Comparing High Achievers to Low Achievers: An Examination of Student Inputs versus School Inputs in the Educational Outcomes of English Adolescent

Amira Elasra

July 2024 No: 1503

Warwick Economics Research Papers

ISSN 2059-4283 (online)
ISSN 0083-7350 (print)

Comparing High Achievers to Low Achievers: An Examination of Student Inputs versus School Inputs in the Educational Outcomes of English Adolescent

Amira Elasra[1]

Abstract

This paper investigates the association between sets of inputs and the educational outcomes of English adolescents. By linking the Longitudinal Study of Young People in England and Ofsted data, the paper employs the Context-Input-Process-Outcome model to compare the correlation of students and school inputs with their cognitive and non-cognitive outcomes. Using Nonlinear Canonical Correlation Analysis, the paper compares the characteristics of the high achievers to those of the low achievers revealing consistency with current findings in the literature. The results reveal that student inputs exert a greater influence than school inputs in revealing these characteristics. Specifically, unlike low achievers high achievers tend to exhibit positive attitudes toward school, benefit from supportive home learning environments, express greater eagerness to pursue university education, and belong to higher socio-economic backgrounds.

Keywords: Educational outcomes, Nonlinear Canonical Correlation Analysis, English adolescents.

1. Introduction and literature review

Numerous studies examined the impact of student and school inputs on students' outcomes. In a comprehensive synthesis of the influence of school inputs on student outcomes, Glewwe, et al. (2011) provided a meta-analysis spanning both educational and economic literature from 1990 to 2010. Investigating school infrastructure, pedagogical resources, and the attributes of educators, including teachers and principals, the study revealed predominantly positive effects on student's achievement. Conversely, the impact of various school organizational inputs was noted to be less straightforward, exhibiting a degree of ambiguity. Furthermore, certain studies exploring educational production functions may disproportionately emphasize the impact of resource inputs on school efficiency, neglecting the significance of school context, processes, and student inputs (Levaččićć and Vignoles, 2002).

In comparison, findings in the literature indicate that students' inputs outweigh school inputs in explaining variations in academic performance. However, the precise impact of certain student inputs remains debateable. For instance, the effect of ethnic composition on student outcomes in countries like the USA and the UK has been a topic of debate. In the UK, Afro-Caribbean students tend to underperform compared to their white counterparts, while Asian

[1] Assistant Professor, Department of Economics, University of Warwick. A.Elasra@warwick.ac.uk.

students often outperform white students (Bradley and Taylor, 2004). However, these findings are not universal. In a study comparing predominantly white to predominantly minority schools in Bradford and Leicester, variations were observed where Indian students generally outperformed white students, while Pakistani students' performance fluctuated depending on the ethnic composition of the school (Johnston, Wilson, and Burgess, 2006). Such correlations may reflect either causal relationships or school-choice dynamics, highlighting the complex nature of ethnicity's influence on student outcomes.

The impact of welfare support programs on educational attainment has yielded equivocal findings. Ku and Plotnick (2003) reported a significant negative relationship between exposure to welfare and schooling in the USA, but this relationship became insignificant after controlling for income. Similarly, Newman and Harkness (2000) found a positive association between living in public housing and low educational attainment, which disappeared after adjusting for demographic and family background factors. They argued that disadvantaged family backgrounds, rather than housing status, explained lower educational attainment.

Neighborhood effects on educational outcomes have also been scrutinized. While some studies in the USA suggest diminishing neighborhood effects after accounting for family background factors (Solon, Page, and Duncan, 2000), others assert the independent influence of neighborhood poverty on educational attainment (Galster et al., 2007). Socioeconomic characteristics, including parental education levels, are deemed crucial inputs in education production, with their exclusion potentially leading to model misspecifications (Gyimah-Brempong and Gyapong, 1991). However, in the UK, the effects of school and neighborhood are dominated by parental education in explaining variances in children's attainment (Chowdry, Crawford, and Goodman, 2009).

Parental socioeconomic factors, such as education, income, and occupation, significantly influence children's educational attainment. parental education has a persistent positive impact on children's attainment, with higher maternal education levels linked to better outcomes (Connelly and Vernon, 2019). The educational transmission effect is stronger from mothers than fathers in countries like Germany, Norway, and Sweden (Anger and Heineck, 2009). In the UK, parental education, particularly maternal education, is associated with higher children's attainment, with maternal education having a causal relationship with academic success (Chowdry, Crawford, and Goodman, 2009).

