The Path to College Education: The Role of Math and Verbal Skills*

Esteban M. Aucejo†
Arizona State University

Jonathan James‡
California Polytechnic State University

January 22, 2019

Abstract

This paper studies the formation of math and verbal skills during compulsory education and their impact on adult outcomes. We introduce a novel method to estimate dynamic, nested CES production functions. Using a rich panel database that follows a cohort of students in England from elementary school to university, we find that the production functions of math and verbal skills are inherently different, where cross-effects are only present in the production of math skills. Results on long-term outcomes indicate that verbal skills play a substantially greater role in explaining university enrollment than math skills. This finding, combined with the large female advantage in verbal skills, has key implications for gender gaps in college enrollment and field of study. Finally, we show that students stuck in low quality schools have lower skill levels at the end of compulsory education compared to students attending high quality schools, with these skill deficits leading to a 30 percentage point gap in college enrollment among these students. Simulation results show that about 15% of this gap is due to differences in skill levels at the beginning of compulsory education while about 20% of this gap is attributable to the differences in school quality, which indicates that policies aiming to improve school quality could help to overcome initial skill disadvantages.

*We thank Peter Arcidiacono, Ghazala Azmat, Caroline Hoxby, Monica Langella, Alan Manning, Guy Michaels, Steve Pischke, Tyler Ransom, Matteo Sandi, and Zachary Tobin for valuable comments. We also thank seminar participants at California Polytechnic State University, the Federal Reserve Bank of St. Louis, Royal Holloway, London School of Economics and Political Science, University College London, University of Essex, Arizona State University, NBER meeting: Economics of Education Program, Middlesex University London, Aalto University-University of Helsinki, University of Western Ontario, University of Rochester, University of Surrey, CESifo conference, Inter-American Development Bank, ACLEC conference, UCLA, American Economic Association meetings, University of Wisconsin-Madison, and Yale University. This version of the paper supersedes a previous version under the title "The Path to College Education: Are Verbal Skills More Important than Math Skills?" All errors are our own.

†Department of Economics, W.P. Carey School of Business, Arizona State University & CEP Email: E.M.Aucejo@lse.ac.uk

‡Department of Economics, California Polytechnic State University. Email: jjames04@calpoly.edu. Homepage: `http://www.calpoly.edu/~jjames04`

1 Introduction

The employment prospects of less-educated workers have worsened significantly since the early 1980's (Autor and Wasserman, 2013). As formal education becomes an increasingly important determinant of lifetime income (Castex and Dechter, 2014), understanding the factors that influence schooling decisions is essential from a policy perspective. A large literature has established that cognitive skills play an important role in explaining educational attainment (Heckman et al., 2006; Cawley et al., 2001; Cameron and Heckman, 2001). However, because skills are multiple in nature (Cunha and Heckman, 2007), more attention is needed to understand exactly which types of skills have the greatest influence on post-secondary educational outcomes and how these specific skills are produced over the schooling career.

The aim of this paper is threefold. First, we study how math and verbal skills develop during compulsory education with particular attention to understanding the role of schools. Unpacking skills in the math-verbal dimension provides insight into how formal education operates and will allow us to disentangle how these two skills impact educational decisions. Our work compliments a growing literature that studies how childhood investments contribute to skill production (Cunha and Heckman, 2008b; Cunha et al., 2010; Heckman et al., 2013; Attanasio et al., 2015; Agostinelli and Wiswall, 2016). Second, we study the differential role of math and verbal skills on university enrollment, field of study, and graduation outcomes. These questions are of central policy relevance as they shed light on how changes to school curricula could help to improve educational attainment of students.[1] In particular, policymakers have mainly prioritized the further development of math skills over verbal skills (e.g. the "Algebra-for-All" movement, Loveless (2008), Long et al. (2012)), however little is known about the differential role of these skills in explaining educational outcomes. Finally, we analyze how differences in math and verbal skills between males and females contribute to the well established gender gaps in college enrollment and in STEM (Science, Technology, Engineering, and Math) major choice. Evidence from many developed countries (OECD, 2012) has shown that males are less likely to attend university than females, but males are more likely to

[1] For example, Loveless (2008) claims that "the nation's push to challenge more students by placing them in advanced math classes in eighth grade has had unintended and damaging consequences, as some 120,000 middle-schoolers are now struggling in advanced classes for which they are woefully unprepared."

enroll in STEM fields. However, there is a lack of evidence regarding the role of different types of skills in explaining these gaps.

We study these questions using a large administrative dataset that covers an entire cohort of public school students in England. This dataset tracks the history of educational outcomes from age 5 to 22 of approximately 500,000 students. For each student we observe a rich set of demographic characteristics, their neighborhood and school attended at each grade, as well as university records that include enrollment, institution, field of study, and graduation. Most importantly, for each student we observe scores from more than 30 subject-specific exams taken over the period of compulsory education.[2] We use these performance measures to estimate a dynamic factor model of skill formation following the spirit of Cunha and Heckman (2008b); Cunha et al. (2010); Attanasio et al. (2015); Agostinelli and Wiswall (2016), which allows us to recover the latent skills (i.e. math and verbal) for each individual at different points in the schooling career and provides a theoretical framework to understand the process of skill formation. We contribute to this literature in two important dimensions. First, we estimate nested CES production functions with three inputs, originally described in Sato (1967), where we use the school accountability report card (Ofsted reports) as a measure of school quality for the third input. The benefit of the nested CES production function is that it allows the elasticity of substitution to vary across pairs of inputs.[3] Second, we propose a novel two-step estimation method for dynamic factor models that takes a similar approach to the multi-step estimators for static factor models in Heckman et al. (2013) and Bakk et al. (2013). Our estimator shares similarities with Attanasio et al. (2015), though it differs from theirs in several important dimensions. In the first step, we use a large measurement system to non-parametrically recover the distribution of the latent skills. In particular, we implement a minorization-maximization algorithm to overcome dimensionality problems (i.e. recover a mixture of ten normals on a sample of approximately a half million students) and censoring of measurement variables.[4] In the second step, we use the estimates from the first stage to take draws

[2]Our data contain student performance on 70 different tests, of which only a subset are mandatory.

[3]Attanasio et al. (2017) also estimates a nested CES production function. We also relax the assumption of constant returns to scale imposed in Cunha et al. (2010); Attanasio et al. (2015).

[4]Attanasio et al. (2015) do not directly address problems that arise from non-continuous data and estimate a mixture of two normals.

from individual-specific distributions of the latent factors and estimate the structural parameters of the production function, treating these draws as observed data.[5] This second stage requires joint numerical optimization of eight production functions that each account for more than 23,000 neighborhood fixed effects and allows for correlation in the unobserved component of total factor productivity.

