

Family Background and Socioeconomic Status Effects on Educational Performances by Data Mining Methods: A Case Study in Iran

Seyed Bagher Mirashrafi¹, Ebrahim Khodaie², Ehsan Jamali²

¹Karlsruhe Institute of Technology (KIT), Germany and University of Mazandran (UMZ)

²National Organization of Educational Testing (NOET)
Iran

Abstract

This paper analyzes the effects of different aspects of family background such as family income, parental education, parental job, family size, and the socioeconomic status of parents on educational performance of higher education applicants in Iran. The gathered data from the Iranian university and college applications in 2010 is analyzed using data mining methods to investigate the effects of family background variables on entrance examination grades. The results of this analysis show that parental education, parental job, and the socioeconomic status of family have large effects on entrance examination grades, and as a result on university and college acceptance and also on educational performances. Results of this study provide a better understanding of the factors that could have significant impact on the applicant's admission, and as a result on the higher education achievement in Iran. Such insights can assist responsible people to make good decision about the relative advantages and limitations for higher education attainment.

1. Introduction

The Iranian university and college admission system requires prospective college and university students ranking up to one hundred majors in order of their preference after receiving their examination grade. The National Organization of Educational Testing (NOET) processes applications for all different kinds of university admissions at all levels.

The nationwide exam is held in five groups: Mathematics and Physics, Empirical Sciences, Human Sciences, Art and Foreign Language. Typically, seven to ten subjects are examined in each group. Table 1 shows these examined subjects. Four general subjects are common in all groups: Farsi Literature, Arabic Language, Islamic Literature, and Foreign Language.

Once assessed, for each subject, the examination results are used to produce a score between -33 and 100 per candidate. The structure of exams is based on multiple choices and every three wrong answers

are considered as one negative point. In other words, if a candidate has three wrong answers and one correct answer, his or her mark is equal to zero.

Table 1. Names of specialized subjects in each group

Testing Group	Examined Subjects
Mathematics and Physics	Mathematics, Physics, Chemistry
Empirical Sciences	Mathematics, Physics, Chemistry, Biology, Geology
Human Sciences	Mathematics, Economics, History and Geography, Social Sciences, Philosophy and Logic, Psychology
Art	Mathematics, Art Information, Technical Drawing, Music, Picture and Imagination ability, Play Skills
Foreign Language	Specialized Language

Subsequent to examination grading, the NOET fixes a total mark for each candidate following a certain process. The points requirements are set so that applicants are offered a place in the highest preference major for which they are eligible; in the case of candidates being tied for the last position in a field, both are offered a place. It should be emphasized that candidates do not know the field point's requirement prior to completing their application or taking their examination. The point's requirement is influenced by the examination results of candidates who applied for each major and by the number of available positions in that major.

Some fields have minimum entry standards. For example, sufficient knowledge of mathematics may be required for an engineering level. A few fields in the Art group have interviews, but these are not common. The candidates who apply within the first three groups are also allowed to apply for the Art and Foreign Language groups, but not vice versa.

2. Literature Review

After a brief review of the Iranian university and college admission system, the impact of family background on applicants' examination result is discussed. This paper is a contribution to general debate of educational attainment. Social scientists have developed sophisticated models for educational attainment using different causal variables [1, 5]. Although, there are variations regarding race and sex e.g. [6, 13], the same causal variables have been applied generally. Considerable variation in educational outcomes is not explained by the basic attainment model [3, 9]. Past research has indicated an academic achievement gap between the sexes, with boys ahead of girls. However, more recent study has shown that the achievement gap has been narrowing and that in some instances girls have higher academic achievement than boys [2]. Additionally, studies show that girls perform better in reading than boys. However, males are found to outperform females in mathematics and science [4]. Research has found that socioeconomic status, parental involvement, and family size are particularly important factors [14]. Lochner & Belley [12] find that post-secondary (PS) attendance (and attendance at four-year PS institutions) is strongly positively related to parental income in the U.S., even after controlling for similar measures of family background and adolescent cognitive achievement. The effect of parental income PS attendance relationship in Canada is also positive, but substantially weaker. The findings of Pedrosa et al. [18] indicate that students coming from a disadvantaged environment, in socioeconomic and educational terms, perform relatively better than those coming from higher socioeconomic and educational strata. More interestingly, from an educational public policy viewpoint, students who came from public schools had a relatively better performance than those who had studied at private schools. On the other hand, the expansion of universities caused some social changes in middle and low class categories. Hence, acceptance in the universities became one of the possible ways of having a better job and other economic opportunities. Getting a university degree may increase the chance of a change of social class from a low class to a high class [11]. Therefore, as mentioned above in the literature, there is a variety of theories on the effects of family background in educational attainment. For example, Khodaie [10] shows that parental education has positive effects on the children's success, from an educational point of view. Moreover, Jamali shows the increase of parent's educations level and father job's causes the increase of educational performances [7, 8]. Additionally, all of these results in this study supported as well as the Mirashrafi et al. studies [15, 16, 17]. Sacker and her colleagues [19] set out to test the model shown in Figure 1. They set