Family income plays a pivotal role in children's education, with higher incomes associated with better academic performance in various countries (Chevalier, et al., 2013; Dahl and Lochner, 2008). Economic segregation, while not significantly affecting overall educational attainment, exacerbates inequalities between low and high-income children (Mayer, 2000; 2001). Similarly, in the UK, family income correlates strongly with children's educational outcomes, with students from low-income families achieving lower academic attainment compared to their affluent counterparts (Chowdry, Crawford, and Goodman, 2009).

Job loss experienced by parents in the USA is associated with reduced likelihood of children pursuing postsecondary education, particularly among black children (Kalil and Wightman, 2011). Conversely, the impact of mothers' work on children's attainment is inconclusive, with some studies indicating a positive relationship but attributing it to unobserved family heterogeneity (Huang, 2003).

Students' behaviors and prior attainment significantly influence their educational outcomes. Early cognitive ability strongly correlates with future academic progression, with parental wealth positively influencing children's attainment (Glick and Sahn, 2010). Changes in students' attitudes and behaviors, especially regarding higher education, are closely associated with changes in educational attainment, highlighting the pivotal role of student attitudes in shaping academic success (Chowdry, Crawford, and Goodman, 2009).

The previous findings are predominantly based on educational production functions of single output methodologies (Fortune, and O'Neil, 1994; Knoeppel, Verstegen, and Rinehart, 2007). Such methodologies may overlook potential relationships among various educational outputs. Employing simultaneous equations modeling to circumvent the singularity issue may presuppose causal relationships, which may not always hold true within the educational framework (Gyimah-Brempong and Gyapong, 1991).

To investigate the intricate interplay between educational inputs and multiple outcomes the 'Context – Input – Process – Outcome' (CIPO) model proposed by Teddlie and Reynolds (1999) is used. This model captures the various types of inputs that impact students' educational outcomes. It illustrates how school context, schooling processes, student inputs and resource inputs, shape students' educational outcomes.

A method that tend to be overlooked in the Economics of Education literature is linear canonical analysis (Knoeppel, Verstegen, and Rinehart, 2007), which investigates the correlations among sets of inputs and outputs (Thompson, 1991). This paper contributes to the existing literature by analyzing the correlations between the sets of the Context-Input-Process-Outcome model using Nonlinear Canonical Correlation Analysis (NLCCA). The main advantage of NLCCA is its ability to derive the clusters identifying how the variables in these sets correlate. This in turn helps reveal the disparities in characteristics of the high and low achievers.

The dataset and models are described in the subsequent section, while section three explains the NLCCA methodology. The results section discusses the principal findings, and section five provides the paper's conclusion.

2. Dataset and Models

The paper uses three databases to investigate the Context-Input-Process-Outcome model: the Longitudinal Study of Young People in England (LSYPE), the National Pupil Database (NPD), and the Ofsted database. Specifically, data from 4237 students born in 1989-1990 (DfE, 2011) who completed their Key Stage Four (KS4) in 2005/06 is analyzed. The dataset links their cognitive outcome from the NPD with their characteristics from waves one (2004), two (2005), and seven (2010) of the LSYPE, as well as the characteristics of the schools they attended using the Ofsted school reports.

The analysis investigates the relationships between the five components (sets) of the CIPO model depicted in figure (1). These are school context variables, student's inputs, resources inputs, school process variables and student outcomes. The analysis examines two types of outcomes: cognitive outcome and non-cognitive outcomes. The cognitive outcome is measured by their KS4 total GCSE/GNVQ score. The cognitive outcome model examines the correlation between 38 variables. The non-cognitive outcomes are measured using three variables: the type of higher education qualification pursued, average annual salary, and standard occupational classification, all measured at the age of 19. This model examines the correlations between 31 variables. To overcome the missing data problem, multiple imputation of missing data has been implemented using a Fully Conditional Specification method based on an iterative Markov Chain Monte Carlo (MCMC) simulation using a Gibbs sampler algorithm (Elasra, 2022a, b).

Figure 1: CIPO Model description

SCHOOL CONTEXT

- Urban/Rural Indicator.
- Whether YP was at an independent or maintained school.
- IDACI (Income Deprivation Affecting Children Index).
- Phase of education

STUDENT's INPUTS

- Banded total amount of benefits.
- Mean annual salary of household over waves (1-2).
- MP: Frequency of additional private lessons over the last 12 months.
- MP: How YP's expenses would be paid if stayed on in education- Parent(s) will support or give money.
- Highest qualification of father.
- Highest qualification of mother.
- Father NS-SEC class.
- Mother NS-SEC class.
- Whether can access internet from home.
- Whether have home computer in household.
- Type of household tenure.
- DV: Family composition.
- Ethnicity.
- YP gender.
- Special educational needs (SEN).
- YP: Whether has a long-standing physical or mental impairment, illness or disability.
- Prior attainment.
- YP's religion
- YP: Importance of religion to YP's way of life.
- YP: Frequency of going to religious classes or courses in the last 12 months.
- YP: Whether YP thinks their religious beliefs will affect how likely they are to get a job or training place.
- YP age.