Our estimates of the production functions shed light on important aspects of math and verbal skill formation. Self-productivity plays a key role in the formation of skills with larger level effects in secondary school than in elementary school.[6] We also show that the relevance of cross-effects heavily depends on the type of skill.[7] While math skills have no impact on the production of verbal skills, we do find a positive effect of verbal skills on the production of math skills. Moreover, we find that school quality plays an important role in the formation of skills. In particular, we show that giving all students access to high quality schools would substantially increase math and verbal skills, increasing college enrollment in the population by 10%. Finally, we use counterfactual simulations to understand differences in education outcomes between students attending only low quality schools and students attending only high quality schools. Because students in low quality schools have lower levels of skills at the end of compulsory education they are 30 percentage points less likely to enroll in college. While 15% of this gap is due to differences in skill levels at the beginning of compulsory education, 20% of this gap is attributable to the differences in school quality, which indicates some scope for schools to overcome initial skill disadvantages.

The analysis of adult outcomes shows that verbal skills at the end of compulsory education have a substantially stronger influence on university enrollment and graduation than math skills. Specifically, we find that the marginal effect of verbal skills on college enrollment is more than twice as large as the effect of math skills. The larger impact of verbal skills is robust to different model specifications and accounts for multiple sources of endogeneity.[8] We provide evidence that this finding is *not* driven by conflating factors like socio-emotional skills or family background

[5] Attanasio et al. (2015) uses minimum distance to estimate the joint distribution of the latent factors and all other variables in the model, while our estimator follows a maximum likelihood approach.

[6] Self-productivity refers to the effect of math (verbal) skills at time t on math (verbal) skills at time $t+1$.

[7] Cross-effects refers to the effect of math (verbal) skills at time t on verbal (math) skills at time $t+1$.

[8] This result is consistent with the findings in Chetty et al. (2014), which shows that a high quality English teacher leads to a greater increase in students' college quality (almost twice) compared to a high quality math teacher.

characteristics. Furthermore, we confirm similar patterns with data from the United States. A possible explanation of this result is that students at or below the margin of college enrollment are less likely to enroll in STEM fields, therefore math skills may play less of a role in this subgroup of the population.[9]

Our finding on verbal skills and college enrollment adds to a large literature that studies the impact of math and verbal skills on wages and subject specific schooling curriculum on wages, which has primarily focused on the importance of math (Levine and Zimmerman, 1995; Rose and Betts, 2004; Joensen and Nielsen, 2009; Altonji et al., 2012; Dougherty et al., 2015). One consequence of our result is that if verbal skills have a predominant effect on college enrollment, then caution needs to be exercised when interpreting regressions of skills on labor market outcomes that also control for the endogenous variable education (Betts, 1995; Neal and Johnson, 1996). Regressions with skills and curriculum effects that control for years of education may mute one of the main channels through which verbal skills influence labor market outcomes, i.e. increasing formal education. In this paper, using data from the United States, we demonstrate that when the control for level of education is taken out of these regressions, the labor market return to verbal skills increases substantially. This alternative interpretation suggests that verbal skills may play a more important role in the labor market than previously thought.

Finally, we document that while gender differences in math skills are small, females have a large advantage in verbal skills. This fact, combined with our main result that verbal skills have a disproportionate influence on college enrollment suggests that gender differences in verbal skills are a key driver of the gender gap in college enrollment. Females' advantage in verbal skills also has implications for the gender gap in STEM majors, as we show that comparative advantage in math influences the decision to major in STEM. Since females and males have similar distributions of math skills, the male disadvantage in verbal skills translates to a comparative advantage in math, leading to an increase in male representation in STEM.

The rest of the paper is organized as follows. Section 2 describes the data and the institutional

[9] STEM fields require in general a strong background in math. Our data indicate that among those college enrollees that have a *below* average population probability of attending college, only 23% of them enrolled in STEM fields. While among those that have *above* average population probability, 31% enrolled in STEM fields.

setting of education in England. Sections 3 and 4 describe the empirical model and our estimation approach. Section 5 shows the main results regarding production function estimates and the importance of skills for educational attainment. Section 6 presents model simulation results. Section 7 explores external validity of our main results and discusses practical implications of the main findings for empirical research. Section 8 studies gender gaps in college enrollment and field of study. Section 9 concludes.

2 Institutional Setting and Data Summary

This section describes the institutional features of the English education system and the data we use in our analysis.

2.1 The English School System

Compulsory education in England is organized in four Key Stages (KS). Each stage ends with nationally assessed standardized tests, in addition to teacher assessments on different subjects.[10] Table 1 summarizes the stages of the English compulsory education system. Students enter school during the Foundation Stage, at age 4, then proceed to Key Stage 1 (KS1) during ages 5-6, and then to Key Stage 2 (KS2) during ages 7-11.[11] At the end of KS2, students move to secondary school, where they progress to Key Stage 3 (KS3, ages 12-14) and Key Stage 4 (KS4, ages 15-16). In KS4, students tailor their curriculum by specializing in six to eight subjects. At age 16, when compulsory education ends, students decide to either exit formal education or continue their studies for two more years, called A-levels (ages 17-18) where they choose a vocational or academic curriculum, which typically concludes with qualifying exams. Most students study three or four A-level subjects concurrently during year 12 and year 13, either in a secondary education institution or in a Sixth Form College. Finally, higher education usually begins at age 19 with a three-year bachelor's degree, where admissions to university are mainly determined by A-level performance.

[10] Recently, a series of reforms regarding the assessment of students have been implemented. However, these reforms were not in place for the years of our analysis.

[11] KS1 is equivalent to grades 1 and 2 in the US school system and KS2 to grades 3, 4, 5 and 6.

Table 1: Key Stages in English Education System

Stage	Age	Years	Test
Key Stage 1	5-7	1 and 2	National Program of Assessment at the end of year 2 in Math, English, and Science (carried out by the teacher) and annual teacher assessments in each subject.
Key Stage 2	8-11	3-6	National Program of Assessment at the end of year 6 in Math, English, and Science. Teacher assessment is also provided.
Key Stage 3	12-14	7-9	National Program of Assessment at the end of year 9 in Math, English, and Science. Teacher assessment is also provided.
Key Stage 4	15-16	10 and 11	General Certificate of Secondary Education (GCSE), generally taken at the end of year 11. End of compulsory education

2.2 Data

Our analysis uses individual-level administrative panel data for the cohort of students who completed their compulsory education in the academic year 2006/07. The final dataset contains information on approximately 500,000 students, which only excludes students in independent (i.e. private) schools because these schools are not covered in the census.[12] Our database links information from the census of all state (i.e. public) school children in England with information from the Higher Education Statistics Agency (HESA).[13] HESA collects information on all students in publicly funded universities. The dataset allows us to track pupils over their entire academic career, containing detailed information on student demographics; neighborhood characteristics and schools attended; exam performance, teacher assessments, and school absences; as well as post-secondary education outcomes. Overall, we observe on average 33 performance measures for each student out of a total of 70 possible measures. The difference occurs because students take different combinations of subject tests in KS4. However, the math and English subject tests are mandatory in KS4 for all students. In all, this provides us with about eight comprehensive measures of verbal and

[12] The independent sector educates around 6.5% of the total number of school children in the United Kingdom.