out to examine how inequalities in educational achievement and adjustment come about. It has been well known for decades that pupils' educational achievement is related to parents' social class yet the mechanisms that form this relationship are not well understood. How does social class influence school achievement?

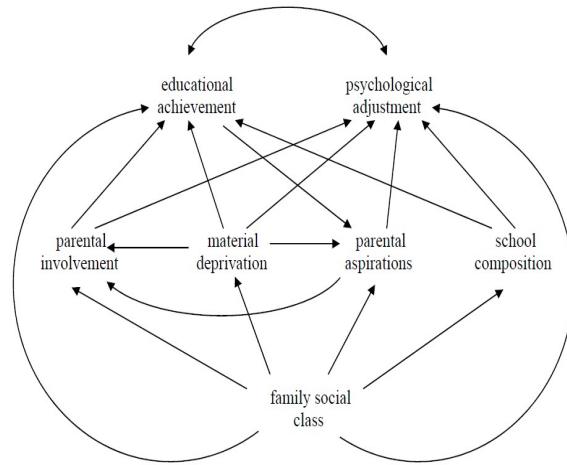


Figure 1. Sacker et al. model of the relationship between family social class, and pupil achievement and adjustment

Furthermore, the growth of the population rate and decline in the mortality rate at the same time, have led to an increased demand for education. Therefore, governments invest more in higher education. This situation happened in Iran after the Islamic Revolution of 1979. Because of some unsuccessful policies, the growth of the population rate went up sharply and, simultaneously, the demands for higher education increased considerably. Hence, the Iranian government started to expand universities to deal with the high demand for higher education.

3. Methodology

This study deals with two main questions. First, does family background such as size of family, parental education, family's income and parental job have any effects on Iranian universities admission? Second, is there any difference between percentages of becoming admitted among different social classes in Iran? In this study we consider acceptance in universities, total grade and total rank as dependent variables and size of family, parental education, family's income, parental job, gender and age of participants as independent variables. As a result, we use data mining methods to answer those two questions. For univariate analysis of variance, descriptive of the full sample and socioeconomic status of low, mid, and high levels were performed. The descriptive factors included min, max, mean, median and standard deviation. Furthermore, analysis

of variance (ANOVA) was run across the low, mid, and high socioeconomic status levels for each family background factor.

4. Data and Descriptive Statistics

The dataset, which is used in this analysis, contains two parts. Part one consists of applicants' specifications such as gender, testing group, age, total grade and acceptance as well as information about their application in 2010 from the NOET's original data file. The second part of data is extracted from a questionnaire with six questions regarding applicants' family background, which is filled by all applicants during the nationwide examination. These two files are merged into one dataset according to the applicants' ID. Table 2 shows a description of variables used in the analysis.

Table 2. Description of variables used in analysis

Variable Name	Values	Measure
Age	Min	14.00
	Max	75.00
	Mean	20.48
	Median	19.00
	Std. Dev.	3.71
Gender	Female	62.2%
	Male	37.8%
Testing Group	Mathematics	22.8%
	Empirical Sciences	32.0%
	Human Sciences	32.1%
	Art	3.6%
	Foreign Language	9.5%
Total Grade	Min	52.00
	Max	13816.00
	Mean	5783.65
	Median	5551.00
	Std. Dev.	1557.85
Acceptance	No	48.1%
	Yes	51.9%
Father's Education	No Education	13.5%
	Primary School	32.5%
	High School	34.8%
	University Degree	19.2%
Mother's Education	No Education	21.6%
	Primary School	37.0%
	High School	31.0%
	University Degree	10.4%
Father's Job	Workless or Other job	14.4%
	Private Sector Employee	53.8%
	Government Employee	25.7%
	Teacher or Lecturer	6.2%
Mother's Job	Housewife	88.8%
	Private Sector Employee	2.9%
	Government Employee	3.8%
	Teacher or Lecturer	3.8%
Family's Income (yearly in USD)	Weak (<4200)	24.9%
	Average (4200-5400)	30.8%
	Good (5400-8700)	29.5%
	Very Good (>8700)	14.8%
Family Size	Three Persons	2.8%
	Four Persons	19.0%
	Five Persons	22.4%
	Six Persons or More	55.8%

