCSEs are public examinations and mark the first major branching point in a young person's educational career. In Britain poor GCSE attainment is a considerable obstacle which precludes young people from pursuing more advanced educational courses. Young people with low levels of GCSE attainment are usually more likely to leave education at the minimum school leaving age, and usually never return to education. Low levels of GCSE attainment frequently disadvantages young people in the labour market, and are also likely to have a longer term impact on experiences in the adult labour market. Therefore, we argue that gaps in GCSE attainment are sociologically important.

Table 1 summarises distributional data on the sample analysed in this paper. Across the YCS cohorts we have access to information on 55120 youths who provided valid occupational data on at least one parent. The key outcome measure reported in most analyses below is GCSE points score, the average of which increases slightly for each school year (cohort), and its pattern is also associated with the respondent's gender and ethnicity. Therefore we include these controls within the analyses.

TABLE 1 ABOUT HERE

3. Operationalisation and analysis of occupation-based measures

We consider a number of alternative classifications of parental occupations, which feature substantial variation in their level of detail. As mentioned above, our starting point is, for each parent, detailed occupational unit group descriptions in the UK's Standard Occupational Classification 1990 (SOC90) (see OPCS, 1990). In addition there is a two category measure of whether or not this parent is self-employed, and if they are working part time or full time. This data facilitates a number of possible measures which we define in terms of alternative 'background measures', 'base units', and 'agglomerate schemes', and the permutations thereof. Table 2 describes the measures considered.

TABLE 2 ABOUT HERE

Historically, social scientists ordinarily favoured simpler occupation-based measures, not least because it has been believed that approaches other than using small numbers of nominal or ordinal categories are cognitively challenging (Hildebrand et al., 1977). However with expanding microdata, metadata and software resources, it is plausible that much more can be achieved with complex occupational records (Lambert et al., 2007). Administrative and operational factors may present obstacles for less detailed occupation-based measures. The practical processes of coding occupational data, and then recoding them into alternative schemes, are vulnerable to unintended aggregations (e.g. coding errors, or collapsing categories inconsistently between measures). Therefore a more detailed occupational scheme has a better chance isolating or identifying problematic categories rather than amalgamating them.

Background measures

An analysis of parental occupational influences could reasonably take account of data about the father, the mother, or both parents. There is a plausible argument that data from both parents can make a distinctive contribution to measured differences in outcomes. Equally however there are grounds to suspect that simplified or single-parental measures may prove adequate indicators in many circumstances.

In addition, an expression of a family background influence is most easily analysed in the form of one rather than two measured variables, since a dual variable approach adds modelling intricacy. In practice this is further complicated because many respondents do not have data available from both parents. We followed the well-worn strategy of deriving a single measure of 'parental occupational background', of which we used and compared four alternatives (cf. Erikson, 1984). First, a 'conventional' measure using father's occupation unless that data was missing in which case mother's occupation was used. Second, a 'family dominance' measure which prioritises full time jobs held by either parent. If more than one full time job is held (or two part time jobs), this approach prioritises the job which is more advantaged according to a stratification criteria (we used CAMSIS score on the relevant gender scale). Third, a measure of father's occupation only, ignoring mother's occupational information. Fourth, a measure of mother's occupation only, ignoring father's occupational information.

Stata syntax for deriving these measures is included in our do file appendix (www.dames.org.uk/docs/conf_papers/rc28_2011_ycs/rc28_2011_gayle_lambert.do). The specification of these four background measures led to four permutations of each subsequent measure. The four permutations are labelled in our results as conventional (cv), family dominance (pd), father (pa) and mother (ma).

Base units

For any given background measure, we had the option of using several plausible detailed occupational unit measures. The data was originally supplied in SOC90 units, which are one of three base units we analysed. SOC90 units group together occupations on the basis of a variety of criteria including shared tasks, training requirements and industrial sectors (OPCS, 1990). These distinctions are plausible empirical indicators of different occupations. The demarcation lines between the 371 different SOC90 units are, in some instances, a product of historical and administrative practices. They may not always coincide with substantially important occupational divisions. For example there are five different SOC codes corresponding to different tasks within textile processing operative occupations, but most national level sociological accounts would not be as discerning.

A number of SOC90 units are sparsely represented in the YCS dataset, which exposes the analyses to the risk of model over-fit when testing the influence of these measures. We therefore compared two other base units in our analyses. These were the 3-digit ‘minor’ groups of ISCO88, which comprise 99 different units in our analysis. The majority of ISCO88 minor groups might be thought to define distinctive occupational clusters (ILO, 1990). However, many ISCO88 minor groups are relatively heterogeneous in character, and the use of this baseline unit may occlude some detailed occupational differences.

In seeking to identify a sociologically informed fine-grained measure of occupational positions, we also operationalised a version of the 82 category ‘microclass’ scheme described by Jonsson et al. (2009).[1] We were not aware of a previous attempt to operationalise this scheme for UK data so constructed our own derivation algorithms for this purposes.[2] The mircoclass scheme is intended to differentiate detailed differences between occupations according to criteria of shared occupational cultures, associations and labour organisations. Nevertheless the scheme used by Jonsson et al. (2009) has acknowledged, pragmatic limits and features several heterogeneous categories, as well as being vulnerable to ambiguities in translation.

Agglomerate schemes

Defined on the basis of algorithms exploiting SOC90 codes and/or employment status data, there are numerous further occupation-based measures which can be linked to any given background measure. We call such measures ‘agglomerate’ schemes since they usually serve to represent occupational differences through a relatively small number of parameters. These measures have two forms. They either contain the categories of a ‘big class’ scheme, by which we mean a categorisation of occupations into a small range of large groups (cf. Weeden and Grusky, 2005). Or they are based on the scaling of occupations in a single dimension such as a ‘stratification position’, ‘socio-economic status’ or ‘prestige’. We label

[1] See also www.classmobility.org.
[2] The coding scheme we developed is made available at www.geode.stir.ac.uk. In a related project (www.camsis.stir.ac.uk/sonocs) we have also derived cross-walks from several other occupational unit group schemes to the microclass scheme of Jonsson et al. (2009). In the case of our derivation for the UK SOC90 scheme, we initially ran a translation macro from SOC90 to ISCO-88 (see OIU, 2001), then linked microclass schemes using ISCO-88 categories according to the macro generated by the SONOCS project, then undertook a manual review of the relationship between SOC90 codes and the allocated microclass units, and made a number of re-classifications according to our own judgement of the most suitable microclass for the unit group in hand.

such agglomerate schemes as ‘occupation-based social classifications’ (cf. Lambert and Bihagen, 2011).

Many different agglomerate measures could have been operationalised. In a fairly recent analysis Lambert and Bihagen (2007) compared 35 different measures. For simplicity we elected to compare twelve measures which include some of the most widely used and influential measures in social stratification research. These measures emerge from a range of theoretical frameworks and represent a variety of functional forms.

Analytical methods

Across the range of measures defined (see Table 2), we have several analytical possibilities open to us to describe empirical differences in outcomes. Both the base units and agglomerate schemes could be analysed in a descriptive way, exploring differences such as the average educational attainment across different categories, or by presenting summary statistics describing the overall scale of correlations between the educational and occupational measures. Descriptive comparisons could also be made to explore the internal differences within agglomerate schemes according to base units, such as assessing the scale of difference between SOC90 units within particular categories of an aggregate scheme. A few salient examples are presented in section 4.1.

Modelling strategies can also be used to review results, though the approach is made more complicated by the large number of categorical units featured in the three base units. Fortunately however, a range of general linear and mixed model formulations can now be readily estimated in statistical software. This allows analysts to take full advantage of the extended information available in fine-grained occupational data.

The analyses presented regress on a linear outcome measure of educational attainment (GCSE score). They then fit four different forms of statistical model, which are described algebraically in Table 3. The models comprise:

TABLE 3 ABOUT HERE

(1) Linear regression models with dummy variables for every one of either the base units or the categorical agglomerate schemes (or a single variable for the metric agglomerate schemes). This is tractable even for large numbers of base units, and is made easier by using Stata's 'areg' (Stata, 2009) estimators for this purpose.
(2) Linear regression models which feature measures of both the base units and the agglomerate measures. These analyses provide a crude test of whether or not there is added explanatory value in using the base units in addition to the agglomerate measures. The base unit effects can reasonably be regarded as a disaggregation within the agglomerate patterns.
(3) Random effects mixed models with clusters defined by the categories of either the base units or the categorical agglomerate schemes. This provides a different form of control for the differences from category to category, modelling the general structure of the difference in responses but not explicitly fitting a fixed parameter for each different category. These models, estimated using 'xtmixed' routines in Stata, provide an indication of the extent to which patterns of the response variables are differentiated in terms of their within group rather than their between group variation. In this paper we restrict analysis to 'random intercepts' random effects models which provide a basic control for the scale of 'shared variance' from category to category. We note that the estimation of more complex random effects models, allowing the modelling of complex variance structures at the level of occupational difference, could further be developed as a means of investigating how processes of occupational difference are related to other explanatory variables. For example we might be able to use random coefficients models to explore whether ethnic group influences on attainment were themselves differing in nature from occupation to occupation.
(4) Random effects mixed models with clusters defined by the categories of the base units plus fixed effects measures for the agglomerate units. This provides a second means of summarizing the additional explanatory power associated with taking account of base unit differences over and above the impact of agglomerate units.

4. Results

4.1Occupational differences

The overall message is that there is a relatively strong, and persistent, association between parental occupations and filial GCSE attainment. Young people from families in which parents have less advantaged occupations consistently perform less well at GCSE. The association is stronger than the associations for both gender and ethnicity. This is worthy of note given the widespread concerns in the UK about the gender gap in educational attainment.

