

Growing Up in Scotland: Early experiences of Primary School

TECHNICAL APPENDIX

Paul Bradshaw, Julia Hall, Tessa Hill, Judith Mabelis and Dan Philo

CONTENTS

1	MULTIVARIATE ANALYSIS	3
1.1	Description of analysis undertaken.....	3
1.2	Interpreting the regression results	3
1.3	Regression tables.....	4
2	EDUCATION ADMINISTRATIVE DATA	6

1 MULTIVARIATE ANALYSIS

1.1 Description of analysis undertaken

Many of the factors we are interested in are related to each other as well as being related to the outcome variables of interest – in this case, low parental involvement in school-related activities, being very satisfied with the child's school and perceived school readiness. For example, younger mothers are more likely to have lower qualifications, to be lone parents, and to live in areas of high deprivation. Simple analysis may identify a relationship between maternal age and parental involvement. However, this relationship may be occurring because of the underlying association between maternal age and education. Thus, it is actually the lower education levels amongst younger mothers which is associated with a greater likelihood of lower involvement in their child's school/education rather than the fact that they are younger in age.

To take these possible confounds into account, multivariate regression analysis was used. This analysis allows the examination of the relationships between an outcome variable and multiple explanatory variables whilst controlling for the inter-relationships between each of the explanatory variables. This means it is possible to identify an *independent* relationship between any single explanatory variable and the outcome variable; to show, for example, that there is a relationship between maternal age and child's cognitive ability that does not simply occur because both education and maternal age are related.

The multivariate analysis undertaken to explore factors associated with low parental involvement and very high parental satisfaction employed stepwise logistic regression. Each model incorporated various socio-demographics and various 'school-related factors' (shown in Tables 1.1 and 1.2). Stepwise regression assesses each variable for significance, entering the most significant variable first and adjusting significance based on variables already entered into the equation, so that the final equation contains only those variables that remain significant when other variables are entered into the model. Backwards stepwise (beginning with all variables in the model and removing one by one those variables that were not significant) was used to check that the same variables were obtained for the final model.

Logistic regression, without a step-wise approach, was used for the multivariate analysis of characteristics related to having an average or higher than average perceived school readiness score. A single model was compiled incorporating a wide range of predictor variables (shown in Table 1.3).

1.2 Interpreting the regression results

The results of the regression analyses are presented in Tables 1.1 to 1.3, showing odds ratios for the models (final models for the stepwise analysis), together with the probability that the association is statistically significant. The predictor variable is significantly associated with the outcome variable if $p < 0.05$. The models show the odds of being in the particular category of the outcome variable (e.g. having average or higher than average school readiness) for each category of the independent variable (e.g. parental education categories). Odds are expressed relative to a reference category, which has a given value of 1. Odds ratios greater than 1 indicate higher odds, and odds ratios less than 1 indicate lower odds. Also shown are the 95% confidence intervals for the odds ratios. Where the interval does not include 1, this category is significantly different from the reference category.

To understand an odds ratio we first need to describe the meaning of odds. The definition of odds is similar but significantly different to that of probability. This is best explained in the form of an example. If 200 individuals out of a population of 1000 experienced persistent poverty, the probability (p) of experiencing persistent poverty is 200/1000, thus $p=0.2$. The probability of not experiencing persistent poverty is therefore $1-p = 0.8$. The odds of experiencing persistent poverty are calculated as the quotient of these two mutually exclusive events. So, the odds in favour of experiencing persistent poverty to not experiencing persistent poverty, is therefore $0.2/0.8=0.25$. Suppose that 150 out of 300 people living in social rented housing experience persistent poverty compared to 50 out of 150 who live in owner occupied housing. The odds of a person living in social rented housing of experiencing persistent poverty are $0.5/0.5=1.0$. The odds of a person living in owner occupied housing of experiencing persistent poverty is $0.3333/0.6666=0.5$. The odds ratio of experiencing persistent poverty is the ratio of these odds, $1.0/0.5=2.0$. Thus the odds of experiencing persistent poverty are twice as high among people who live in social rented housing (compared to people who live in owner occupied housing – the ‘reference category’).

1.3 Regression tables

Table 1.1 Factors independently associated with low parental involvement (involvement in one activity or no involvement): full results of stepwise logistic regression

	Odds ratio	P-value	95% confidence interval	
Household socio-economic classification (ref: Managerial/professional)				
Intermediate	1.30	0.18	0.88	1.93
Small employers and own account workers	1.26	0.44	0.70	2.26
Lower supervisory and technical	1.90	0.02	1.13	3.19
Semi-routine and routine	1.56	0.06	0.98	2.49
Never worked	4.36	0.00	1.77	10.75
Parental level of education (ref: Degree level qualifications)				
No qualifications	2.60	0.02	1.19	5.70
Lower SGs or other	2.01	0.02	1.14	3.54
Upper SGs or Int VQs	1.96	0.00	1.25	3.08
Higher Gs or upper level VQs	1.38	0.12	0.91	2.07
Tenure (ref: Owner-occupied)				
Social rented, private rented or other	1.55	0.01	1.10	2.17

Table 1.2 Factors independently associated with being ‘very satisfied’ with the child’s school: full results of stepwise logistic regression