The used data file is the total number applicants and is not a sample. The variables in this study allow a reasonable replication of the educational attainment models, most commonly estimated. (The categorization of variables such as family's income, size of family, parental education and parental job were coded in the NOET).

In 2010 the total number of applicants in the Iranian nationwide university entrance examination as 1,099,320 persons. Foreign Language and Art groups were floating groups and candidates could choose one of main groups (Mathematics and Physics, Empirical Sciences, Human Sciences) as well as either or both of these two groups. Therefore, after removing duplicated candidates 955,633 actual candidates were used in our analysis. Table 3 shows the number and percentages of candidates and entrants by each group. It can be seen from this table that from 955,633 candidates 496,211 persons were accepted in universities; of which 34.8% are from group one, 27.8% are from group two, 37.4% are from group three. Moreover, Table 3 shows the chance of entrance for main groups one, two and three are 68.8%, 39.2% and 52.6% respectively. It shows that the group one candidate has a higher chance for entrance to university according to the other group candidates.

Table 3. The number and percentage of candidates and entrants by main testing group

Group:	Math. & phys.		Empirical sciences		Human sciences		Total	
	No.	%	No.	%	No.	%	No.	%
Candidates	250,940	26.3	351,486	36.8	353,208	37.0	955,633	100
Entrants	172,608	34.8	137,802	27.8	185,801	37.4	496,211	100
Chance		68.8		39.2		52.6		51.9

In the following section, we use the total grade of each candidate as a dependent variable and calculate cross-tabulations between the total grade and family background factors.

4.1. Parental Education

Table 4 and 5 compare total grade of applicants according to their parental education levels and the gender. The percentages of candidates according to the their father education in four categories of university degree holders, high school certificate holders, primary school certificate holders and with no educations are 13.5%, 32.5%, 34.8% and 19.2% respectively. As can be seen from Table 4 and Figure 3, the father's education has a positive effect on the total grades of candidates. With an increase in the father's level of education, the total grades of applicants increase. For example, the means of total grades in four categories are 6244.6, 5651.4, 5473.6, and 5284.9. This pattern is the same for both sexes and the only difference is that the mean of total

grades for females is slightly higher than males. Table 5, shows the mean of total grades of applicants regarding their mother's education and the gender. In this table we have the same pattern as in Table 4. Hence, the education of mother has a positive effect on the mean of total grades of applicants. In other words, the higher level of parental education is, in almost every case, associated with a higher mean of total grades of male and female applicants. The likelihood of entering the university is, in particular, much higher for the children of university-educated parents compared to the children of parents with any other educational level.

Table 4. The mean of total grades of applicants by father's education and the gender of candidates

	No educations	Primary school	High school	Uni. degree	F.	Sig.
Male:	5186.3	5304.5	5454.3	6009.8	4006.6	0.000
Female:	5359.0	5560.9	5751.4	6419.3	10214.4	0.000
Total:	5284.9	5473.6	5651.4	6244.6	13469.9	0.000

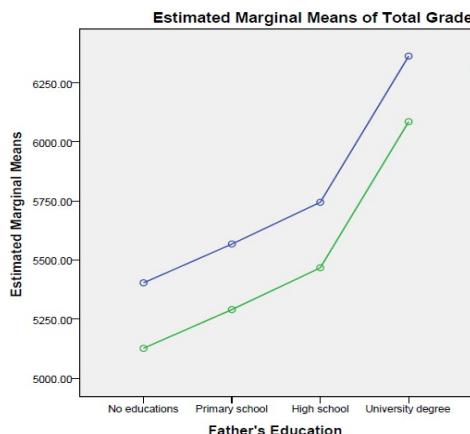


Figure 3. Estimated the mean of total grades by father's education and the gender of candidates