The patterns of association between GCSE attainment and the agglomerate classes are substantively plausible. Descriptive analyses indicate that in the categorical agglomerate schemes there some categories in which the range of the levels of GSCE attainment for pupils is wide. This is illustrated in Figure 1. In NS-SEC category 2 *lower managerial and professional occupations*, there are large differences between the mean GCSE attainment for the children of secondary school teachers and children of publicans, even though they are in the same 'big class' category. Similarly, the children of other teachers (e.g. dance teachers) outperform their counterparts in NS-SEC category 3 *intermediate occupations*. These apparent differences motivate us to explore occupations in more detail.

FIGURE 1 ABOUT HERE

We are interested in identifying a sociologically informed fine-grained measure of occupational positions through a version of the 82 category 'microclass' scheme described by Jonsson et al. (2009). They assert that 'the microclass approach shares with the big-class model the presumption that contemporary labor markets are balkanized into discrete categories, but such balkanization is assumed to take principally the form of institutionalized occupations (e.g., doctor, plumber, postal clerk) rather than institutionalized big classes (e.g., routine nonmanuals, proprietors)' (Jonsson et al. 2009, pp.982-983).

It is possible that a microclass approach might be well suited to the analysis of the relationship between parental occupations and filial attainment. Jonsson et al. (2009) argue that microclasses might be more appropriate for understanding the mechanisms behind the transmission of human capital, cultural capital, social networks and economic resources. They further argue that the microclasses approach is fruitful in the examination of the transmission of occupation-specific skills, occupation-specific cultures, occupation-specific networks and of fixed resources.

We were not aware of a previous attempt to operationalise this scheme for UK data so for illustrative purposes we provide a table of descriptive statistics for the microclass (see Appendix 2). As we might anticipate the offspring of health professionals (e.g. doctors), and professors perform well in GCSE examinations. In contrast the children of launders and dry-cleaners, and fishermen have low levels of attainment. An aspect of the utility of the microclasses approach is illustrated in Figure 2. The children of other teachers (e.g. dance teachers) would ordinarily be placed in NS-SEC category 3 *intermediate occupations*. The mean GCSE score for NS-SEC category 3 is much lower than the mean for other teachers (e.g. dance teachers).

FIGURE 2 ABOUT HERE

In the microclass approach the offspring of other teachers would be included within a microclass that includes elementary and secondary school teachers and this appears to be a more appropriate classification of this occupation. It is plausible that the parents who are in other teaching jobs (e.g. dance teachers) are in a position to transmit occupation-specific knowledge that is not available to other parents in their ‘big class’, but that is similar to counterparts in the related occupations that form their microclass.

The descriptive differences from occupation to occupation are predictable and consistent with an emphasis on the impact of particular occupational skills and cultures. These preliminary patterns are encouraging and lead us to anticipate that there may be analytical benefits to employing a microclass approach to the analysis of filial attainment. We also tentatively speculate that a microclass approach might have applications in other research areas such as the study of youth transitions.

4.2 Model comparisons

Tables 4a and 4b summarise one illustrative example from the array of different statistical models estimated. In this example, the table shows the ten different models estimated for the conventional background measure, using the SOC90 baseline unit, and where the agglomerate measure is the RGSC categorical scheme. The models feature coefficient effects associated with each unit of analysis, and summary statistics which tell us about overall model fit. In this example, the standard regression model with dummy variables for categories of the agglomerate measure alongside controls is a relatively good fit (R^2 =.18), but an improvement to the model fit can be achieved by explicitly modelling occupational positions as either fixed or random effects. In this example we see modest, but not significant, differences in the coefficients of the control variables between models, suggesting robustness in this instance. More generally there is a possibility that interpretations of other results could be influenced by the way in which occupational data is controlled for.

The graphs supplied in Appendix 1 attempt to summarise aggregate findings on model fit from all of the models estimated (for the interested reader, the full array of aggregate statistical values arising from these models are available within the data file www.dames.org.uk/docs/conf_papers/rc28_2011_ycs/all_stats_2.dta). The twelve different graphs correspond to the permutations across the three different base units (SOC90, ISCO88 3-digit, and Microclass) and the four different background measures, conventional (cv), parental dominance (pd), fathers (pa) and mothers (ma).

The graphs summarize a number of different model results. The bars show the 'standard' statistical results obtained from using agglomerate measures as either random or fixed effects respectively. A high R^2 value and a high value of Rho could be regarded as an indication of the quality of an agglomerate measure. The gaps between the bars and the lines, however, show us the volume of information about occupations which is potentially being missed across the different possible agglomerate measures (the gap represents the difference between the R^2 and Rho statistics using the more detailed base units compared with the agglomerate units). The circle plots indicate the residual occupational clustering which persists after controlling through dummy variables (or linear parameters) for the agglomerate measures. That these values are consistently well above zero reinforces the idea that no agglomerate measure ever fully accounts for all the baseline information.

The figures can be examined on the one hand to review differences between agglomerate measures within a particular baseline measure and base unit. An interesting story of moderate differences between schemes emerges. Approximately, across the different figures, we see typical variation by about as much as one third to one quarter of the total R^2, and as much as half of the intra-cluster correlation (Rho), from different agglomerate schemes. This is clearly evidence of differences between the measures in which empirical patterns they identify. On the other hand, the general impact of all the different measures is broadly the same, and the differences in fit statistics do not seem strongly aligned with theoretical differences between the schemes. We interpret these patterns in a similar way to the conclusions drawn in an earlier paper (Lambert and Bihagen, 2007). This leads us to suspect that differences between agglomerate schemes are often largely unimportant to sociological conclusions. We further suspect that they cannot therefore be used to adjudicate between different theoretical mechanisms, but that the choice of scheme could be influential to any more precise interpretations of model coefficients that may be considered. We envisage that policy-oriented claims about small changes in the patterns of background influence over time, which are common in the UK, may be particularly vulnerable in this respect. Accordingly, in methodological terms, such differences between the properties of agglomerate schemes make a strong case that a sensitivity analysis comparing different measurement tools should routinely feature in any analysis of stratification effects.

In addition, looking across the figures, it is possible to compare between baseline measures and between base units. Consistently we see that SOC90 base units capture more information than the ISCO-88 3-digit and microclass units. This pattern is unsurprising since the latter units involve fewer categories, and could indeed reflect some instances of model over-fit. From a different perspective, both the microclass and ISCO-88 3-digit base units also come reasonably close to capturing similar levels of variance as the SOC90 base units, and provide a clear statistical gain compared with the agglomerate schemes. The organisation of SOC90 units can reasonably be viewed as part substantive and part administrative. Whereas the other base units, and in particular the microclass base unit, have tried to define the differences between categories on consistent criteria these might be regarded as favourable base units of analysis. On balance we suspect that there are relevant differences in occupational units (some of which may perhaps be unique to the UK) which are captured by SOC90 but not by the other base units. Therefore for some empirical investigations SOC90 is probably the most favourable of the three alternatives considered.

When comparing between different background measures, we also see relatively little difference in overall model properties. The explanatory value of the conventional measure is generally highest (with the highest R^2 for the agglomerate measure and base units), which might suggest that this approach remains a valid empirical strategy despite criticisms which have emerged on both theoretical and political grounds. There is however little difference between the conventional, dominance and fathers measures. On the other hand, using the mother's occupation alone seems to be problematic for this analysis since it tends to lead to lower model fit statistics (suggesting in general a lower level of empirical explanation). This may be unsurprising given that occupational gender segregation in Britain in general tends to mean less precise differentiation in occupation-based measures for women than for men. This may suggest that parental occupation-based measures are as problematic for measuring the social position of mothers as they are for women in general.

5. Discussion

We conclude that there is a relatively strong, and persistent, association between parental occupations and filial GCSE attainment. These effects are greater than the effect of both gender and ethnicity. This relationship is observed irrespective of however GCSE attainment and parental occupations are measured. The relationship is empirically stronger when the data on parental occupations recognises specific and detailed differences between jobs. For example the difference between school teachers and publicans, which in many UK socio-economic schemes are assigned to the same classification.

Amongst the range of measures used, fine-grained occupation-based schemes such as the original SOC90 units or the microclass scheme bring a substantial, parsimonious improvement to the empirical description of parental influences upon filial educational attainment. The differences from occupation to occupation seem predictable and consistent with an emphasis on the impact of occupation-specific influences. The results presented favour the arguments developed by Jonsson et al. (2009) and Grusky and Weeden (2006; 2005; 2002). Our analyses convince us that microclass approaches may be applicable more generally in stratification research.

Nevertheless a large proportion of the social origins influence is adequately measured by quite simplified occupation-based schemes. This seems to suggest that much, but not all, of the influence of parental occupations is a direct function of average position within the stratification structure. In these analyses of filial educational attainment the divergence between schemes is minimal. This does not support the claim that different stratification measures measure different aspects of stratification (cf. Rose and Harrison 2010; Lambert and Bihagen 2007).

In the UK, much contemporary social policy interest is directed to the influence of family background on young people's educational attainment. Numerous strategies are under consideration intended to improve the attainment of those from relatively disadvantaged backgrounds. Conclusions have overwhelmingly been based upon analysis which used 'big class' schemes or other simplified measures of parental background. First, our results suggest that such research is likely to detect the bulk of the empirical patterns of difference in a reasonably robust way. It would be misleading however to attribute the differences observed to any more specific causal mechanisms than the general influence of stratification position. Second, our results indicate that a further substantial component of background influence is not well captured by such measures. Heterogeneity within 'big' categories could obviate the intended policy impacts of targeted measures, yet reasonably tractable measures of occupational background could well provide sufficient tools to derive more focussed targeted policies.

References

Chan, T. W., & Goldthorpe, J. H. (2007). Class and Status: The Conceptual Distinction and its Empirical Relevance. *American Sociological Review, 72*, 512-532.