[13] The final census entails data from the National Pupil Database (NPD) and the Pupil Level Annual School Census (PLASC), that has been replaced in 2007 by the School Census.

math ability at each of the four Key Stages for each student.[14]

Table 2 presents summary statistics of the key variables in our data. The top panel shows information on student background characteristics for KS4, the last stage of compulsory education, though we also have similar information for earlier stages. First, we report the average Index of Multiple Deprivation (IMD), which indicates the degree of deprivation in the neighborhood where the student lives, with higher scores corresponding to more impoverished areas.[15] The IMD is a composite of seven indices that measure different forms of deprivation in a neighborhood. In our empirical analysis we use the individual domains rather than the composite to more robustly control for contextual factors.[16] The data also identify students who meet eligibility requirements for free school meals (FSM). According to Hobbs and Vignoles (2007), FSM status proxies for children in households with family incomes below £200 (US$300) per week. The special education needs (SEN) variable indicates whether a child has learning difficulties or disabilities. Overall, the data show that 11% of the students in KS4 are eligible for FSM, where differences between genders are small. On the contrary, the indicator for SEN shows that only 12% of female students are included in this category compared to 20% of male students. Finally, 95% of the students in our sample have a mother who speaks English, and more than 89% of students are white.

Pupils aged 5-16 in state schools must be taught the National Curriculum. The second panel of Table 2 shows performance in math and English at each Key Stage.[17] The scale of the math and English scores have been mapped into national levels of performance that are comparable across schooling years. The National Curriculum sets standards of achievement in each subject for pupils aged 5 to 14. These standards, for most subjects, range from Levels 1 through 8. For example, most 7 year olds are expected to achieve Level 2, most 11 year olds are expected to achieve Level 4, and most 14 year olds are expected to achieve Levels 5 or 6.[18] A similar mapping exists with

[14] This paper does not use the A-level information because only those students who are college bound will continue to A-level, producing selection bias issues.

[15] Neighborhood denotes a lower layer super output area (LSOA), which contains about 1,500 people, which is roughly equivalent to a census block group in the US. This information is calculated by the Department for Communities and Local Government in England.

[16] See Footnote 46 for a description of the individual domains and to see how they are used in the analysis.

[17] In Table 2 we only report performance in math and English, while in our analysis we use information on a wide range of test scores.

[18] Depending on their level of achievement, students can also be classified as performing *below expectation* or *above expectation*. A description of the mapping of achievement levels to these classifications

Table 2: Summary Statistics: Overall and by Gender

	All		Female		Male	
	Mean	Std.	Mean	Std.	Mean	Std.
	Background Characteristics (KS4)					
Index of Multiple Deprivation	22.05	15.92	22.27	16.03	21.84	15.80
Free School Meal	0.11	0.31	0.11	0.32	0.11	0.31
Special Education Needs	0.16	0.37	0.12	0.33	0.20	0.40
Mother Tongue English	0.95	0.22	0.95	0.23	0.95	0.22
	Racial Distribution					
White	89.5%		89.3%		89.6%	
Asian	5.7%		5.8%		5.7%	
Black	2.4%		2.5%		2.4%	
Other	2.4%		2.5%		2.3%	
	Key Stage (KS) Math Scores - National Curriculum Levels					
KS1	2.0	0.6	2.0	0.6	2.1	0.7
KS2	4.0	0.8	4.0	0.8	4.0	0.8
KS3	5.6	1.2	5.6	1.2	5.6	1.3
KS4	6.6	1.8	6.6	1.8	6.6	1.8
	Key Stage (KS) Verbal Scores - National Curriculum Levels					
KS1	1.8	0.6	1.9	0.6	1.7	0.6
KS2	4.0	0.8	4.1	0.8	3.9	0.8
KS3	5.2	1.0	5.3	1.0	5.0	1.0
KS4	6.8	1.6	7.1	1.5	6.6	1.6
	Average Number of GCSE (KS4) Exams Taken					
	8.1	2.0	8.3	1.9	7.9	2.0
	Average Proportion of Sessions Absent in KS4					
Authorized Absences	0.07	0.07	0.08	0.08	0.07	0.07
Unauthorized Absences	0.02	0.06	0.02	0.06	0.02	0.06
	Distribution of Primary School Quality (KS2)					
Inadequate	6.3%		6.3%		6.2%	
Satisfactory	36.9%		36.9%		36.9%	
Good	45.6%		45.6%		45.6%	
Outstanding	11.3%		11.2%		11.4%	
	Distribution of Secondary School Quality (KS4)					
Inadequate	8.0%		7.8%		8.3%	
Satisfactory	32.7%		31.5%		33.8%	
Good	43.4%		44.1%		42.7%	
Outstanding	15.9%		16.6%		15.2%	
	University Outcomes					
Enrollment	0.36	0.48	0.40	0.49	0.33	0.47
Graduation	0.25	0.43	0.29	0.45	0.22	0.41
Enrollment STEM	0.11	0.31	0.09	0.28	0.13	0.34
Graduation STEM	0.07	0.26	0.07	0.25	0.08	0.28
Enrollment Top 24	0.07	0.26	0.08	0.27	0.07	0.25

KS4 (i.e. GCSE) scores.[19] These levels make it possible to track the progress of students over time. For example, an eleven year old student that achieves level 2 in English while in KS2 shows the same proficiency level of an average seven year old student in KS1. Therefore, a subgroup of measures of performance in math and English in our database are reported on a scale that has a specific meaning allowing us to recover latent factors that share the same scale at different points in the schooling career.[20] Table 2 shows that females outperform males on English exams at each Key Stage, while there are almost no gender differences in math. To conclude, the remaining rows of the second panel indicate that students take on average 8.1 GCSE subject-exams, which includes the compulsory math and English exam, and that the average proportion of authorized and unauthorized school sessions absent are 7% and 2%, respectively.