	Odds ratio	P-value	95% confidence interval	
Confidence in helping with homework (ref: Confident in all subjects or tasks) Confident in some but not in others, or not confident at all	0.55	0.00	0.39	0.78
Has received information from school on how to help child’s learning (ref: Yes) No	0.66	0.00	0.55	0.79
How useful found parents’ evening (ref: Very useful) Quite useful	0.31	0.00	0.24	0.39
Not very useful or not at all useful	0.16	0.00	0.10	0.25
Has not attended a parents’ evening or a parents’ evening has not taken place yet	0.56	0.03	0.33	0.95
Has spoken to any teachers about how their child is doing at school outside of parents evenings or similar events (ref: Yes) No	1.56	0.00	1.25	1.95
How easy was it or would it be to approach teacher (ref: Very easy) Quite easy	0.26	0.00	0.20	0.34
Not very or not at all easy	0.15	0.00	0.07	0.31
Family type (ref: Lone parent) Couple family	1.49	0.02	1.08	2.05
National statistics socio-economic classification (ref: Managerial and professional occupations) Intermediate occupations	0.84	0.22	0.64	1.11
Small employers and own account workers	1.13	0.59	0.72	1.76
Lower supervisory and technical occupations	1.42	0.10	0.93	2.17
Semi-routine and routine occupations	1.65	0.01	1.17	2.33
Never worked	0.74	0.43	0.35	1.57

Table 1.3 Factors independently associated with having an average or above average school readiness score: full results of logistic regression

	Odds ratio	P-value	95% confidence interval	
Child's gender (ref: male)				
Female	1.09	0.30	0.93	1.28
Child's age at school entry (ref: between 5 and 5.5)				
Under 5	0.61	0.00	0.51	0.72
Over 5.5	0.68	0.01	0.52	0.90
Household equivalised income (ref: lowest income quintile)				
2nd Quintile (>=£11,875<£19,444)	0.82	0.19	0.60	1.11
3rd Quintile (>=£19,444< £25,625)	0.95	0.74	0.70	1.29
4th Quintile (>=£25,625< £37,500)	1.14	0.37	0.86	1.51
Top Quintile (>=£37,500)	1.32	0.10	0.95	1.83
Parental level of education (ref: no qualifications)				
Lower SGs or VQs or 'Other' quals	0.62	0.20	0.30	1.30
Upper level SGs or Intmed VQs	1.19	0.51	0.70	2.01
Higher Grades or Upper level VQs	1.36	0.26	0.79	2.34
Degree level academic or VQs	1.30	0.33	0.76	2.24
Pre-school type (ref: nursery class attached to school)				
Local Authority nursery school	0.92	0.54	0.70	1.20
Private nursery school	0.90	0.46	0.67	1.20
Other provider	0.87	0.61	0.51	1.50
Weekly duration of pre-school (ref: 12 or 12.5)				
Less than 12 hours of pre-school	0.74	0.04	0.55	0.99
Between 12.5 and 15 hours	0.74	0.03	0.57	0.97
15 hours or more	0.84	0.22	0.63	1.11
Perception of school readiness score (ref: average or above)				
Below average score	0.26	0.00	0.22	0.31
Score on Strengths and Difficulties Scale (ref: Normal)				
Moderate	0.69	0.09	0.44	1.06
Severe	0.47	0.00	0.30	0.72
Vocabulary ability (ref: Below average)				
Average or above	1.23	0.04	1.01	1.49
Problem solving ability (ref: Below average)				
Average or above.	1.19	0.05	1.00	1.42

2 EDUCATION ADMINISTRATIVE DATA

The report incorporates analysis of education administrative data drawn from the ScotXed databases held by the Scottish Government Education Directorate. Permission to link survey information with this administrative data was obtained from parents at sweep 6 of the study. Of the 3657 parents interviewed at sweep 6, 97% (n=3534) gave permission to obtain data on their child from ScotXed. Of these, a successful match between GUS details and education records was made for 3433 children (97% of those who consented).

Matching of school-level data was undertaken using the school's unique 'SEED code' – a numeric identifier used by ScotXed. This code was already available for the vast majority of the GUS sample members as it is included on the school look-up table used when the child's school details are being collected during the survey interview.

Matching of pupil level data was undertaken using four variables taken from the GUS dataset – child gender, date of birth, SEED code and child's home postcode. ScotXed do not hold the child's name so this could not be used for individual level matching purposes.

The process of matching was undertaken in six stages. Details of each stage including the number of cases requiring matching and the proportion successfully matched are included in Table 2.1. The first stage looked for unique matches for all four variables – taken from the sweep 6 dataset - with ScotXed's pupil level census data from September 2010 (all children in GUS were eligible to have started school by August 2010 though some had started by August 2009). This stage produced 2992 successful matches. The remaining cases were then matched, using the same four variables but this time taken from the sweep 5 data, against the pupil level census data from September 2009. This resulted in a further 42 successful matches. To allow for children who had moved home between the GUS data collection and the pupil census – thus having different postcodes on each database – the next stage involved matching on the basis of gender, date of birth and SEED code only against both the 2010 and 2009 census datasets. These stages produced an additional 268 matches. Finally, to allow for children having moved school – thus having different SEED codes on each database – the match was run using only gender, date of birth and postcode resulting in a further 63 matches.

In total, 3365 GUS cohort members were successfully identified in the ScotXed database, representing 95% of those who provided consent and 92% of all cases achieved at sweep 6.

Table 2.1 Breakdown and results of process through which GUS cohort members were identified in ScotXed administrative records

Stage description	No. requiring match	No. matched	% of cases with consent matched at this stage	Cumulative % matched
Match Sw6 {gender / DoB / School ID / home postcode} with Census 2010	3534	2992	85%	85%
Match Sw5 {gender / DoB / School ID / home postcode} with Census 2009	542	42	1%	86%
Match Sw6 {gender / DoB / School ID} with Census 2010	500	263	7%	93%
Match Sw5 {gender / DoB / School ID} with Census 2009	237	5	<1%	93%
Match Sw6 {gender / DoB / home postcode} with Census 2010	232	47	1%	94%
Match Sw5 {gender / DoB / home postcode} with Census 2009	185	16	<1%	95%
Remaining cases	169	-	-	-