Croxford, L., Iannelli, C., & Shapira, M. (2007). Youth Cohort Time Series for England, Wales and Scotland, 1984-2002. ESRC Grant R000239852. Swindon: ESRC.

Erikson, R. (1984). Social Class of Men, Women and Families. *Sociology - The Journal of the British Sociological Association, 18*(4), 500-514.

Goldthorpe, J. H. (2002). Occupational Sociology, Yes: Class Analysis, No: Comment on Grusky and Weeden's Research Agenda. *Acta Sociologica, 45*(3), 211-217.

Goldthorpe, J. H. (2007). *On Sociology, Second Edition: Volume Two - Illustration and Retrospect*. Stanford: Stanford University Press.

Hildebrand, D. K., Laing, J. D., & Rosenthal, H. (1977). *Analysis of Ordinal Data*. Thousand Oaks, California: Sage.

ILO. (1990). *ISCO-88 : International Standard Classification of Occupations*. New York: International Labour Office.

Jonsson, J. O., Grusky, D. B., Di Carlo, M., Pollak, R., & Brinton, M. C. (2009). Microclass Mobility: Social Reproduction in Four Countries. *American Journal of Sociology, 114*(4), 977-1036.

Lambert, P. S., & Bihagen, E. (2007). *Concepts and Measures: Empirical evidence on the interpretation of ESeC and other occupation-based social classifications*. Paper presented at the International Sociological Association, Research Committee 28 on Social Stratification and Mobility, Montreal (14-17 August).

Lambert, P. S., & Bihagen, E. (2011). *Stratification research and occupation-based social classifications*. Stirling: University of Stirling, Technical Paper 2011-1 of the Data Management through e-Social Science Research Node (www.dames.org.uk).

Lambert, P. S., Tan, K. L. L., Turner, K. J., Gayle, V., Prandy, K., & Sinnott, R. O. (2007). Data Curation Standards and Social Science Occupational Information Resources. *International Journal of Digital Curation, 2*(1), 73-91.

OIU. (2001). *OOSS User Guide 01.1 : Mapping of Standard Occupational Classification 1990 (SOC1990) to International Standard Classification of Occupations European Version (ISCO-88COM)*. Titchfield, Fareham, Hampshire: Occupational Information Unit, Office for National Statistics.

OPCS. (1990). *Standard Occupational Classification, Volume 1: Structure and definition of major, minor and unit groups*. London: Office for Population Censuses and Surveys.

Rose, D.,& Harrison, E. (2010). *Social Class in Europe: An Introduction to the European Socio-economic Classification.* London: Routledge.

StataCorp. (2009). Stata Statistical Software, Release 10.1. College Station, TX: StataCorp LP.

Weeden, K. A., & Grusky, D. B. (2005). The Case for a New Class Map. *American Journal of Sociology, 111*(1), 141-212.

Data Citation

Youth Cohort Time Series for England, Wales and Scotland, 1984-2002 [computer file]. First Edition, Colchester, Essex: UK Data Archive [distributor], November 2007. SN 5765.

Tables and figures referred to in the text

Table 1: Descriptive statistics for original variables

	#categories	mean	min	max	Valid n
GCSE points score	85	39.0	0	84	55120
Respondent's cohort (year when aged 16)	5	1994.7	1990	1999	55120
Respondent is male	2	0.459	0	1	55120
Respondent ethnicity: Black	2	0.0138	0	1	55120
Respondent ethnicity: Indian	2	0.0236	0	1	55120
Respondent ethnicity: Pakistani	2	0.0130	0	1	55120
Respondent ethnicity: Bangladeshi	2	0.0040	0	1	55120
Respondent ethnicity: Other Asian	2	0.0095	0	1	55120
Respondent ethnicity: Other	2	0.0089	0	1	55120
Father's SOC-90 (pa)	366		100	999	49863
Mother's SOC-90 (ma)	312		100	999	45679
Conventional measure of SOC-90 (cv)	370		100	999	55120
Parental dominance SOC-90 (pd)	369		100	999	55120
Father part-time	2	0.107	0	1	49435
Mother part-time	2	0.347	0	1	44786
Father self-employed	2	0.240	0	1	48828
Mother self-employed	2	0.093	0	1	44172

Source: Youth cohort study of England and Wales, cohorts of youths aged 16 in 1990, 1993, 1995, 1997 and 1999, Harmonised data from Croxford et al. (2007). Unweighted analysis, sample of those reporting at least one valid parental occupational detail and valid GCSE attainment data.

Table 2: Descriptive statistics for selected occupation-based measures

Abbrev.	Description	#units	% missing (if any)	R^2 GCSE Score
	Background measure=Parental dominance			
soc_pd	SOC-90 unit	369		16.9
micro_pd	Microclass unit	81	0.4	14.7
isco88_pd	ISCO-88 minor group	102		14.7
mcam_pd	UK CAMSIS (male scale)	(scale)		13.3
fcam_pd	UK CAMSIS (females scale)	(scale)		12.2
isei_pd	International Socio-Economic Index	(scale)		12.2
siops_pd	Standard International Occupatonal Prestige Scale	(scale)		11.6
nes_pd	New Earnings Survey (average income)	(scale)		11.8
rgsc_pd	Registrar General's Social Class	6	0.3	11.0
egp11_pd	Erikson-Goldthorpe 11-category scheme	11		11.7
ns8_pd	National Statsitics Socio-Economic Scheme	8		11.4
ns3_pd	``(3 category version of NS-SEC)	3		10.1
esec_pd	European Socio-economic Classification	9	0.4	10.7
skill_pd	Skill classification using Elias (1993)	4		9.3
mnm_pd	Manual-Non-manual classification	2		7.8
	Background measure=Conventional			
soc_cv	SOC-90 unit	370		15.1
micro_cv	Microclass unit	81	0.5	13.0
isco88_cv	ISCO-88 minor group	102		13.1
	Background measure=Father's occupation			
soc_pa	SOC-90 unit	366	9.5	15.0
micro_pa	Microclass unit	81	10.1	13.0
isco88_pa	ISCO-88 minor group	102	9.5	13.2
	Background measure= Mother's occupation			
soc_ma	SOC-90 unit	312	17.1	14.4
micro_ma	Microclass unit	79	17.5	12.3
isco88_ma	ISCO-88 minor group	99	17.1	12.3

Source: As table 1. Agglomerate measure summaries from other background measures are not included.

Table 3: Summary of models used in the regression analyses

Model	Stata 10 estimation routines and example formulation
(1) $\mathbf{Y_i = \beta X_i + \gamma O_i + \varepsilon_i}$	xi: regress gcse fem black i.class areg gcse fem black, absorb(microclass)
(2) $\mathbf{Y_i = \beta X_i + \gamma A_i + \zeta B_i + \varepsilon_i}$	xi: areg gcse fem black i.class, absorb(microclass)
(3) $\mathbf{Y_{io} = \beta X_{io} + \mu_o + \varepsilon_{io}}$	xtmixed gcse fem black ||microclass: ,
(4) $\mathbf{Y_{ib} = \beta X_{ib} + \gamma A_{ib} + \mu_b + \varepsilon_{ib}}$	xi: xtmixed gcse fem black i.class ||microclass: ,

Y = GCSE attainment; X = other controls; ε = individual level error; μ = error component modelled at occupational level; i indexes different respondents; b indexes different base units; o indexes different agglomerate units or base units; B represents dummy variables for base units; A represents dummy variables or scale parameters for agglomerate units; O represents dummy variables for either base or agglomerate units; β, γ and ζ represent vectors of model coefficients. Estimation examples show 'fem' and 'black' as control variables, 'class' as agglomerate scheme and 'microclass' as base unit.

Table 4a: Example models (random effects) estimating GSCE score

	(3) agg	**(3) agg**	**(3) base**	**(3) base**	**(4) base**
	(model format as Table 3, models (3) fits random effect to either agglomerate or base units)				
Cohort 1993		5.29***		5.12***	5.14***
Cohort 1995		9.72***		9.51***	9.53***
Cohort 1997		8.41***		8.18***	8.19***
Cohort 1999		12.8***		12.7***	12.7***
Boys		-3.36***		-3.45***	-3.45***
Black		-4.17***		-4.23***	-4.19***
Indian		1.72***		2.55***	2.57***
Pakistani		-4.02***		-2.93***	-2.86***
Bangladeshi		-1.38		-1.16	-0.878
Other Asian		4.17***		4.69***	4.77***
Other		-0.704		-0.43	-0.407
RGSC II					-4.49***
RGCS III(N)					-7.39***
RGSC III(M)					-14***
RGSC IV					-15.7***
RGSC V					-18***
Intercept	39.4***	33.8***	39.3***	34***	43.9***
Level 2 variance	6.29***	6.17***	6.89***	6.75***	3.25***
Level 1 variance	15.95***	15.33***	15.64***	15.03***	15.03***
Rho	0.28	0.29	0.31	0.31	0.18
Log-likelihood	-207079	-205004	-206398	-204321	-204130
n	49373	49373	49373	49373	49373