The third panel of Table 2 reports the distribution of students by school effectiveness. School quality is determined by Ofsted reports, which are produced by a governmental office responsible for inspecting a range of educational institutions, including state schools. This office carries out regular inspections of each school in England, resulting in a published evaluation of the effectiveness of the schools inspected. Inspections generally consist of three-day visits, with two days' notice. They focus on examining how well the school is managed, and what processes are in place to ensure standards of teaching and improvement of learning.[21] After each inspection schools are classified as inadequate (1), satisfactory (2), good (3), or outstanding (4). The data indicate that 6.3% of primary school students and 8% of secondary school students attended inadequate schools.[22]

Finally, the last panel of Table 2 provides an overview of post-secondary education outcomes. Around 36% of the students in our sample enrolled in university, with females being 7 percentage

can be found at `http://webarchive.nationalarchives.gov.uk/20140109220530/http://www.education.gov.uk/schools/performance/archive/ks3_05/k5.shtml`.

[19] `http://webarchive.nationalarchives.gov.uk/20140109214956/http://www.education.gov.uk/schools/performance/archive/schools_10/s11.shtml` and the document "A Guide to Understanding Student Progress at KS4" extracted from `http://www.abbeygrangeacademy.co.uk/assets/key-stage-4-progress-information.pdf` suggest how to link Key Stage levels to GCSE scores, which involves rescaling KS4 scores to make them comparable over time.

[20] Agostinelli and Wiswall (2016) provide a thorough discussion on the importance of having access to variables that share the same scale over time when estimating dynamic factor models.

[21] Ofsted inspectors look at school inputs such as the quality of teaching, safety of pupils, and the quality of leadership and management. An adverse report may include a recommendation for further intervention in the running of the school.

[22] School quality is defined in this study as the average of the Ofsted reports between 2001 and 2007.

points more likely to enroll than males. On the other hand, males account for nearly 60% of total enrollment in STEM fields. Finally, the variable "University Enrollment Top 24" denotes the proportion of students attending the most selective institutions in the United Kingdom (the so-called Russell group).[23] Approximately 7% of students enroll in these institutions.

To conclude this section, we use simple regressions to describe two of the key aspects of the data we intend to study more deeply with our main analysis. Table 3 shows the average marginal effect of KS4 math and English achievement levels on the probability of university enrollment and university graduation from linear probability models estimated separately for males and females, which also include a robust set of controls for student characteristics (see note in Table 3). These results show that while both math and English test scores have an effect on university enrollment and graduation, English scores appear to have a slightly larger effect than math. While this analysis is informative, it has serious limitations beyond the usual possible problems of endogeneity because using raw levels of overall performance in math and English only proxies for skills, leading to problems of measurement error that are remarkably salient in our context. In particular, apart from the well known concerns with attenuation bias, Maddala (1992) shows that when two regressors are highly correlated and measured with similar error, then the ratio of their coefficients tend to 1 as the correlation goes to one, regardless of the true ratio.[24] This result suggests that the coefficients reported in Table 3 could be misleading in determining the relative importance of each skill in predicting educational outcomes.[25] While taking simple averages across the many measurements could help to alleviate this problem, it is not clear *a priori* which of the measurements (e.g. more than 30 in KS4) should be used to proxy for which skills, and any (naive) aggregation would impose arbitrary and strong assumptions on the weighting scheme that best captures skills.[26] Therefore,

[23]The Russell Group represents 24 leading UK universities: University of Birmingham, University of Bristol, University of Cambridge, Cardiff University, Durham University, University of Edinburgh, University of Exeter, University of Glasgow, Imperial College London, King's College London, University of Leeds, University of Liverpool, London School of Economics & Political Science, University of Manchester, Newcastle University, University of Nottingham, University of Oxford, Queen Mary University of London, Queen's University Belfast, University of Sheffield, University of Southampton, University College London, University of Warwick, University of York.

[24]See Appendix A for a formal proof of this proposition.

[25]We show in the results section of the paper that bias due to measurement error understates the role of verbal skills in educational attainment.

[26]Given the high correlation among these scores, a regression that includes all of them would be difficult to understand, with many of the coefficients possibly having the wrong sign due to multicollinearity.

Table 3: Linear Probability Model: University Outcomes

	Female		Male	
	Enrollment	Graduation	Enrollment	Graduation
Average Marginal Effect of Test Scores				
KS4 Math Achievement Level	0.075 (0.001)	0.062 (0.001)	0.068 (0.001)	0.048 (0.001)
KS4 Verbal Achievement Level	0.102 (0.001)	0.079 (0.001)	0.096 (0.001)	0.075 (0.001)
Observations	238,574		239,082	

Note: Additional controls include mother tongue, race, KS4 free school meal eligibility, subindices that comprise the index of multiple deprivation, KS4 special education needs, KS4 school absences, number of GCSE exams taken, and secondary school fixed effects. Standard errors are reported in parentheses.

we outline in the next section a dynamic factor model of skill formation that allows us to optimally extract skills from these many measures and provides a framework to comprehensibly analyze how skills are formed during the schooling career.

3 Dynamics of Skill Formation

To characterize how math and verbal skills develop during compulsory education, we estimate a dynamic factor model of skill formation similar to Cunha et al. (2010). This approach provides a natural framework to understand the role of cross effects, self-productivity, school quality, and complementarities in the production of skills at different stages of the schooling career. Moreover, with over 70 measures of student performance from the beginning of elementary school to the end of compulsory education, a dynamic factor framework facilitates an efficient use of this large data set.

We modify and expand the Cunha et al. (2010) framework in several dimensions. First, we propose and estimate a production function of skill formation that corresponds to a nested CES with three inputs, originally described in Sato (1967). The advantage of a two-level CES is that it allows the elasticities of substitution to vary across pairs of inputs. With the exception of Attanasio et al. (2017), nested CES production functions have not been widely estimated within this literature, however many papers in macroeconometrics (e.g., Acemoglu (1998); Krusell et al. (2000); Pandey (2008)) have used them where input factors need further differentiation (e.g. wage differentiation between skilled and unskilled labor). Second, we relax the assumption of constant returns to scale imposed in Cunha et al. (2010) and Attanasio et al. (2015) and directly estimate this parameter. Third, the model allows for total factor productivity (TFP) terms, which account for the remaining components of the skill production functions that are not captured by the inputs.[27] Fourth, we propose a novel two-step estimation method for dynamic factor models along the lines of the multi-step estimators for static factor models in Heckman et al. (2013) and Bakk et al. (2013). Our estimator shares similarities with Attanasio et al. (2015), though our approach differs from them in several important dimensions as we describe in Section 3.4. Finally, each period of skill development in our model maps to a specific stage of compulsory education in England (i.e. Key Stages) providing clear guidance to distinguish the different periods of skill formation.