Table 5. The mean of total grades of applicants by mother's education and the gender of candidates

	No educations	Primary school	High school	Uni. degree	F.	Sig.
Male:	5237.6	5346.5	5601.1	6162.5	3490.0	0.000
Female:	5411.1	5610.1	5918.9	6737.8	10318.1	0.000
Total:	5339.3	5522.7	5808.4	6468.9	12865.4	0.000

4.2. Parental Job

Table 6 shows the mean of total grades of applicants relative to their father's job. The father's job is categorized into four categories: education related (teachers, university lecturers), governmental employee, private sector, and other jobs. The means of total grades were 6071.8, 5733.3, 5626.3, and 5254.4, for applicants, whose fathers' jobs are educational related, government employee, private sector employee, and other respectively. The mean of

total grades for applicants whose fathers are teachers or university lecturers is more than other categories. However, the mean of total grades for those applicants whose father is a governmental employee is greater compared to the other two groups. Probably government employees provide better quality of care to children and try to construct a supportive environment for their children by encouraging them to study.

Table 6. The mean of total grades of applicants by father's job and the gender of candidates

	Other jobs	Private sector	Gov. emp.	Teacher or lecturer	F.	Sig.
Male:	5148.1	5437.1	5566.2	5833.5	1608.0	0.000
Female:	5333.0	5716.0	5830.0	6243.7	4188.4	0.000
Total:	5254.4	5626.3	5733.3	6071.8	5762.3	0.000

4.3. Family's Income

Table 7 shows the mean of total grades in 2010 regarding the applicants' family income and their gender. The family's income is grouped into four categories which are weak (<4200 USD), average (4200-5400 USD), good (5400-8700 USD), and very good (>8700 USD) yearly. The mean of total grades in four categories are 5385.8, 5611.7, 5965.8, and 6186.8 for applicants, whose family incomes are weak, average, good, and very good respectively.

Table 7. The mean of total grades of applicants by family's income and the gender of candidates

	Weak	Average	Good	Very good	F.	Sig.
Male:	5231.3	5438.6	5806.9	5984.9	2882.9	0.000
Female:	5479.5	5706.4	6058.0	6326.7	5974.3	0.000
Total:	5385.8	5611.7	5965.8	6186.8	8673.4	0.000

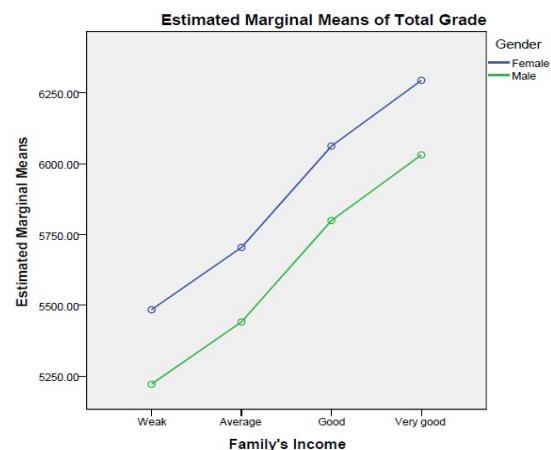


Figure 4. Estimated the mean of total grades of applicants by family's income and the gender

The mean of total grades for applicants whose family's income is good or very good are greater than the other categories. This pattern is the same for both sexes and the only difference is that the mean of total grades for females are slightly higher than for

males'. As can be seen from figure 4 and the ANOVA result that the means of total grades of applicants have gone up with the family's income increase.

5. Inferential Analysis

In this section the result of some inferential analysis such as logistic regression, linear regression and neural networks are presented.

5.1. Binary logistic regression

In this analysis, the dependent variable is the acceptance in universities and colleges, while the independent variables or covariates are age, gender, father's job, father's education, mother's education and family's income. In this analysis the stepwise method is used to determine the order of the variables importance.