Table 4b: Example models (OLS) estimating GSCE score

	(1) agg	(1) agg	(1) base	(1) base	(2)
	(model format as Table 3, models (2) fits dummy variable effects to either agglomerate or base units)				
RGSC II	-3.88***	-3.6***			-0.927
RGSC III(N)	-6.67***	-6.39***			0
RGSC III(M)	-13.9***	-13.3***			12.5
RGSC IV	-14.7***	-14.4***			11.5
RGSC V	-17.5***	-16.9***			22.8
Cohort 1993		5.29***		5.12***	5.12***
Cohort 1995		9.72***		9.51***	9.51***
Cohort 1997		8.41***		8.17***	8.17***
Cohort 1999		12.8***		12.7***	12.7***
Boys		-3.36***		-3.45***	-3.45***
Black		-4.17***		-4.26***	-4.26***
Indian		1.72***		2.65***	2.64***
Pakistani		-4.01***		-2.73***	-2.74***
Bangladeshi		-1.38		-0.957	-0.985
Other Asian		4.17***		4.77***	4.76***
Other		-0.704		-0.431	-0.444
Intercept	48.8***	43***	39.6***	34.2***	27.8***
Log-likelihood	-207055	-204981	-205768	-203684	-203683
bic	414175	410145	411546	407497	407538
n	49373	49373	49373	49373	49373
R^2	.106	0.178	0.151	0.22	0.22

Figures referred to in the text

Figure 1

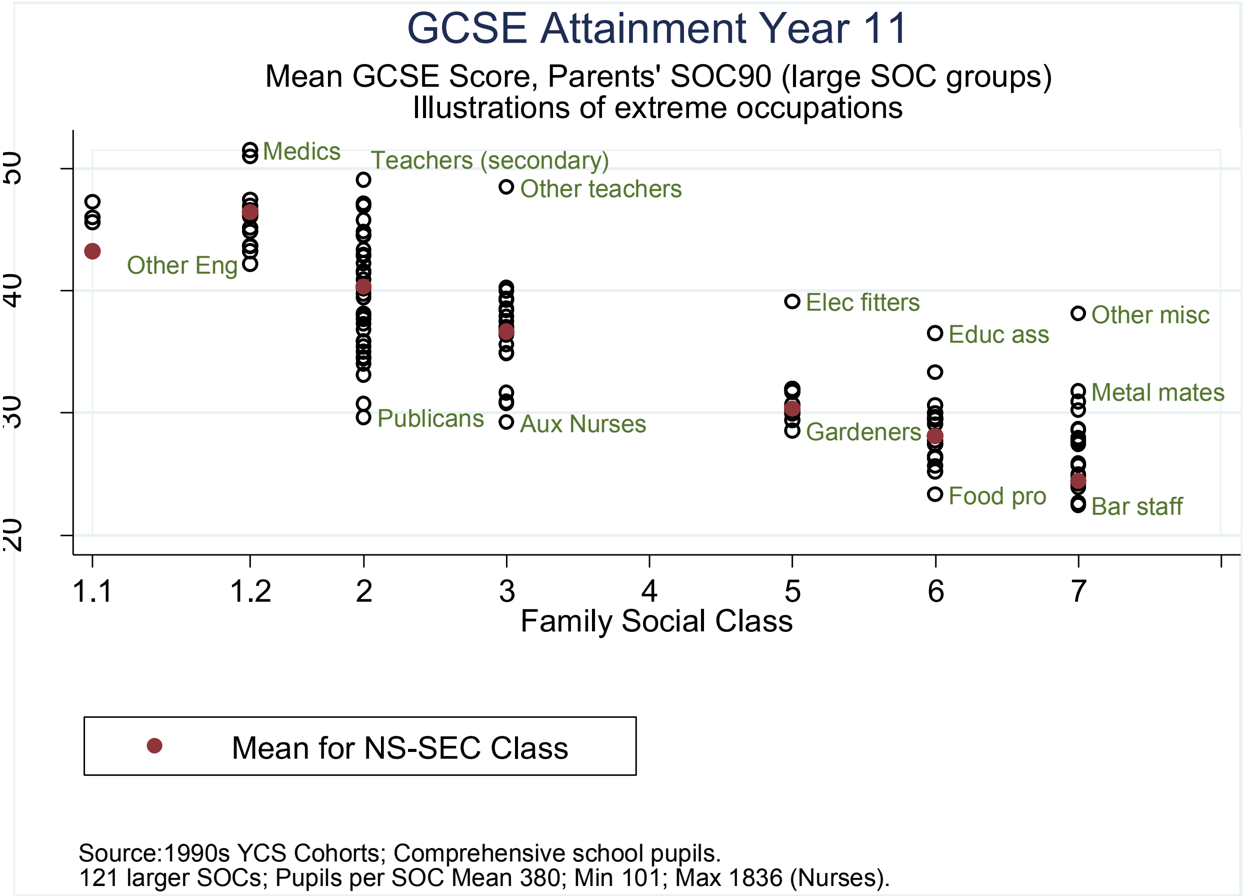
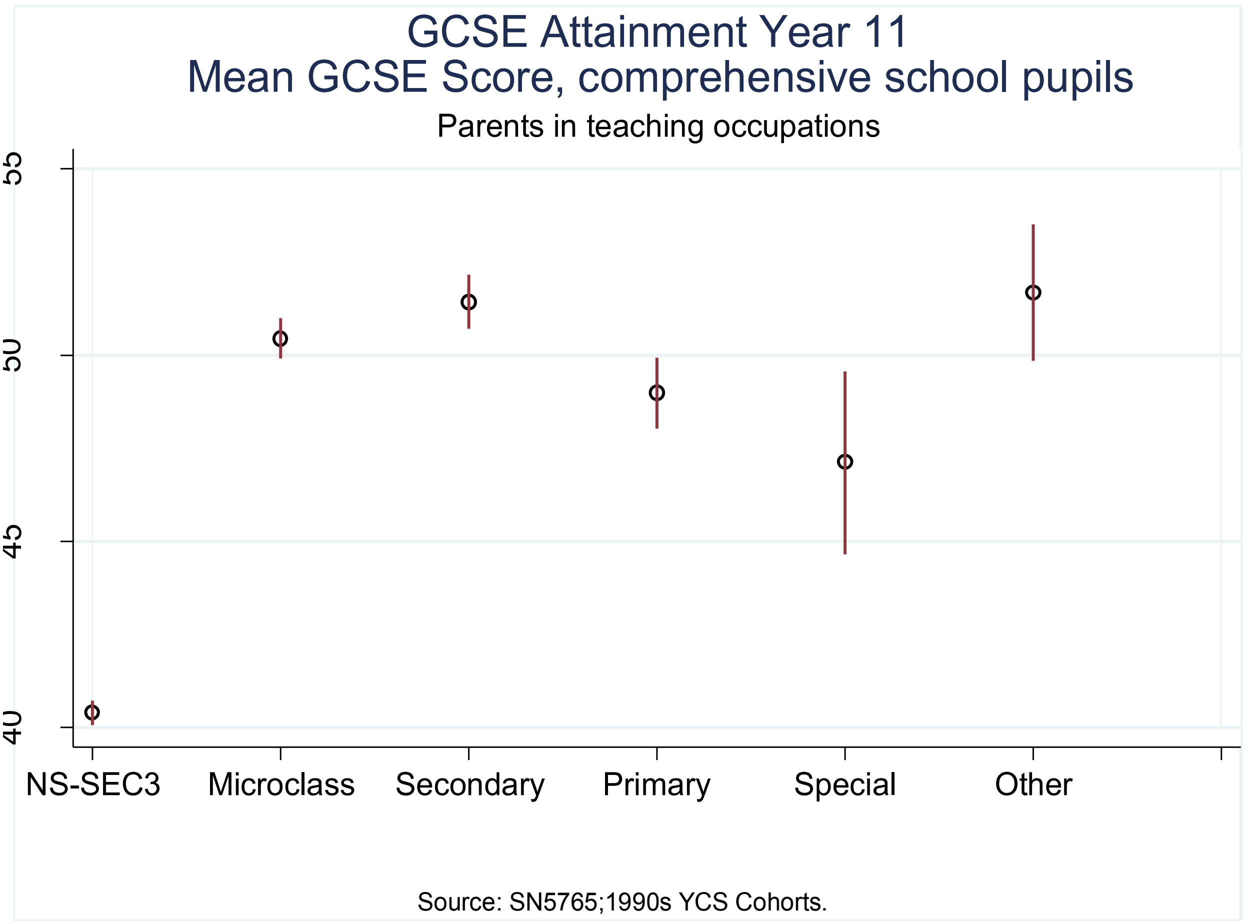
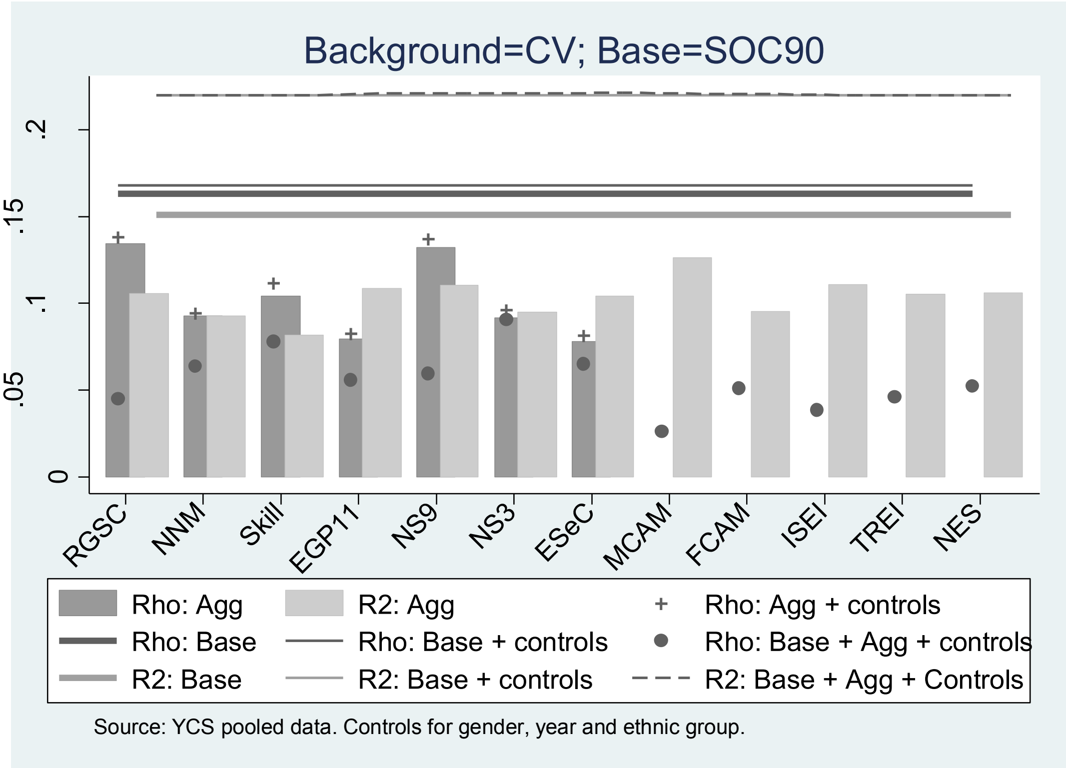
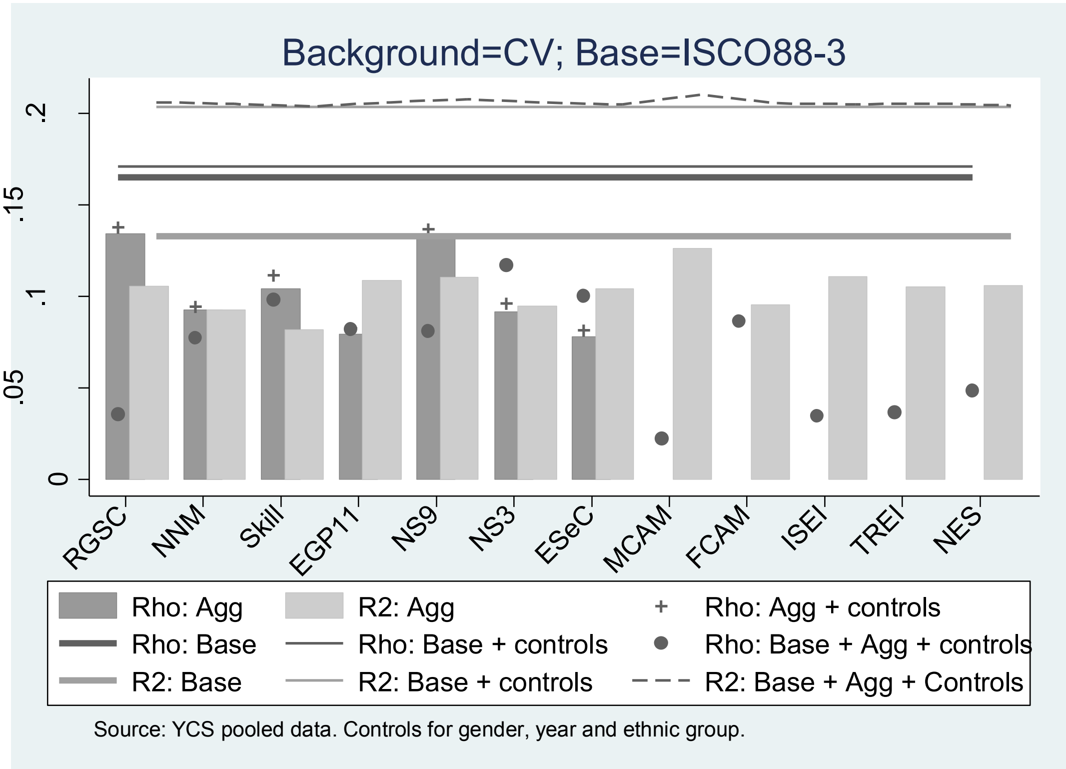