3.1 Model

During compulsory education, students take a series of national tests at different Key Stages that represent periods of skill development in our model.[28] These tests provide noisy information on the students' math (m) and verbal (v) skills.[29] Time is indexed by t where $t \in \{1, 2, 3, 4\}$ represents each of the four Key Stages of compulsory education. At time period t, skill $k \in \{m, v\}$ for student i is denoted as $\Theta^k_{i,t}$, where $\Theta^k_{i,t} > 0$. We assume skills for $t \in \{2, 3, 4\}$ develop following a nested CES

[27] Agostinelli and Wiswall (2016) describe the conditions under which a TFP term can be identified. Attanasio et al. (2015) also model a TFP term.

[28] Table 1 reports the school grade of the students at each Key Stage.

[29] As we described in Section 2 certain measures of skills are on a scale that is comparable over time (i.e. National Curriculum levels).

production function with three inputs, while $t = 1$ corresponds to the initial conditions period:[30]

$$\Theta_{i,t}^k = A_{it}^k \left[\delta_t^k \left(\Omega_{it}^k\right)^{\rho_t^k} + (1-\delta_t^k)\,(Q_{it})^{\rho_t^k}\right]^{r_t^k/\rho_t^k} \tag{1}$$

where

$$\Omega_{it}^k = \left(\alpha_t^k \left(\Theta_{i,t-1}^m\right)^{\gamma_t^k} + (1-\alpha_t^k)\left(\Theta_{i,t-1}^v\right)^{\gamma_t^k}\right)^{1/\gamma_t^k}$$

$$A_{it}^k = \exp\left(x_{it}'\psi_t^k + \pi_i^k + \nu_{it}^k\right)$$

$$\delta_t^k \in [0,1];\ \alpha_t^k \in [0,1];\ \rho_t^k \le 1;\ \gamma_t^k \le 1$$

(production function)

Equation (1) corresponds to a 3-input nested CES function. The skill aggregator Ω_{it}^k represents the inner CES nest, which describes how the prior period's stock of math and verbal skills contribute to the production of skills in period t. The outer CES nest is a function of the skill aggregator Ω_{it}^k and the quality of the school attended by the student in period t, Q_{it}.[31] The main advantage of the nested CES relative to a traditional CES with multiple inputs is that it allows the elasticity of substitution to differ between the three inputs rather than imposing a single elasticity of substitution for all inputs.[32] We include school quality as an input in the production functions for two reasons. First, schools are an important contributor to the development of math and verbal skills (Hanushek, 2005). Second, using the estimates from our model, we are interested in performing counterfactual simulations on relevant policy parameters, and school quality is a natural instrument that policymakers can influence.

The remaining components contributing to skill production that are not captured by our three main inputs are represented by the total factor productivity (TFP) term A_{it}^k. TFP is individual, time, and skill specific, and is a function of both observed and unobserved variables. First, a vector of observed variables, x_{it}, which includes student background characteristics (i.e. race, mother

[30] While gender subscripts are omitted in Eq. (1), we estimate production functions separately for males and females.

[31] School quality, Q_{it}, is obtained from school report cards developed by Ofsted.

[32] For example, in the nested CES the substitutability of math skill and school quality is allowed to differ from the substitutability of math skills and verbal skills, while in the traditional CES with multiple inputs the elasticity of substitution is assumed to be the same.

tongue, free school meal eligibility, special education needs) and neighborhood fixed effects (nearly 23,000). Next, the term π_i^k represents an unobserved component of TFP that is persistent over time and correlated across skills. Similar to Cunha et al. (2010), this term accounts for endogeneity between the inputs and the unobserved component of TFP. Finally, the last term in the TFP, ν_{it}^k, represents an idiosyncratic shock to skill production that is independent across skills and time.

Following the literature (Cunha et al., 2010; Attanasio et al., 2015; Agostinelli and Wiswall, 2016), we estimate the log of the production function in Eq.(1), so it is convenient to define the natural log of the skills as $\theta_{i,t}^k$ (i.e. $\theta_{i,t}^k = \ln \Theta_{i,t}^k$).

3.2 Measurement System

The main challenge in estimating the parameters of our model is that the skills we aim to study are not directly measured. We overcome this problem by using a factor model approach that allows us to extract these unobserved skills from a large set of observed data. One shortcoming of factor models is that the factors have no natural scale, making interpretation of results challenging. A second issue relevant to dynamic factor models is whether the factors can be compared over time. An important feature of our data that allows us to address these two issues is that at least one of our measures for each skill at each Key Stage is scaled based on the Department of Education's National Curriculum scale. This grading system is designed in a specific way so that test scores are comparable across Key Stages, and furthermore, their numerical values have a tangible interpretation (i.e. *below expectations, at level expected, or above expectations*). Normalizing the log of our factors to this National Curriculum scale addresses both of the previously stated concerns with factor models. First, it facilitates valid comparison of our estimated skills over the schooling career (Agostinelli and Wiswall, 2016). Second, it offers a clear-cut interpretation of the factors. For example, we will be able to study how moving students to higher quality schools impacts their ability to perform at *expected level*.[33]

Let $\theta_i = \begin{bmatrix} \theta_{i,1}^m & \theta_{i,1}^v & \cdots & \theta_{i,4}^m & \theta_{i,4}^v \end{bmatrix}'$ represent student i's complete vector of realized log math

[33] Alternatively, if the factors were normalized to a measure that lacked an interpretable scale, for example an exam with arbitrary scale, then we would need to anchor our factors to an interpretable adult outcome as in Cunha et al. (2010). However, in our analysis anchoring is not necessary since our factors are interpretable by themselves based on the standards set out by the Department of Education's National Curriculum scale.

and verbal factors. While these factors are not observed, our data contain frequent and extensive measures for each student that we use to recover the latent math and verbal skills, $\theta^k_{i,t}$, in each period. Let w^*_{ij} for $j = 1, 2, \ldots, J$ denote the J measures for individual i. Each measure is determined by:

$$w^*_{ij} = \mu_j + \lambda'_j \theta_i + \eta_{ij}$$

(measurement equation)

Where μ_j is the mean of the j^{th} measure and λ_j contains the loadings on the factors for measurement j. Our factors span multiple time periods, so only the factors that are associated with a particular skill and time period will have non-zero loadings on the measurements. Finally, η_{ij} is the remaining portion of the measurement that is not explained by the factors and is assumed to be independent and normally distributed with mean zero and variance σ^2_j.

Identification of factor models requires normalizations to set the scale and the location of the factors. As stated earlier, we set the scale and location for the eight factors by normalizing the constant equal to zero and the factor loading equal to one for the eight measures that are given on the Department of Education's National Curriculum scale, which normalizes our factors on this scale and allows our factors to be comparable across skills and across time. In our measurement system we use at least three dedicated measures per period that only load on one factor. For example in Key Stage 1 we observe for each student, scores in reading, writing, and spelling as well as a Key Stage 1 English teacher assessment. These four noisy measures only load on the student's latent verbal skills in Key Stage 1.[34] Appendix E, Tables 17 and 18, provides a full list of our measurements and the normalizations we make on the factor loadings.