RESOURCE INPUTS

- How effectively and efficiently resources, including staff, are deployed to achieve value for money?

SCHOOL PROCESS

School level

- MP: How involved is the MP in YP's school life?
- Procedures for safeguarding learners meet current government requirements
- How well does the school work in partnership with others to promote learners' well-being?
- How well equality of opportunity is promoted and discrimination tackled so that all learners achieve as well as they can.
- Education for all learners aged 14-19 provides an understanding of employment and the economy.

Pupil level

- MP: How YP's expenses would be paid if stayed on in education - YP get job or work part time.
- YP: Likelihood of YP applying for university
- Mean score of Young person attitude to school over waves (1-2).
- The extent to which learners adopt safe practices.
- How well learners develop workplace and other skills that will contribute to their future economic well-being.
- The extent to which learners make a positive contribution to the community.
- How well learners with learning difficulties and disabilities make progress.

STUDENT OUTCOMES

Cognitive outcomes

- KS4 Total GCSE/GNVQ new style point score (KS4).

Non-Cognitive outcomes

- Higher Education Qualification Being Studied at 19 (HED).
- Average Yearly Pay (PAY).
- Major groupings for young person SOC (SOC).

3. Nonlinear Canonical Correlation Analysis

Nonlinear multivariate analysis offers researchers a more comprehensive simulation of reality, facilitating broader generalizations. Nonlinear canonical correlation analysis extends regression analysis to encompass multiple dependent variables. This approach involves exploring correlations among various sets of variables to ascertain both the primary factors driving the relationships between these sets and the distinguishing characteristics of different groups or clusters of objects, such as students (Van der Burg, De Leeuw, and Dijksterhuis, 1994). A notable advantage of NLCCA is its freedom from distributional assumptions or predefined models (Yazici, et al., 2009). Moreover, the outcomes derived from NLCCA remain consistent even when subjected to certain nonlinear transformations of the data.

One method under the NLCCA family is based on the minimization of a loss function using an alternating least squares algorithm (ALS)[2] (Yazici, et al., 2009). It investigates multiple sets of categorical variables that are optimally scaled. By finding the correlations between these sets the algorithm identifies not only the most important variables in explaining the relationships between these sets, but also the characteristics of different objects (i.e students) based on the grouping of the categories of all variables within all sets (Van der Velden and Takane, 2012).

To explain the ALS algorithm, let J denotes the categorical *variables* for N *objects* (students), where variable $j \in J = \{1, 2, \ldots, J\}$ has l_j categories. The J set of variables is classified into K subsets $J(1), \ldots, J(k), \ldots, J(K)$. The algorithm constructs a low dimensional joint map of objects and categories in a p dimensional space R^p, where $p \prec J$. X is the $N \times p$ *object scores matrix* containing the coordinates of the objects vertices in R^p. Y_j is the $l_j \times p$ *category quantification matrix* containing the coordinates of the l_j category vertices of variable j. G_j is an indicator matrix of variable j with entries $G_j(i,t) = 1$, $i = 1, \ldots, N$, $t = 1, \ldots, l_j$ if object i belongs to category t and $G_j(i,t) = 0$ if it belongs to some other category. The G matrix is the *adjacency matrix* of the bipartite graph. If edges are used to connect each category, the loss function that needs to be minimized would be the average squared edge length (over all variables) and would be given by

$$\sigma(X; Y_1, \ldots Y_j) = K^{-1} \sum_{k=1}^{K} SSQ(X - \sum_{j \in J(k)} G_j Y_j) \quad (1)$$

where $SSQ(H)$ is the sum o f squares of the elements of matrix H (Gifi, 1981) and is subject to two constraints given by;

$$X^{'} X = NI_p \quad (2)$$

$$u_N^{'} X = 0 \quad (3)$$

[2] Also known as OVERALS.

where u is a column with n elements equal to one. The first constraint (2) standardizes the squared length of object scores to be equal to N and in two or more dimensional spaces requires the columns of X to be orthogonal. Such constraint helps bypassing the trivial solution corresponding to $X = 0$, and $Y_j = 0$ for every $j \in J$. It also makes the columns of X uncorrelated, with variances equal to one. The second normalization constraint (3) requires the graph plot to be centred around the origin (Van der Burg, De Leeuw, and Verdegaal, 1984; 1988).