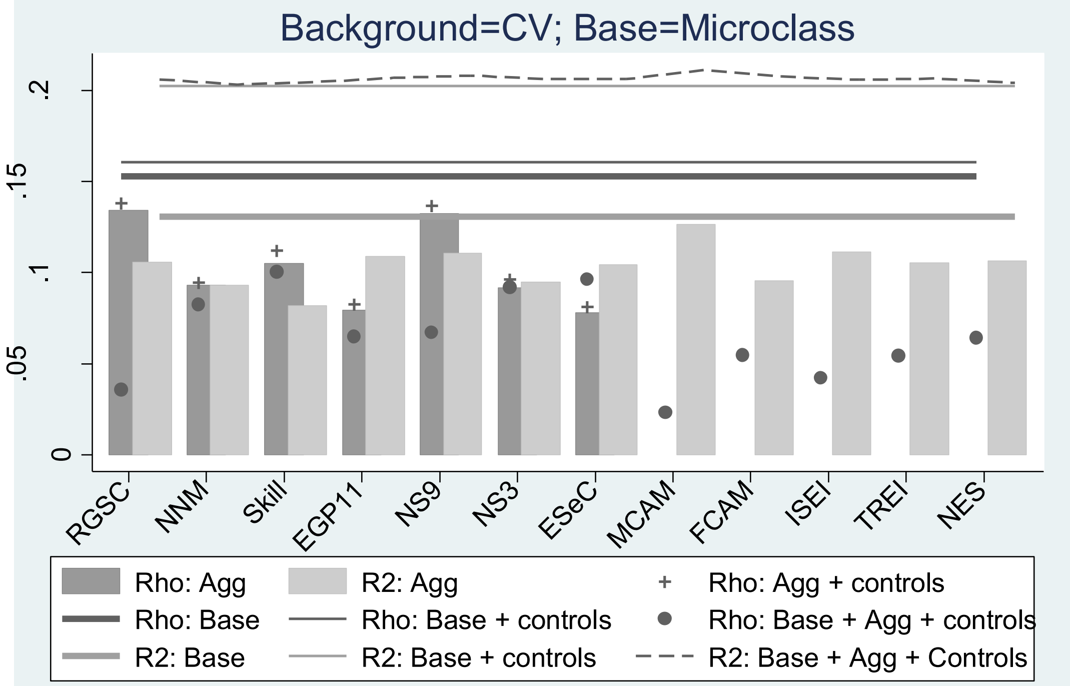
Figure 2


GCSE Attainment Year 11
Mean GCSE Score, comprehensive school pupils
Parents in teaching occupations
40
45
50
55
NS-SEC3
Microclass
Secondary
Primary
Special
Other
Source: SN5765;1990s YCS Cohorts.

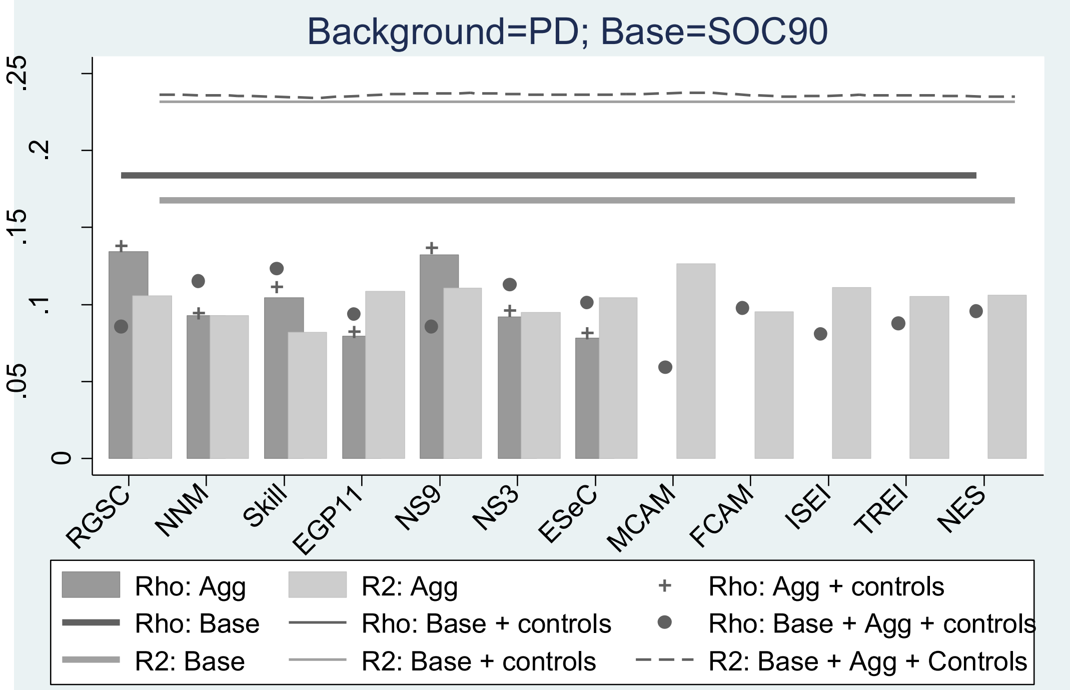
Appendix 1


Background=CV; Base=SOC90
0
.05
.1
.15
.2
RGSC
NNM
Skill
EGP11
NS9
NS3
ESeC
MCAM
FCAM
ISEI
TREI
NES
Rho: Agg
R2: Agg
Rho: Agg + controls
Rho: Base
Rho: Base + controls
Rho: Base + Agg + controls
R2: Base
R2: Base + controls
R2: Base + Agg + Controls
Source: YCS pooled data. Controls for gender, year and ethnic group.