[34]While many of our measures load only on a single factor, we allow several measures to load on multiple factors (e.g. geography), when *a priori* it is not straightforward to determine which skills these measures should load on.

3.2.1 Conditional Probabilities

A subgroup of measurements in our database cannot be characterized as continuous variables given that they only take a limited number of values.[35] Unfortunately, this type of data is generally inconsistent with our model. Given that w_{ij}^* is a linear function of the factors plus continuous measurement error, it is unlikely that this model could generate a distribution with a finite number of mass points. As we discuss later in the estimation section, our approach requires us to allow for a non-parametric distribution of the factors. Therefore this type of discrete data needs to be treated carefully in estimation by recognizing that its mass points occur through the construction of the variable, not because the underlying distribution of the factors is discrete. In this regard, we treat these variables as interval censored data.

For each j, we denote w_{ij} as the observed measure for individual i, which may differ from the true measure w_{ij}^* in the measurement equation. For estimation, we treat all measures that have more than 10 finite values as continuous variables. Let $\mathcal{J}_{cont}$ denote the set of observed continuous measures. For these measures we assume $w_{ij} = w_{ij}^*$, thus the probability of an observed measure conditional on a given value of the factors can be written as:

$$\Pr(w_{ij}|\theta) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{\left(w_{ij} - \mu_j - \lambda_j'\theta\right)^2}{2\sigma_j^2}\right) \qquad \text{for } j \in \mathcal{J}_{cont}$$

The remaining set of measures, $\mathcal{J}_{cens}$, which all have 10 or less unique values, are treated as interval censored data, where the observed value w_{ij} is a lower bound on the true measurement w_{ij}^*. Constructing the likelihood for these data depends on the number of censoring points, G_j, and the spread of the observed values. Let $\{\zeta_1^j, \zeta_2^j, \ldots, \zeta_{G_j}^j\}$ denote the observed values of measure j. If $w_{ij} = \zeta_g^j$, then we assume this indicates that $w_{ij}^* \in [\zeta_g^j, \zeta_{g+1}^j)$. Thus for $j \in \mathcal{J}_{cens}$ we define the

[35] Figure 3 in Appendix B illustrates this case with specific examples. Notice that this type of data is very common in many data sets, for example any data with letter grades A–F.

probability of the observed measure as:

$$\Pr(w_{ij}|\theta) = \begin{cases} \Phi\left(\frac{\zeta_2^j - \mu_j - \lambda_j'\theta}{\sigma_j}\right) & \text{if } w_{ij} = \zeta_1 \\ \Phi\left(\frac{\zeta_{g+1}^j - \mu_j - \lambda_j'\theta}{\sigma_j}\right) - \Phi\left(\frac{\zeta_g^j - \mu_j - \lambda_j'\theta}{\sigma_j}\right) & \text{if } w_{ij} \in \left\{\zeta_2, \ldots, \zeta_{(G_j-1)}\right\} \\ 1 - \Phi\left(\frac{\zeta_{G_j}^j - \mu_j - \lambda_j'\theta}{\sigma_j}\right) & \text{if } w_{ij} = \zeta_{G_j} \end{cases}$$

Where $\Phi(\cdot)$ corresponds to the cumulative distribution function of a standard normal distribution. Importantly, even though w_{ij}^* is not observed directly, since the censoring points are known *a priori*, we are able to recover both the factor loadings and the scale of the measurement error for these measurement equations.

3.3 Selection Correction

As mentioned, while the math and English tests in Key Stage 4 are mandatory, students have the opportunity to tailor their curriculum by specializing in six to eight additional subjects. This means that not all outcomes are observed for all students during this period of compulsory schooling because test scores are only observed for the subjects in which a student enrolls.[36] Since the score on these subject tests will be used to recover the latent skills, this creates a potential selection problem.

We address this potential problem by modeling Key Stage 4 course selection directly and including a selection correction term in our likelihood. Since we observe multiple subject choices for each student, this allows us to recover an additional level of student heterogeneity in Key Stage 4, which we label $\theta_{i,4}^{sel}$. This additional factor can be estimated under arbitrary correlation with the other factors and is identified from residual correlation in subject choices that cannot be explained by selection on math and verbal skills. For further identification, we use the number of excused and unexcused absences in Key Stage 4 as additional measures that only load on this additional factor. This selection factor serves two empirical purposes. First it allows us to account for selection in the

[36] This is not an issue for Key Stages 1 to 3 because for these periods we have nearly universal coverage of these measures. For the few individuals where some of the Key Stage 1 to 3 measurements are missing, we assume it is missing at random conditional on observables.

estimation of the parameters. Second, it serves as an additional control to be included in the TFP in Key Stage 4 to address potential omitted variable bias in the estimation of the KS4 production functions.

Let $a_{ic} = 1$ if student i enrolls in subject c in Key Stage 4. Students have a total of $C = 36$ possible subjects in which to enroll. Since subject selection occurs at the beginning of Key Stage 4 we assume that these decisions are a function of Key Stage 3 skills in addition to the selection factor. We assume the probability of choosing a subject, which will be included as a selection correction term in our likelihood, can be represented with a conditional logit:

$$\Pr(a_{ic}|\theta) = \frac{\exp(\tau_{c0} + \tau_{c1}\theta_3^m + \tau_{c2}\theta_3^v + \tau_{c4}\theta_4^{sel})^{a_{ic}}}{1 + \exp(\tau_{c0} + \tau_{c1}\theta_3^m + \tau_{c2}\theta_3^v + \tau_{c4}\theta_4^{sel})}$$

3.4 Two-Step Estimation

Our estimation strategy for the dynamic factor model is motivated by the multi-step approaches for estimating static factor models in Heckman et al. (2013) and Bakk et al. (2013). It also shares similarities to the estimator proposed by Attanasio et al. (2015).[37] To summarize our two-step estimator, in the first step, we use a large measurement system to non-parametrically recover the distribution of the latent skills. In the second step, we estimate the parameters of the production function using draws from the factor distribution obtained in the first step as observed data.