Table 8. The coefficients and variables in the logistic regression model

Variable name	B	S.E.	Wald	Sig.	Exp(B)
Age	-0.127	0.001	24986.0	0.000	0.881
Mother's Ed.	0.102	0.003	1017.9	0.000	1.107
Family's Inc.	0.070	0.003	683.8	0.000	1.072
Father's Job	0.055	0.003	311.1	0.000	1.057
Father's Ed.	0.037	0.003	124.0	0.000	1.037
Gender	-0.029	0.004	45.8	0.000	0.972
(Constant)	1.536	0.019	6211.3	0.000	4.644

Table 8 shows how two variables, namely age and gender, have a negative effect on the acceptance in the universities. That means if the age of a candidate increased, he would have a less chance to get accepted in universities. The variables in Table 8 are ordered according to their importance in the model. That means, in this model, the age is the most important factor affecting the acceptance in the universities and after that mother's education has the maximum effect and the gender has the minimum effect.

5.2. Linear regression

The linear regression analysis is used in which the dependent variable is the mean of total grade.

Table 9. The variables and coefficients in the linear regression model

	Unstandardized Coefficients		Standardized Coefficients		
Variable name	B	Std. Error	Beta	t	Sig.
Age	-116.77	0.403	-0.271	-290.0	0.000
Family's Inc.	141.99	1.893	0.078	75.0	0.000
Father's Ed.	71.08	2.319	0.043	30.7	0.000
Gender	-223.81	2.984	-0.069	-75.0	0.000
Mother's Ed.	83.45	2.264	0.047	36.8	0.000
Father's Job	38.87	2.197	0.020	17.7	0.000
(Constant)	7629.16	11.041		691.0	0.000

This analysis is performed like the logistic regression in section 5.1, where the independent variables are age, gender, father's job, father's education, mother's education and family's income. Table 9 shows the order of the variables according to their importance in the linear regression model.

5.3. Neural networks

The neural networks have been used for predictive purposes, i.e. not only for classification but also for regression of continuous target attributes. Figure 5 shows the most important variables based on the final model, which are age of participant, mother's education, family's income, father's job, father's education and gender respectively. The analysis of this model has an approximate 63.6% accuracy in classification of the acceptance in the universities.

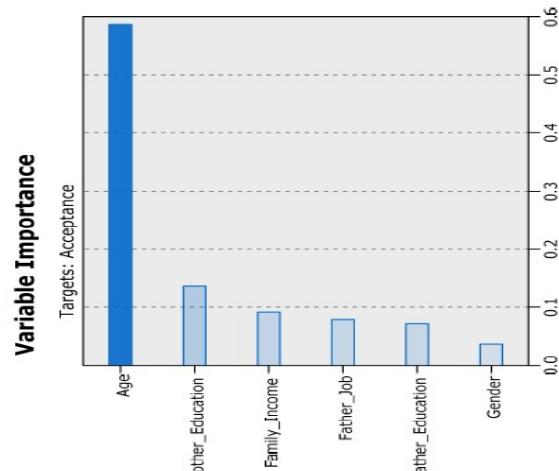


Figure 5. The neural networks results for classification

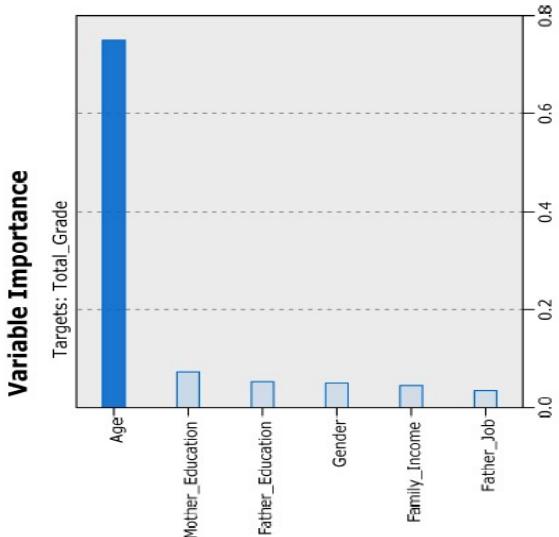


Figure 6. The neural networks results for prediction

Figure 6 shows the neural networks results of prediction model, in which independent variable is the mean of total grade. Due to this model age, mother's education, father's education, gender, family's income and father's job are considered the most important variables respectively. The analysis of this model has an approximate 41% linear correlation in prediction of the mean of total grade.