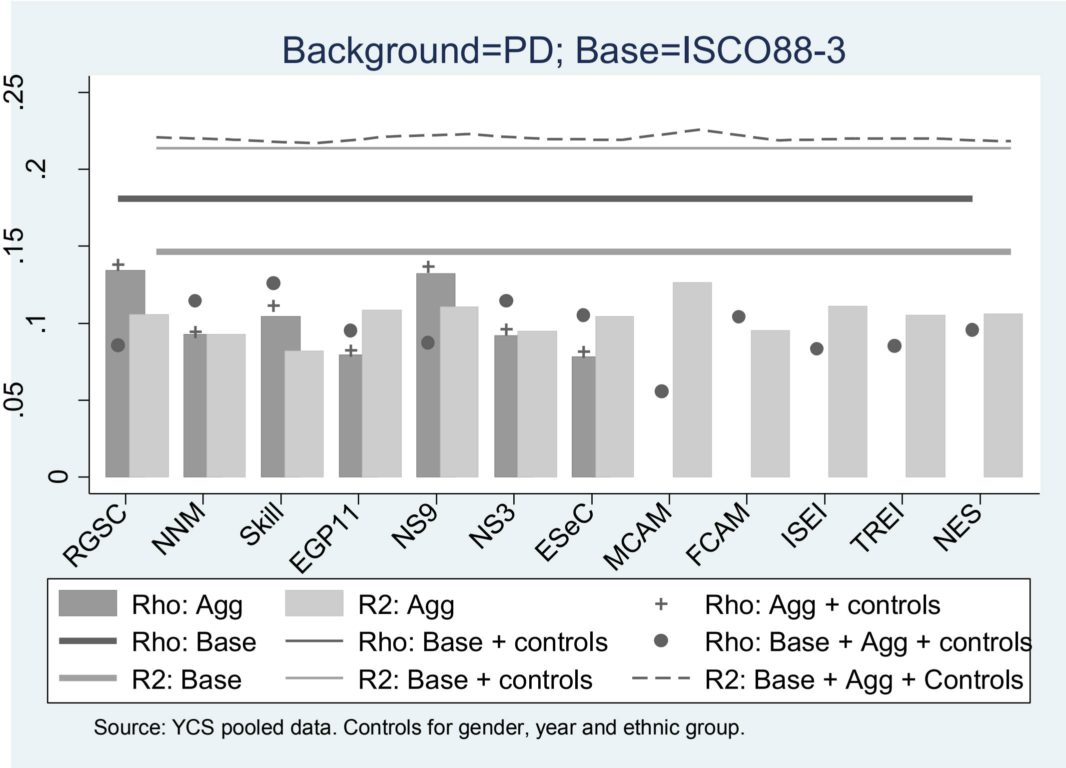

Background=CV; Base=ISCO88-3
0
.05
.1
.15
.2
RGSC
NNM
Skill
EGP11
NS9
NS3
ESeC
MCAM
FCAM
ISEI
TREI
NES
Rho: Agg
R2: Agg
Rho: Agg + controls
Rho: Base
Rho: Base + controls
Rho: Base + Agg + controls
R2: Base
R2: Base + controls
R2: Base + Agg + Controls
Source: YCS pooled data. Controls for gender, year and ethnic group.



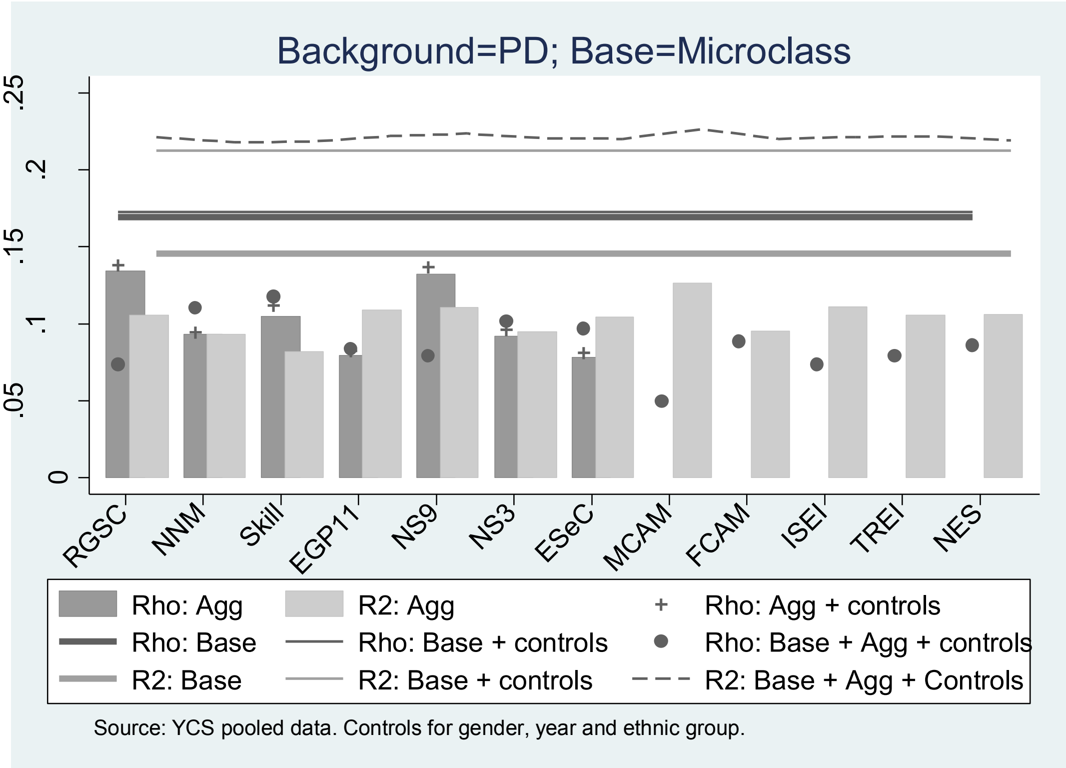
Source: YCS pooled data. Controls for gender, year and ethnic group.



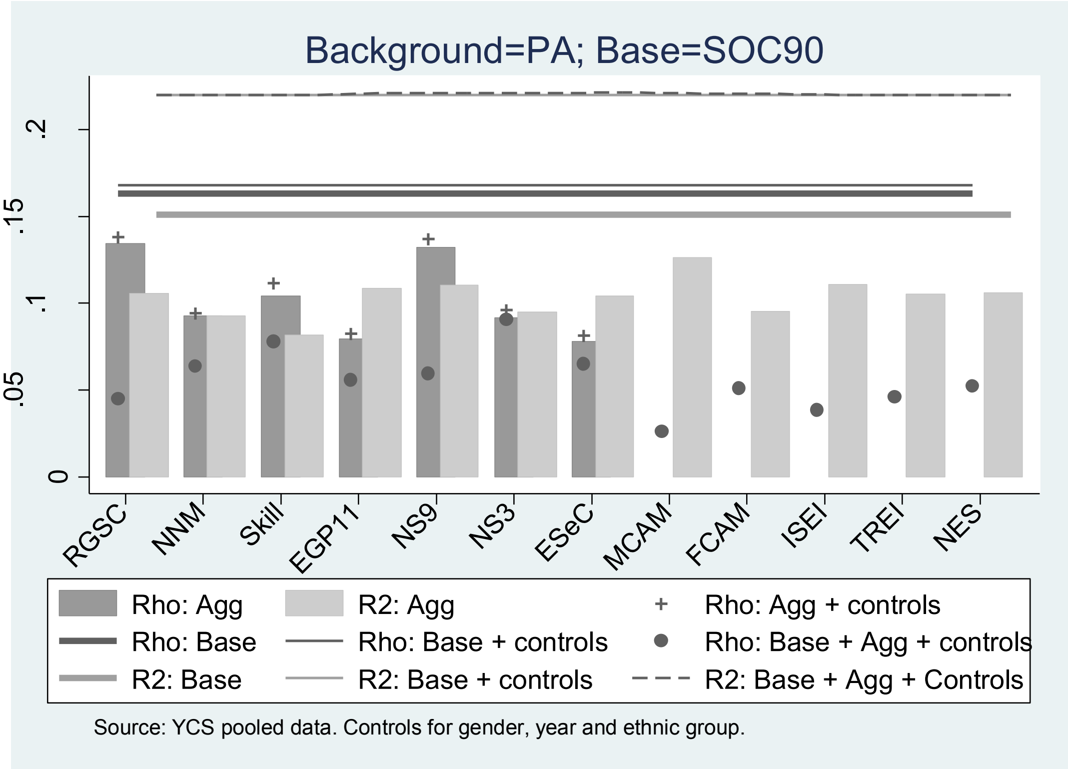
Source: YCS pooled data. Controls for gender, year and ethnic group.