Let $\theta \in \{\theta_1^m, \ldots, \theta_4^v, \theta_4^{sel}\}$ denote a possible realization of the vector of the unobserved log factors. We can write the probability of the observed measures and observed Key Stage 4 subject choices conditional on θ as $p(w_i, a_i|\theta) = \prod_{j=1}^{J} \Pr(w_{ij}|\theta) \prod_{c=1}^{C} \Pr(a_{ic}|\theta)$. Given our model, each of these probabilities are conditionally independent once we account for the unobserved factors. The production function outlined in Section 3.1 gives rise to a functional form of the conditional distribution of the factors. Consider $F(\theta|x_i, Q_i)$ as the conditional probability of the log of the

[37] Attanasio et al. (2015) proposes a three stage estimator. Our approach differs from theirs in a number of ways. First, we combine the first and second stage of Attanasio et al. (2015) into a single step. Second, a step of their estimation algorithm relies on a minimum distance estimator while we use maximum likelihood in all stages leading to efficiency gains. Third, our algorithm directly handles issues that arise from including non-continuous variables in the measurement system. Fourth, our approach can easily accommodate discrete variables and a large set of covariates (including a large vector of neighborhood fixed effects) given that it does not require us to model the entire distribution of the population characteristics. Fifth, we allow for a larger number of components in our Gaussian mixture distribution when recovering the non-parametric distribution of the latent factors.

factors based on the production functions in Eq.(1), where x_i is the vector of observed covariates in TFP, and Q_i is student i's vector of school quality indices for the schools which she attended from Key Stage 1 to Key Stage 4. Since the transitory productivity shocks, ν are i.i.d., this joint probability can be written as the product of marginal probabilities integrating over the multi-dimensional vector of unobserved persistent productivity shocks π:

$$F(\theta|x_i, Q_i) = \int_{\pi} f(\theta_4^v|\theta_3^m, \theta_3^v, x_{i4}, Q_{i4}, \pi) \cdots f(\theta_1^m|x_{i1}, Q_{i1}, \pi)p(\pi)d\pi \tag{2}$$

Where $f(\theta_t^k|\cdot)$ refers to the marginal distribution of the log of skill θ_t^k and $p(\pi)$ corresponds to the probability density function of π, which we parameterize in Section 3.4.2.

Since θ is not observed, estimation will be based on the maximization of an integrated likelihood function. Given observations of n individuals and letting Ψ denote the parameters of the model, the log-likelihood is defined as:

$$\begin{aligned} LL(\Psi) &= \sum_{i=1}^{n} \ln L(w_i, a_i, x_i, Q_i) \\ &= \sum_{i=1}^{n} \ln \left[\int_{\theta} p(w_i, a_i|\theta) F(\theta|x_i, Q_i) d\theta \right] \end{aligned} \tag{3}$$

Rather than direct maximization of Eq. (3), we undertake a two-step estimation approach. Our two-step estimator separates the estimation of the measurement system and production function into two stages, which leads to a simple and straightforward estimation procedure. In the first stage, we non-parametrically estimate the unconditional distribution of the log factors, $p(\theta)$, along with the parameters of the measurement system. In the second stage, we construct for each individual $h(\theta|w_i, a_i)$ which is the conditional distribution of individual i's unobserved log factors conditional on the observed measures and estimate the parameters of the production function using maximum likelihood, integrating over these densities:

$$LL_{\substack{\text{second}\\ \text{stage}}} = \sum_{i=1}^{n} \int_{\theta} \ln \left[F(\theta|x_i, Q_i) \right] h(\theta|w_i, a_i) d\theta \tag{4}$$

Where the integral is outside of the log function, which drastically simplifies the estimation of the production function parameters.

The second stage of our estimator centers on the construction of the conditional density of the unobserved factors given the individual's observed measures and their subject choices in Key Stage 4. Letting $p(\theta)$ represent the unconditional distribution of the unobserved factors, the conditional density is given by:

$$h(\theta|w_i, a_i) = \frac{p(w_i, a_i|\theta)p(\theta)}{p(w_i, a_i)} \tag{5}$$

Where $p(w_i, a_i) = \int_\theta p(w_i, a_i|\theta)p(\theta)d\theta$. To see how this density is used in estimation, we re-write the log-likelihood in Eq. (3) replacing $F(\theta|x_i, Q_i) = p(x_i, Q_i|\theta)p(\theta)/p(x_i, Q_i)$, which yields:[38]

$$\begin{aligned}
LL &= \sum_{i=1}^{n} \ln\left[\int_\theta p(w_i, a_i|\theta)p(x_i, Q_i|\theta)p(\theta)/p(x_i, Q_i)d\theta\right] \\
&= \sum_{i=1}^{n} \ln\left[\int_\theta h(\theta|w_i, a_i)p(w_i, a_i)p(x_i, Q_i|\theta)/p(x_i, Q_i)d\theta\right] \\
&= \sum_{i=1}^{n} \ln\left[p(w_i, a_i)\right] + \sum_{i=1}^{n} \ln\left[\int_\theta p(x_i, Q_i|\theta)h(\theta|w_i, a_i)d\theta\right] - \sum_{i=1}^{n} \ln\left[p(x_i, Q_i)\right] \\
&= \underbrace{\sum_{i=1}^{n} \ln\left[\int_\theta p(w_i, a_i|\theta)p(\theta)d\theta\right]}_{LL_{\text{first stage}}} + \sum_{i=1}^{n} \ln\left[\int_\theta p(x_i, Q_i|\theta)h(\theta|w_i, a_i)d\theta\right] - \sum_{i=1}^{n} \ln\left[p(x_i, Q_i)\right] \qquad (6)
\end{aligned}$$

Equation (6) serves as the basis for the first stage of our two-step estimation approach. Because of the additive separability of the log-likelihood, this equation shows that the parameters of the measurement system as well as the unconditional distribution of the factors can be estimated by only considering the first element in this function. While this estimator is unbiased and consistent, as pointed out in Heckman et al. (2013) for example, it is less efficient because it ignores the information in the second component of the log-likelihood function.[39] Specifically, these estimates are not the full information maximum likelihood estimates because they neither incorporate all of the functional form assumptions of our production function nor the additional information that x_i and Q_i may

[38] This occurs from writing the joint distribution using the conditional distributions $p(\theta, x, Q) = p(x, Q|\theta)p(\theta) = p(\theta|x, Q)p(x, Q)$.

[39] The third component is a constant which drops out in maximization.

contribute to the identification of the factors, θ. Although this partial likelihood approach is less efficient, the fact that the first stage does not rely on any functional form assumptions of the production function could be viewed as an advantage. Specifically, identification of the latent factors is more transparent, since they are exclusively estimated based on the measurement system and are robust to any functional form assumption on the production function. If the functional form assumptions of the production function were imposed in the estimation of the factor model, not only would the factors have to be re-estimated for each model, which would be computationally intensive, but also each model would potentially be recovering different latent factors making comparisons across models difficult (Bakk et al., 2013).