The algorithm solves the optimization problem in three steps. First, it minimizes equation (1) with respect to Y_j for fixed X and provides the solution:

$$Y_j = D_j^{-1} G_j^{'} X \qquad (4)$$

where $D_j = G_j^{'} G_j$ is the $l_j \times l_j$ diagonal matrix containing on its diagonal the relative frequencies of the categories of variable j. Equation (4) shows that one category quantification is in the centroid of the object scores that belong to it corrected for the influence of other variables in the set. Second, it minimizes equation (1) with respect to X for fixed Y_j's and provides the solution:

$$X = K^{-1} \sum_{k=1}^{K} \sum_{j \in J(k)} G_j Y_j \qquad (5)$$

Equation (5) shows that an object score is the average of the quantifications of the categories it belongs to. That is object scores are simply the averages of quantified variables (Van de Geer, 1987). In the third step the object scores X are columned centered by setting $W = X - u_N (u_N^{'} X / N)$, and then orthonormalized by the modified Gram-Schmidt procedure $X = \sqrt{N} GRAM(W)$, so that both of the normalization constraints (2 and 3) are satisfied (Michailidis and De Leeuw, 1998; Yazici, et al., 2009).

Simply put, the correlations between the multiple sets of inputs and outputs are identified by how X (the object scores) and Y_j (the category quantification) interact together. Such correlations identify the characteristics of different groups of objects (students) based on the grouping of the categories of all variables within all sets.

The relationships among the sets of variables are explained by the *fit* and *loss* values. The *fit* value shows how much the model accounts for variations in the object scores. It is the sum of eigenvalues (i.e how much of the relationships between the sets is shown by each dimension) and has a maximum of the number of dimensions p. A maximum fit value indicates that the relationship is perfect. The *loss* value is the difference between the maximum and the actual fit value and is provided for each dimension and set. It indicates the proportion of variation in the object scores that cannot be accounted for by the weighted combinations of variables in that set. Another related measure is the multiple correlations between weighted sums of optimally scaled variables and dimensions indicating how much each set accounts for the variance in each dimension (Van der Burg, De Leeuw, and Dijksterhuis, 1994).

Another key statistic is the *component loading* of each variable that reflects its importance in explaining the relationships between the sets of variables (Michailidis and De Leeuw, 1998). It represents the correlation between the object scores and variables. That is, on a two-dimensions component loadings plot, the closer the loadings of a variable to the origin, the less important it is in explaining the relationships between the sets of variables and vice versa. Moreover, the closer the variables' loadings of different sets to each other, the more related those sets are to each other. To identify clusters of categories that characterizes the high achievers and the low achievers, the algorithm provides the *category centroid.* It is the centroid of all objects scores that share the same category (Yazici, et al., 2009). It explains how categories of variables correlate with each other for groups of objects (i.e. students).

4. Findings

Cognitive Outcome

This analysis starts with examining students' cognitive outcomes measured by their key stage four total GCSE/GNVQ score (KS4). The model's fit value indicated in table (1) is 0.977, which means it accounts for 49% of the variation in the object scores. That is 51% of the variation in the object scores cannot be explained by the weighted combinations of variables in all sets, with a mean loss value of 1.023. The two dimensions account for 55% and 42% of the fit value respectively. The multiple correlations for the five sets (student's inputs, school context variables, resources inputs, school process variables and student outcomes) reveal that the student's inputs set accounts for the highest variance in the two dimensions.

As shown by the component loadings in figure (2), the five most contributing variables (farthest from the origin) in explaining the relationships between the five sets are the urban/rural indicator (UR), KS4 score (KS4), prior attainment (PAT), the type of household tenure (TENURE), and the Likelihood of the student applying for university (UNI) respectively. On the contrary, the least contributing variables (closest to the origin) are the total annual amount of benefits received (BEN), how well the school works in partnership with others to promote learners' well-being (OE3) and the student's age (AGE).

Table 1: Fit, Loss and Correlations Values of the Cognitive Outcome Model

		Dimension		Sum	Multiple correlation	
					Dimension	
		1	2		1	2
Loss	Student's inputs	0.267	0.235	0.502	0.856	0.875
	School context	0.681	0.261	0.942	0.565	0.860
	Resources inputs	0.652	0.895	1.547	0.590	0.324
	School process	0.312	0.503	0.815	0.829	0.705
	Student outcome	0.317	0.992	1.309	0.826	0.089
	Mean	0.446	0.577	1.023		
Eigenvalue		0.554	0.423			
Fit				0.977		

Figure 2: Component Loadings Plot of the Cognitive Outcome Analysis

To identify the characteristics of the high and low achievers, the category centroids help in understanding the correlation between variables. The centroids explain how the categories of variables correlate with each other for groups of students. To better understand the characteristics of these students, clusters of categories are derived representing each group of achievers. The clustering of categories is based on how close the values of their distances from the origin are to each other (Michailidis and De Leeuw, 1998; Yazici, et al., 2009; IBM, 2021).