6. Conclusions

It can be concluded that, in most cases, the mean of total grades for male applicants fell further behind the total grades of their female peers according to the following variables: parental education, parental job, and family's income. Additionally, female grades increased marginally at all higher parental education levels and increased substantially for applicants with university-educated parents. Thus, the data mining methods and regression models show that all of the family background factors influence on participates' educational achievement. Some factors have positive effects on the mean of total grades such as: parental education, parental job and family's income. Additionally, the age of applicants has a negative effect on the candidates' educational achievement. Neural Net analysis indicated an additional point of view in this study. Due to this analysis the important factors on educational achievement are age, father's education, family's income, mother's education, father's job, and gender respectively.

7. Acknowledgements

The authors are very grateful to the National Organization for Educational Testing of IRAN for providing the database of all applicants applying for the nationwide exam in 2010.

8. References

- [1] Blau, P.M., and O.D. Duncan, *The American occupational structure*, New York, Wiley, 1967.
- [2] Chambers, E.A., and J.B. Schreiber, "Girls academic achievement: Varying associations of extracurricular activities", *Gender and Education*, 2004, Vol. 16, pp. 327-346.
- [3] DiMaggio, P., "Cultural capital and school success: The impact of status cultural participation on the grades of U.S. high school students", *American Sociological Review*, 1982, Vol. 47, pp. 189-201.
- [4] Eitle, T.M., "Do gender and race matter? Explaining the relationship between sports participation and achievement", *Sociological Spectrum*, 2005, Vol. 25, pp. 177-195.
- [5] Featherman, D.L., and R.M. Hauser, *Opportunity and change*, New York: Academic Press, 1978.
- [6] Gottfredson D.C., "Black-white differences in the educational attainment process: What have we learned?" *American Sociological Review*, 1981, Vol. 46, pp. 542-557.
- [7] Jmali, E., "The effects of socioeconomic status on educational performance of higher education applicants in Iran." *Journal of Higher Education in Iran*, 2010, Vol. 10, pp. 25-54.
- [8] Jmali, E., "The Trend of Socioeconomic Status of Nation Wide examination Applicants on Their Educational Performances in 2001 to 2009", *Journal of Higher Education in Iran*, 2012, Vol. 16, pp. 25-56.
- [9] Jencks C., J. Crouse, P. Mueser, "The Wisconsin model of status attainment: A national replication with improved measures of ability and aspiration", *Sociological Education*, 1983, Vol. 56, pp. 3-19.
- [10] Khodaie, E., "A study on the relationship between parental socio-economic capital and the probability of students' acceptance in the 2006 nationwide examination of Iran", *Journal of Higher Education in Iran*, 2009, 4, pp. 65-84.
- [11] Khodaie, E., "Effective Factors on Passing in National Entrance Exam in Postgraduate Level", *Quarterly journal of Research and Planning in Higher Education*, 2010, Vol. 15(4), pp. 19-34.
- [12] Lochner, L. and P. Belley, "Educational attainment by parental income: a Canada - U.S. comparison", *The Boston University Econometrics Seminar*, November 2008.
- [13] Marini, M.M., "Sex differences in the determination of adolescent aspirations: A review of research", *Sex Roles*, 1978, Vol. 4, pp. 723-753.
- [14] Majoribanks, K., "Family learning environments and students' outcomes", *Journal of Comparative Family Studies*, 1996, Vol. 27, pp. 373-394.
- [15] Mirashrafi, S.B., E. Khodaie, G. Bol, and G. Nakhaeizadeh, "The impact of family background and socioeconomic status on university admission in Iran", *14th Annual International Conference on Education*, Athens, Greece, May 2012.
- [16] Mirashrafi, S.B., G. Bol, and G. Nakhaeizadeh, "The impact of family background and socioeconomic status on university admission in Iran in 2009", *Canada International Conference on Education*, Ontario, Canada, June 2012.
- [17] Mirashrafi, S.B., G. Bol, and G. Nakhaeizadeh, "The effect of individual factors, family background and socioeconomic status on university admission in Iran in 2007." *Literacy Information and Computer Education Journal*, 2013, Vol. 4(3), pp. 1042-1048.
- [18] Pedrosa, R.H.L., et al., "Educational and socioeconomic background of undergraduates and academic performance: consequences for affirmative action programs at a Brazilian research university", *IMHE/OECD General Conference*, Paris, September 2006.

- [19] Sacker, A., I. Schoon, and M. Bartley, "Social inequality in educational achievement and psychological adjustment throughout childhood: magnitude and mechanisms", *Social Science and Medicine*, 2002, Vol. 55(8), pp. 63-880.