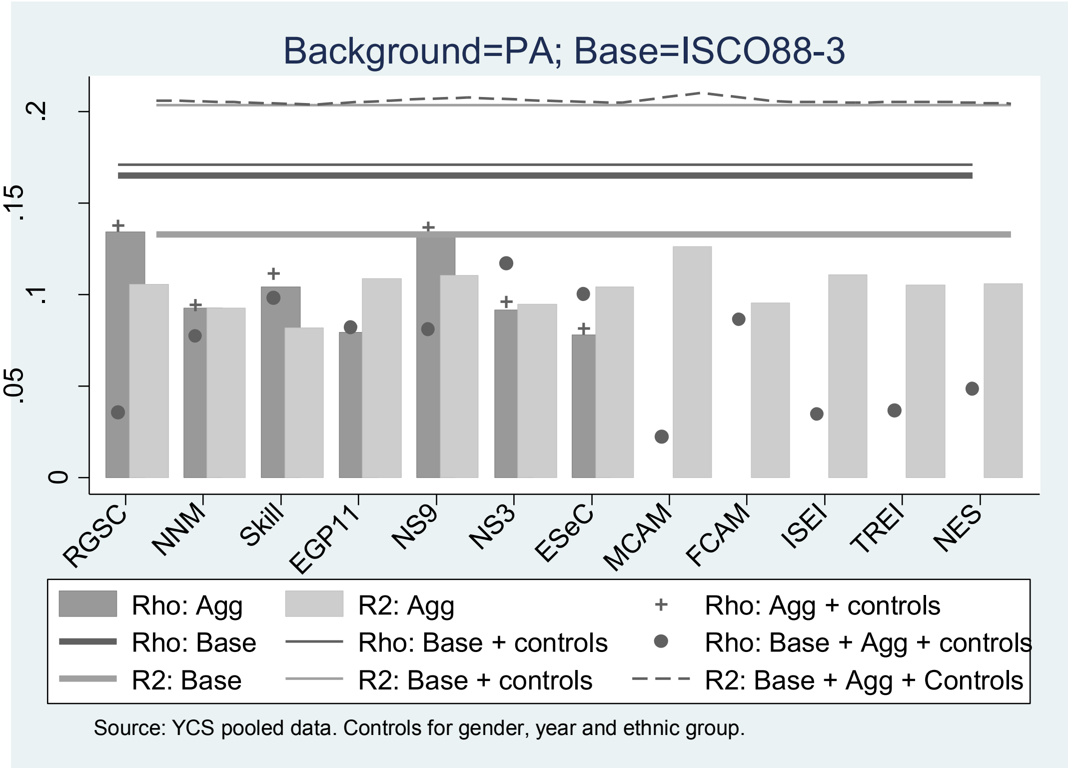


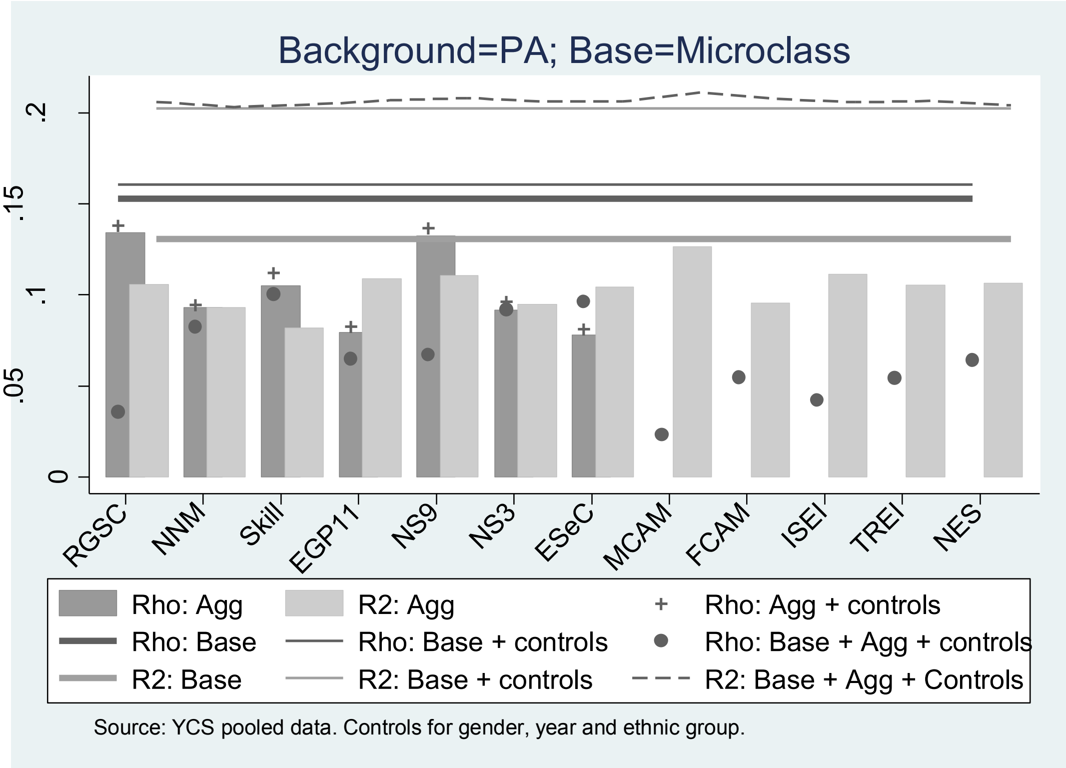
Background=PD; Base=ISCO88-3
0
.05
.1
.15
.2
.25
RGSC
NNM
Skill
EGP11
NS9
NS3
ESeC
MCAM
FCAM
ISEI
TREI
NES
Rho: Agg
R2: Agg
Rho: Agg + controls
Rho: Base
Rho: Base + controls
Rho: Base + Agg + controls
R2: Base
R2: Base + controls
R2: Base + Agg + Controls
Source: YCS pooled data. Controls for gender, year and ethnic group.

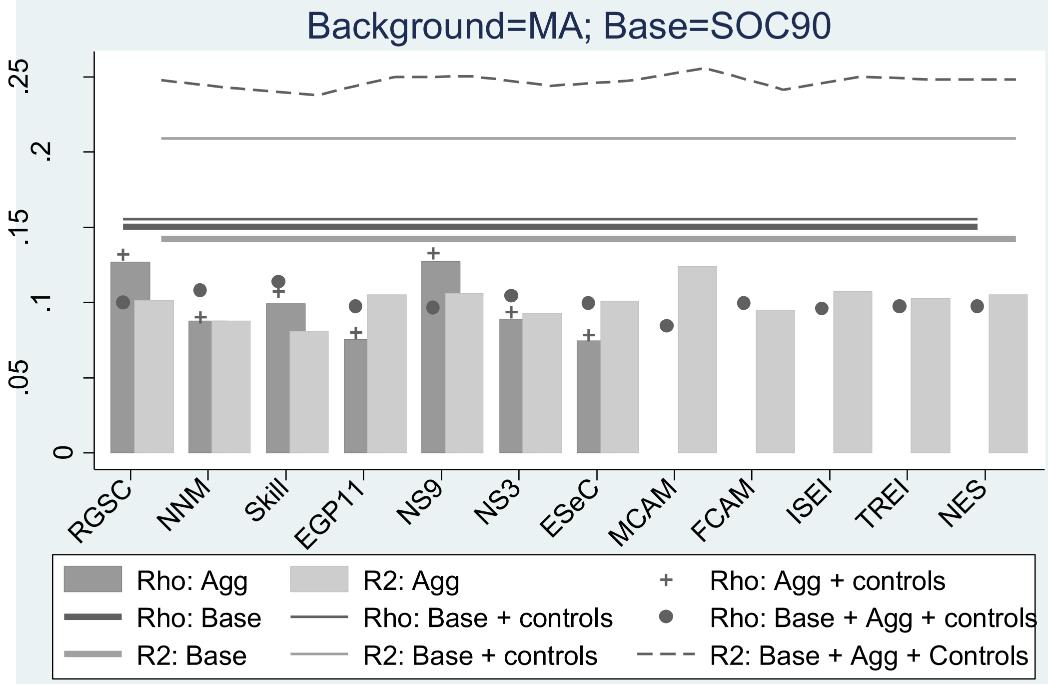


Background=PD; Base=Microclass
0
.05
.1
.15
.2
.25
RGSC
NNM
Skill
EGP11
NS9
NS3
ESeC
MCAM
FCAM
ISEI
TREI
NES
Rho: Agg
R2: Agg
Rho: Agg + controls
Rho: Base
Rho: Base + controls
Rho: Base + Agg + controls
R2: Base
R2: Base + controls
R2: Base + Agg + Controls
Source: YCS pooled data. Controls for gender, year and ethnic group.

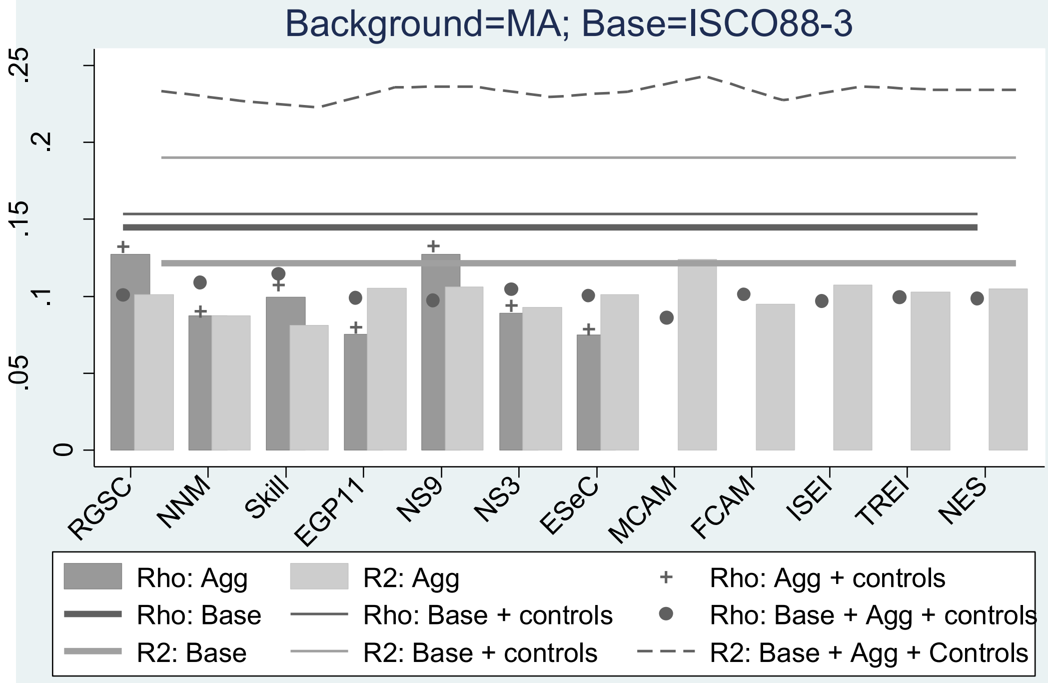

Background=PA; Base=SOC90
0
.05
.1
.15
.2
RGSC
NNM
Skill
EGP11
NS9
NS3
ESeC
MCAM
FCAM
ISEI
TREI
NES
Rho: Agg
R2: Agg
Rho: Agg + controls
Rho: Base
Rho: Base + controls
Rho: Base + Agg + controls
R2: Base
R2: Base + controls
R2: Base + Agg + Controls
Source: YCS pooled data. Controls for gender, year and ethnic group.