Table (2) shows that the top two achievers in prior education usually maintain their position in their KS4. These students identify as Hindu or Sikh, who typically demonstrate positive attitudes toward schooling. Moreover, this subset exhibits a strong inclination toward pursuing higher education, with a high propensity for university applications. Demographically, their parents predominantly consist of married couples residing in housing tenures secured through mortgage or bank loans, while their fathers commonly hold degrees or equivalent qualifications. Conversely, students who exhibit low academic performance have subpar academic performance in their previous education and attend special schools. They typically show minimal interest in pursuing higher education and reside in households without parental figures, often residing in sparse towns or villages with limited parental involvement in their education.

Table 2: Centroids of the high and low cognitive achievers

Variable	Distance from origin
The first and second high achievers	**0.933**
	1.067
have the second highest prior attainment	0.519
have the fourth highest prior attainment	0.973
Hindu	0.854
Sikh	0.355
have fathers have a degree or equivalent	0.769
very likely to apply to university	0.700
have the highest two attitudes to school	0.425
	0.597
live in tenures that are being bought in a mortgage or bank loan	0.406
parents are married couples	0.119
The lowest achievers	**1.561**
go to special schools	1.877
live in a household with no parents	1.823
live in town & fringe–sparse	1.333
live in village-sparse	1.774
live in shared ownership	1.677
live in some other arrangements	1.455
there is no involvement at all from the parents in their children education	1.401
the worst performers in their prior education as well	1.398
not likely at all to apply to university	1.191

The above findings underscore a robust association between students' outcomes and their previous academic performance. Generally, low achievers exhibit negative attitudes toward school, lack access to quality home learning environments, and experience unfavorable family compositions. Conversely, high achievers demonstrate positive attitudes toward school and come from households with comparatively higher socio-economic status.

Non-cognitive Outcomes

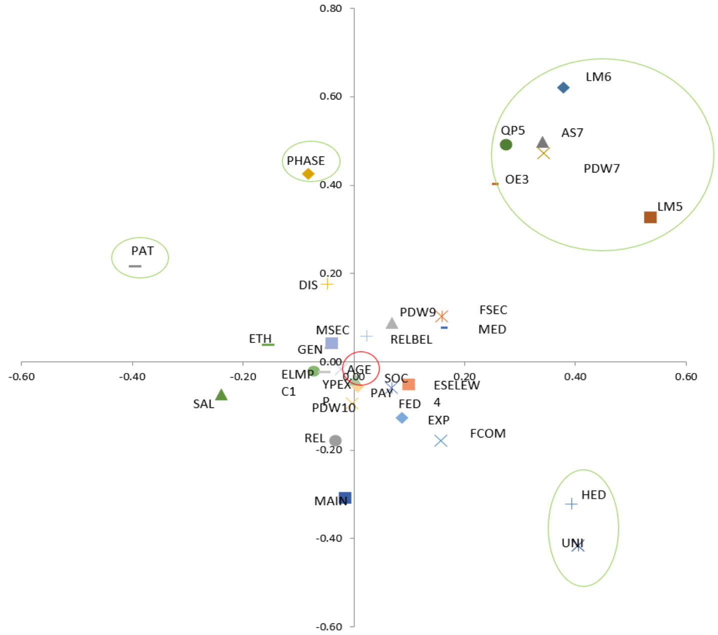
The analysis broadens to explore the characteristics of high and low achievers concerning their non-cognitive outcomes post-secondary education. The goal is to understand student behavior after leaving school at age 19 and how their outcomes—such as the type of higher education qualification pursued (HED), average annual salary (PAY), and standard occupational classification (SOC)—correlate with their observed characteristics.

The model explains 42% of the variation in the object scores. However, unlike the cognitive outcome model the multiple correlations indicate that the school process set accounts for the highest variance in the two dimensions rather than the student's inputs set (table 3). As shown by the component loadings in figure (3), the most important five variables in explaining the relationships between the five sets of variables are how effectively and efficiently resources, including staff, are deployed to achieve value for money in the school (LM6), how well equality of opportunity is promoted and discrimination tackled in the school so that all learners achieve as well as they can (LM5), how well learners with learning difficulties and disabilities make progress in the school (AS7), the extent to which learners adopt safe practices (PDW7), and the likelihood revealed by the student in their KS4 to apply for university (UNI) respectively, while the least important variable is AGE.

Table 3: Fit, Loss and Correlations Values of the Non-Cognitive Outcome Model

Value	Set	Dimension		Sum	Multiple correlation	
					Dimension	
		1	2		1	2
Loss	Set 1 (Student's inputs)	0.365	0.683	1.048	0.797	0.563
	Set 2 (School context)	0.982	0.722	1.704	0.135	0.527
	Set 3 (Resources inputs)	0.583	0.613	1.196	0.645	0.622
	Set 4 (School process)	0.217	0.299	0.516	0.885	0.837
	Set 5 (Student outcomes)	0.526	0.842	1.367	0.689	0.398
	Mean	0.535	0.632	1.166		
Eigenvalue		0.465	0.368			
Fit				0.834		

Figure 3: Component Loadings Plot of the Non-cognitive Outcomes Model



Starting with their salary outcome at age 19, table (4) shows that the lowest salary earners are typically pursuing a degree while being employed in "associate professional and technical occupations" or "sales and customer service occupations." They belong to Indian, Pakistani, Bangladeshi, or Black African ethnic groups, possess the second, third, or fourth highest prior attainment levels in their KS3, and have fathers with degrees or equivalent qualifications.

Individuals with the second lowest salary are engaged in "administrative and secretarial occupations" and are typically Black Caribbean females with no long-standing disabilities. They are moderately inclined to pursue university education and attended secondary schools known for their effective partnership initiatives aimed at promoting learner well-being. Moreover, their parents are married couples who are supportive of their future education, with fathers holding GCE A level qualifications or their equivalent.

Individuals pursuing a "diploma in higher education" or the lowest level of higher education qualifications[3] occupy positions as "managers and senior officials." These individuals typically exhibit the lowest levels of prior attainment and demonstrate a minimal likelihood of applying to university, often compounded by the presence of a disability. Moreover, they commonly reside in households without parental figures.

[3] HND/HNC/RSA or OCR Higher Diploma/NVQ Level 4 or 5/Certificate in Higher Education/Other.

Conversely, the top two earners typically occupy positions in "elementary occupations." They possess the second and third lowest levels of prior attainment and exhibit a low likelihood of applying to university. These individuals are often found residing with single fathers, who possess qualifications at level 1 and below or other equivalent qualifications. Additionally, their mothers are typically employed in lower supervisory and technical roles, and neither parent demonstrates a willingness to support their future education.

Table 4: Actual Centroids of the highest and lowest salary earners

Variable	Distance from origin
have a degree	**0.485**
have the lowest pays	**0.354**
work in 'associate professional and technical occupations' or 'sales and customer service occupations'	**0.420**
	0.320
have the second highest prior attainment	0.335
have the third highest prior attainment	0.442
have the fourth highest prior attainment	0.561
Indian	0.555
Pakistani	0.368
Bangladeshi	0.419
Black African	0.332
fathers tend to have degree or equivalent	0.466
work in 'administrative and secretarial occupations'	**0.169**
the second lowest salary earners	**0.122**
fairly likely to apply to university	0.182
more likely to have a married couple parents	0.168
have the fifth highest prior attainment	0.147
Black Caribbean	0.130
females	0.125
their fathers have GCE A level or equivalent	0.098
parents are willing to support them in their future education	0.067
more likely to have attended secondary schools	0.064
with no long-standing disability	0.054
attended schools that are good in how they work in partnership with others to promote learners' well-being	0.046
have a 'diploma in higher education' or 'HND/HNC/RSA or OCR Higher Diploma/NVQ level 4 or 5/certificate in higher education/other'	**0.508**
	0.765
likely to be working as 'managers and senior officials	**0.525**
are not likely at all to apply to university	0.984
have the lowest prior attainment	0.866
may be living in a household with no parents	0.661
with a disability	0.619
The highest two salary earners	**0.306**
	0.275
work in 'elementary occupations'	**0.279**
have the second lowest prior attainment in KS3	0.431
have the third lowest prior attainment in KS3	0.252
are not very likely to apply to university	0.393
fathers have qualifications at level 1 and below	0.357
fathers have other qualifications	0.366
parents are not willing to support them in their future education	0.348
living with lone fathers	0.211
their mothers work in lower supervisory and technical occupations	0.209

To examine if a change in the data structure would enhance the fitness of the model and to check the robustness of the findings, the two models have been re-specified with three rather than five sets. The three sets include the student cognitive (or non-cognitive) outcomes; the student inputs (combing student's inputs and resources inputs) and the school inputs (combining school context and process). Indeed, the three sets model of both cognitive and (non-cognitive) outcomes explains 71% (62%) of the variation in the data compared to 49% (42%) in the five sets model. Accordingly, it can be suggested that reducing the number of sets improves the fitness of the model. Additionally, the three sets models show similar results to the five sets model in terms of identifying the most and least important variables in explaining the relationships between the sets. Moreover, the student's inputs set accounts for the highest variance in the two dimensions.