Background=PA; Base=ISCO88-3
0
.05
.1
.15
.2
RGSC
NNM
Skill
EGP11
NS9
NS3
ESeC
MCAM
FCAM
ISEI
TREI
NES
Rho: Agg
R2: Agg
Rho: Agg + controls
Rho: Base
Rho: Base + controls
Rho: Base + Agg + controls
R2: Base
R2: Base + controls
R2: Base + Agg + Controls
Source: YCS pooled data. Controls for gender, year and ethnic group.

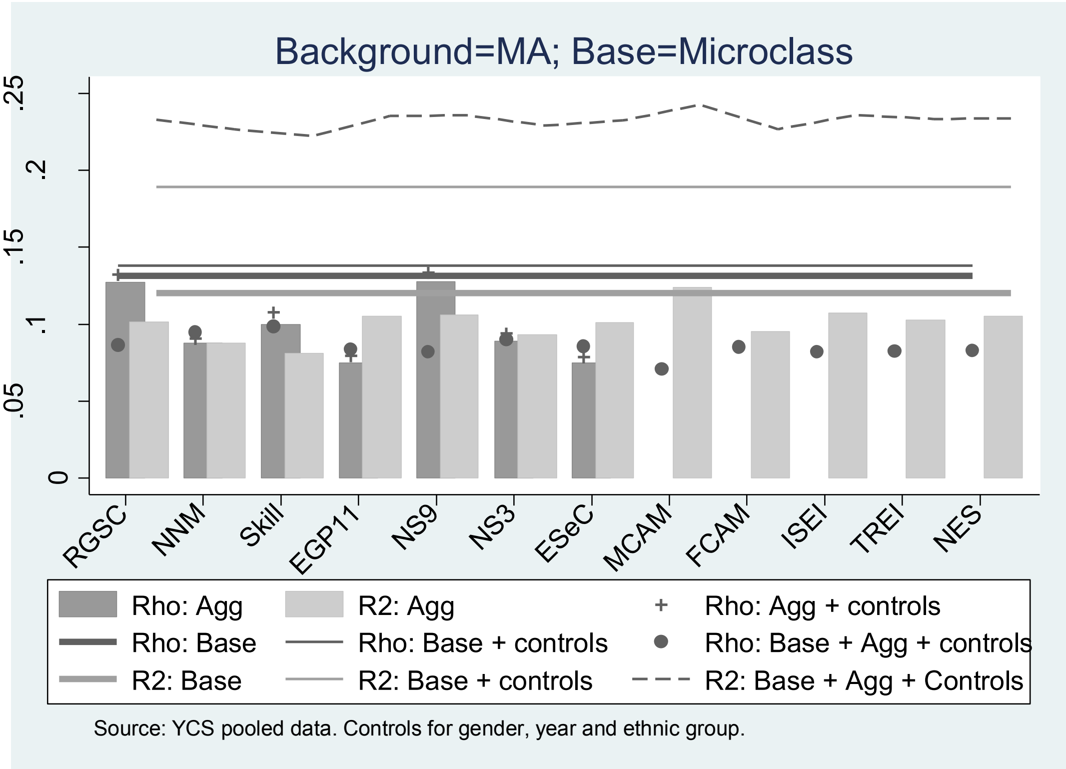

Background=PA; Base=Microclass
0
.05
.1
.15
.2
RGSC
NNM
Skill
EGP11
NS9
NS3
ESeC
MCAM
FCAM
ISEI
TREI
NES
Rho: Agg
R2: Agg
Rho: Agg + controls
Rho: Base
Rho: Base + controls
Rho: Base + Agg + controls
R2: Base
R2: Base + controls
R2: Base + Agg + Controls
Source: YCS pooled data. Controls for gender, year and ethnic group.



Source: YCS pooled data. Controls for gender, year and ethnic group.



Source: YCS pooled data. Controls for gender, year and ethnic group.



Background=MA; Base=Microclass
0
.05
.1
.15
.2
.25
RGSC
NNM
Skill
EGP11
NS9
NS3
ESeC
MCAM
FCAM
ISEI
TREI
NES
Rho: Agg
R2: Agg
Rho: Agg + controls
Rho: Base
Rho: Base + controls
Rho: Base + Agg + controls
R2: Base
R2: Base + controls
R2: Base + Agg + Controls
Source: YCS pooled data. Controls for gender, year and ethnic group.

Appendix 2

Microclasses (parental dominance) mean GSCE scores

Microclass	Mean GCSE score	S.E. Mean GCSE score	Median GCSE score	n
Launderers and dry-cleaners	25.37	2.23	25.00	43
Fishermen	26.00	2.88	23.00	29
Housekeeping workers	26.93	0.58	26.00	859
Miners and related workers	27.28	1.20	27.00	183
Longshoremen and freight handlers	28.47	1.29	29.00	139
Heavy machine operators	28.49	0.80	28.00	445
Textile workers	28.92	1.52	30.00	127
Truck drivers	29.28	0.51	29.50	1062
Food processors	29.31	3.70	29.00	13
Sawyers and lumber inspectors	29.39	1.67	29.00	84
Bakers	29.80	1.63	31.50	86
Craftsmen and kindred workers	30.56	1.32	29.00	179
Food service workers	30.70	0.50	30.00	1157
Mass transportation operators	31.03	0.61	31.00	774
Operatives and kindred workers	31.16	0.41	32.00	1732
Forestry workers	31.22	3.08	31.00	27
Tailors and related workers	31.53	0.98	33.00	302
Painters	31.55	0.69	33.00	581
Hospital attendants	31.64	0.53	33.00	1016
Butchers	31.91	1.30	32.50	162
Guards and watchmen	32.11	0.87	34.00	431
Service workers	32.16	1.23	33.00	210
Janitors and cleaners	32.23	1.11	32.00	217
Metal processors	32.35	0.79	33.00	480
Bricklayers, carpenters, construction workers	32.58	0.31	34.00	2940
Locomotive operators	32.58	1.93	34.00	65
Stationary engine operators	32.70	2.23	33.00	66
Farm labourers	32.71	1.27	32.00	154
Sales workers and shop assistants	32.74	0.45	34.00	1303
Vehicle mechanics	32.78	0.61	33.00	679
Transport conductors	33.21	1.60	35.00	114
Gardeners	33.35	1.13	35.00	249
Plumbers and pipe-fitters	33.49	0.68	33.00	614
Welders and related metal workers	33.50	0.69	35.00	572
Chemical processors	33.89	0.87	33.50	386
Cabinetmakers	34.06	1.83	36.00	83
Postal and mail distribution clerks	34.22	0.88	35.00	334
Printers and related workers	34.65	0.77	35.00	376

Newsboys and deliverymen	34.70	1.94	33.00	83

***(continued)* Microclasses (parental dominance) mean GSCE scores**

Microclass	Mean GCSE score	S.E. Mean GCSE score	Median GCSE score	n
Other mechanics	35.22	1.57	36.00	98
Hairdressers	35.98	1.45	37.00	101
Blacksmiths and machinists	36.02	0.44	37.00	1363
Proprietors	36.07	1.07	37.00	211
Members of armed forces	36.38	1.19	38.00	182
Telephone operators	36.74	1.82	38.00	77
Nursery school teachers and aides	36.89	0.60	39.00	765
Farmers and farm managers	38.16	0.71	39.00	542
Building managers and proprietors	38.35	0.63	39.00	669
Cashiers	38.68	0.60	40.00	754
Social and welfare workers	38.83	0.65	40.00	698
Electronics service and repair workers	38.99	0.39	40.00	1658
Office and clerical workers	39.31	0.23	40.00	4723
Health semiprofessionals	39.34	0.37	41.00	1898
Other agents	40.81	0.43	42.00	1270
Protective service workers	41.47	0.52	43.00	880
Bookkeepers and related workers	41.48	0.46	43.00	1159
Managers	41.54	0.22	43.00	5374
Professional, technical, and related workers	41.68	0.57	43.00	745
Jewelers, opticians, and precious metal workers	42.14	1.92	44.00	86
Creative artists	42.18	1.05	44.00	240
Nonmedical technicians	43.57	0.60	45.00	608
Real estate agents	43.73	1.52	45.00	105
Ship officers	43.86	2.07	45.00	59
Insurance agents	45.08	0.76	48.00	340
Aircraft pilots and navigators	45.53	1.48	48.00	77
Engineers	45.84	0.37	48.00	1643
Commercial Managers	45.86	0.29	47.00	2658
Journalists, authors, and related writers	46.81	1.14	48.00	161
Personnel and labor relations workers	47.18	0.69	48.50	448
Accountants	47.35	0.64	49.00	525
Officials, government and non-profit organizations	48.05	0.76	49.00	396
Architects	48.54	0.64	50.00	471
Librarians	48.81	1.29	50.00	111
Workers in religion	49.01	1.13	52.00	180
Systems analysts and programmers	49.11	0.46	51.00	948
Jurists	49.81	1.03	51.00	219
Elementary and secondary school teachers	50.45	0.28	52.00	2521

Natural scientists	51.12	0.78	53.00	299

(continued) **Microclasses (parental dominance) mean GSCE scores**

Microclass	**Mean GCSE score**	**S.E. Mean GCSE score**	**Median GCSE score**	**n**
Professors and instructors	51.13	0.52	53.00	738
Statistical and social scientists	51.18	1.55	54.00	83
Health professionals	52.96	0.69	55.00	426