5. Conclusion

Numerous studies exploring educational production functions have traditionally relied on analyzing a singular output variable. This paper uses Nonlinear Canonical Correlation Analysis, a technique relatively underutilized in the field of Economics of Education literature. The methodology offers the distinct advantage of facilitating a comprehensive understanding of the intricate correlations among various input and output sets.

Methodologically, the method offers a more nuanced understanding of the educational process. Essentially, its goal is to reveal and depict the underlying structure of categorical multivariate data by transforming it into a lower-dimensional representation. It's important to note that this approach doesn't frame the problem as an estimation task with model parameters and error terms. Instead, it addresses an optimization challenge involving the Gifi loss function, without suggesting any statistical inference. Furthermore, within the Gifi system, stability is indicated when minor alterations in inputs result in similarly minor changes in outputs. The results of this study demonstrate relative stability, yet further exploration into the robustness of these findings through alternative specifications of the Context-Input-Process-Outcome model would be beneficial.

This study adopts the CIPO model to understand the interrelationships among five distinct input and output sets. It examines two models contingent upon the nature of the outcome: cognitive and non-cognitive outcomes. The findings reveal that certain variables emerge as particularly importance. Within the *student's inputs* set, these include student prior attainment, likelihood to apply for university, household tenure type, parental involvement in school activities, and maternal socio-economic class. Notably, the overall quality of the school stands out as the primary variable in the *resources inputs* set. Additionally, the *school context* input of the urban/rural indicator, holds importance in the educational process. Among the *school process* inputs, how effectively and efficiently resources, including staff, are deployed to achieve value for money and how well learners with learning difficulties and disabilities make progress emerges as crucial variables. Conversely, variables such as age, total annual amount of benefits received, and how well the school works in partnership with others to promote learners' well-being are found to be less important.

The findings corroborate the observations from existing literature. The results indicate that the comparison between the two outcomes models reveals a notable degree of stability and consistency in the results. Particularly, a robust association is observed between students' present and past cognitive outcomes, as documented by Glick and Sahn (2010). High achievers exhibit favorable attitudes toward school (Belzil and Leonardi, 2007) and benefit from a supportive home learning environment (Chowdry, Crawford, and Goodman, 2009; Chowdry et al., 2010). They demonstrate a greater propensity to pursue higher education and typically come from households with high socio-economic status (SES) (Mayer, 2000; 2001; Chevalier, et al., 2013, Belzil and Leonardi, 2007; Dahl and Lochner, 2008). Conversely, low achievers display less interest in attaining a higher education degree, have negative attitudes toward school, experience suboptimal home learning conditions, and generally possess lower overall SES levels.

REFERENCES

Anger, S. and Heineck, G. 2009. Do smart parents raise smart children? The intergenerational transmission of cognitive abilities. *Journal of Population Economics*, 23 (October), pp. 1255–1282.

Belzil, C. and Leonardi, M., 2007. Can risk aversion explain schooling attainments? Evidence from Italy. *Labour Economics* 14 (June), pp.957–970.

Bradley, S. and Taylor, J., 2004. Ethnicity, educational attainment and the transition from school. *Manchester School*, 72 (3), pp.317-346.

Chevalier, A., et al. 2013. The impact of parental income and education on the schooling of their children. *IZA Journal of Labor Economics*, 2(1), pp. 1-22.

Chowdry, H., Crawford, C. and Goodman, A., 2009. *Drivers and Barriers to Educational Success Evidence from the Longitudinal Study of Young People in England*. London, Department for Children, Schools and Families.

Chowdry, H. et al., 2010. *Poorer children's educational attainment: how important are attitudes and behaviour?* York: Joseph Rowntree Foundation.

Connelly, R and Vernon, G. 2019. An investigation of social class inequalities in general cognitive ability in two British birth cohorts. The British journal of sociology, 70 (1), pp. 90-107.

Dahl, G. and Lochner, L., 2008. The impact of family income on child achievement: evidence from the earned income tax credit. Massachusetts: National Bureau of Economic Research. Working Paper No.14599.

Department for Education, 2011. *LSYPE User Guide to the Datasets: Wave 1 to Wave 7*. London: Department for Education.

Elasra, A. 2022a. Multiple Imputation of Missing Data in Educational Production Functions". *Computation*, 10 (4), 49.

Elasra, A. 2022b. The Longitudinal Study of Young People in England Multiple Imputed datasets. 2022. The University of Warwick's institutional repository, WRAP (wrap.warwick.ac.uk), reference number 161945.

Fortune, J.C. and O'Neil J.S., 1994. Production Function Analysis and the Study of Educational Funding Equity: A Methodological Critique. *Journal of Education Finance,* 20 (1), pp.21-46.
Galster, G., et al., 2007. The Influence of Neighborhood Poverty during Childhood on Fertility, Education, and Earnings Outcomes. *Housing Studies*, 22 (5), pp.723–751.
Gifi, A., 1981. *Nonlinear Multivariate Analysis*. Leiden: University of Leiden, Department of Data Theory.
Glewwe, et al., 2011. School resources and educational outcomes in developing countries: a review of the literature from 1990 to 2010. Massachusetts: National Bureau of Economic Research. Working Paper No. 17554.
Glick, P. and Sahn, D.E., 2010. Early Academic Performance, Grade Repetition, and School Attainment in Senegal: A Panel Data Analysis. *The World Bank Economic Review*, 24 (1), pp.93–120.
Gyimah-Brempong, K. and Gyapong, A.O., 1991. Characteristics of Education Production Functions: An Application of Canonical Regression Analysis. *Economics of Education Review*, 10 (I), pp.7-17.
Huang, F., 2003. The Impact of Childhood Events on Educational Achievement: A Sibling Study. *Taiwan Economic Review*, 28 (4), pp.425-450.
IBM, 2021. *Nonlinear Canonical Correlation Analysis*. [Online] available at < https://www.ibm.com/docs/sr/spss-statistics/25.0.0?topic=categories-nonlinear-canonical-correlation-analysis-overals> [Accessed March 2024].
Johnston, R., Wilson, D. and Burgess, S., 2005. Ethnic segregation and educational performance at secondary school in Bradford and Leicester. Environment and Planning, 39 (3), pp.609-629.
Kalil, A. and Wightman, P., 2011. Parental Job Loss and Children's Educational Attainment in Black and White Middle-Class Families. *Social Science Quarterly*, 92 (1), pp.57-78.
Knoeppel, R.C., Verstegen, D.A. and Rinehart, J.S., 2007. What Is the Relationship Between Resources and Student Achievement? A Canonical Analysis. *Journal of Education Finance*, 33 (2), pp.183-202.
Ku, I. and Plotnick, R., 2003. Do children from welfare families obtain less education? *Demography*, 40 (1), pp.151-170.
Levaččićć, R. and Vignoles, A., 2002. Researching the Links between School Resources and Student Outcomes in the UK: A Review of Issues and Evidence. *Education Economics*, 10 (3) pp.313-331.
Mayer, S., 2000. Income Inequality, Economic Segregation and Children's Educational Attainment. Chicago: JCPR working paper 209.
Mayer, S., 2001. How economic segregation affects children's educational attainment. Chicago: JCPR working paper 235.
Michailidis, G. and De Leeuw, J., 1998. The Gifi System of Descriptive Multivariate Analysis. *Statistical Science*, 13 (4), pp.307-336.
Newman, S. and Harkness, J., 2000. Assisted Housing and the Educational Attainment of Children. *Journal of Housing Economics*, 9 (February), pp.40–63.
Solon, G., Page, M.E. and Duncan, G.J., 2000. Correlations between neighboring children in their subsequent educational attainment. *The Review of Economics and Statistics*, 82 (3), pp.383-392.
Teddlie, C. and Reynolds, D., 2000. *The International Handbook of School Effectiveness Research*. London: RoutledgeFalmer.
Thompson, B., 1991. A Primer on the Logic and Use of Canonical Correlation Analysis. *Measurement and Evaluation in Counseling and Development*, 24 (2), pp.80-93.
Van de Geer, J.P. 1987. *Algebra and geometry of OVERALS*. Leiden: University of Leiden, Department of Data Theory.
Van der velden, M. and Takane, Y., 2012. Generalized canonical correlation analysis with missing values. *Computational Statistics*, 27 (2012), pp.551–571.
Van der Burg, E., De Leeuw, J., and Verdegaal, R., 1984. *Nonlinear canonical correlation analysis with m sets of variables*. Leiden: University of Leiden, Department of Data Theory.
Van der Burg, E., De Leeuw, J., and Verdegaal, R., 1988. Homogeneity analysis with K sets of variables: An alternating least square method with optimal scaling features. *Psychometrika*, 53 (2), pp.177-197.

Van der Burg, De Leeuw, and Dijksterhuis, 1994. OVERALS Nonlinear canonical correlation with k sets of variables. *Computational Statistics & Data Analysis,* 18 (1994), pp.141-163.
Yazici, A.C., et al., 2010. An application of nonlinear canonical correlation analysis on medical data. *Turkish Journal of Medical Sciences*, 40 (3), pp.503-510.