Once the parameters of the measurement system and the parameters of the unconditional distribution of the factors, $p(\theta)$, are estimated in the first stage, they are used to form the conditional distribution functions in Eq. (5). In the second stage, these densities are used to estimate the parameters of the production function by maximizing Eq. (4). To show how Eq. (4) arises within the context of our empirical strategy, it is necessary to revisit Eq. (3). The maximum likelihood estimate of these parameters is a root of the score function:

$$\begin{aligned}\frac{\partial LL(\Psi)}{\partial \Psi} &= \sum_{i=1}^{n} \frac{1}{L(w_i, a_i, x_i, Q_i)} \int_\theta \frac{\partial \left[p(w_i, a_i|\theta)F(\theta|x_i, Q_i)\right]}{\partial \Psi} d\theta \\ &= \sum_{i=1}^{n} \int_\theta \frac{\partial \ln \left[p(w_i, a_i|\theta)F(\theta|x_i, Q_i)\right]}{\partial \Psi} \frac{p(w_i, a_i|\theta)F(\theta|x_i, Q_i)}{L(w_i, a_i, x_i, Q_i)} d\theta \\ &= \sum_{i=1}^{n} \int_\theta \frac{\partial \ln \left[p(w_i, a_i|\theta)F(\theta|x_i, Q_i)\right]}{\partial \Psi} h(\theta|w_i, a_i, x_i, Q_i) d\theta \\ &= \sum_{i=1}^{n} \int_\theta \frac{\partial \ln \left[p(w_i, a_i|\theta)\right]}{\partial \Psi} h(\theta|w_i, a_i, x_i, Q_i) d\theta + \underbrace{\sum_{i=1}^{n} \int_\theta \frac{\partial \ln \left[F(\theta|x_i, Q_i)\right]}{\partial \Psi} h(\theta|w_i, a_i, x_i, Q_i) d\theta}_{\text{second stage score}}\end{aligned}$$

The density function $h(\theta|w_i, a_i, x_i, Q_i)$ is the conditional density of the unobserved factors conditional on all of the data and the functional form assumptions of the model. Similar to Heckman et al. (2013), our two-step estimation approach is based on the conditional independence assumption that $h(\theta|w_i, a_i, x_i, Q_i) = h(\theta|w_i, a_i)$, that is, once we condition on our observed measurements and selection correction variables, the remaining observed data is independent of the factors.[40] Since $h(\theta|w_i, a_i)$ is recovered in the first stage, this leads to the likelihood problem for the production function parameters in Eq. (4).

The following subsections provide further details regarding implementation of the two-step approach, including discussions of functional form, identification, and endogeneity.

3.4.1 First Stage

Our goal in the first stage is to use the measurement system along with the selection correction term to recover the distribution of the factors in the population. To place minimal parametric assumptions on the distribution of the log factors, $p(\theta)$, we assume that they can be approximated by a mixture of $D = 10$ multivariate normal distributions. Let $K_d(\theta) = K(\theta|\xi_d, \Delta_d)$ denote the probability density function of a multivariate normal distribution with mean ξ and full covariance Δ. We assume that θ_i is drawn from $K_d(\cdot)$ with probability κ_d. Let $\Psi_A \in \{\mu, \lambda, \sigma, \tau, \kappa, \xi, \Delta\}$ denote the parameters to be estimated in the first stage, which includes all of the parameters of the measurement system, the selection correction parameters, as well as the distribution of the factors. Given the functional form assumptions, we estimate these parameters from the data by maximizing the log-likelihood function:

$$\begin{aligned} LL_{\substack{\text{first} \\ \text{stage}}}(\Psi_A) &= \sum_{i=1}^{n} \ln \left[\int_{\theta} p(w_i, a_i|\theta) p(\theta) d\theta \right] \\ &= \sum_{i=1}^{n} \ln \left[\int_{\theta} \left(\prod_{j=1}^{J} \Pr(w_{ij}|\theta) \prod_{c=1}^{C} \Pr(a_{ic}|\theta) \right) p(\theta) d\theta \right] \\ &= \sum_{i=1}^{n} \ln \left[\sum_{d=1}^{D} \kappa_d \int_{\theta} \left(\prod_{j=1}^{J} \Pr(w_{ij}|\theta) \prod_{c=1}^{C} \Pr(a_{ic}|\theta) \right) K_d(\theta) d\theta \right] \end{aligned} \tag{7}$$

Because our measures contain interval censored data and selection equations, maximizing Eq. (7) is quite challenging because the integral does not have a closed form solution as it would in a factor

[40] This assumption follows Heckman et al. (2013), equation A.3. The conditional independence assumption is not necessary for our two-step estimator. In general, without the conditional independence assumption, in the first stage we would include x and Q non-parametrically into the conditional distribution of θ as $p(\theta|x, Q)$, in which our first stage density would become $h(\theta|w, a, x, Q) \propto p(w, a|\theta)p(\theta|x, Q)$.

model with all continuous measures. We overcome this computational challenge by proposing a simple iterative routine based on the minorization-maximization (MM) algorithm developed in James (2017), which incorporates the results in Stewart (1983) to address the interval coded data.[41] The steps of the MM algorithm are outlined in Appendix C.

3.4.2 Second Stage and Endogeneity

Once the parameters of the measurement system and the distribution of the factors have been estimated in the first stage, in the second stage we form the conditional distributions in Eq. (5) that are used to estimate the parameters of the production function in Eq. (4). We estimate the parameters of the production function through maximum likelihood, which requires distributional assumptions about the unobserved random variables, π and ν, as well as assumptions about the initial conditions. We model the initial conditions as:

$$\theta_{i,1}^{k} = x_{i1}' \psi_1^k + \delta_1^k Q_{i,1} + \pi_i^1 + \nu_{i1}^k$$

Similar to the production functions, the initial conditions include observables x_{i1} and the quality of the school attended in Key Stage 1, Q_{i1}.[42] We augment the vector of unobservables, $\pi_i = \{\pi_i^1, \pi_i^m, \pi_i^v\}$ with an additional unobservable π_i^1 that is present in the unobserved component of the initial conditions, which in the estimation is allowed to be arbitrarily correlated with the unobserved variables π^k in the future TFP. Finally, ν represents the idiosyncratic component of the initial conditions.

Assuming ν comes from a mean zero normal distribution with skill and time specific variance, $\nu_{it}^k \sim N(0, \varsigma_{tk}^2)$ for $t \in \{1, 2, 3, 4\}$ and $k \in \{m, v\}$, conditional on $\pi = \{\pi^1, \pi^m, \pi^v\}$, the probability

[41] Both of these methods are based on the expectation-maximization algorithm (Dempster et al., 1977).
[42] The vector x_{i1} only includes race and mother tongue due to data availability in KS1.

density functions $f(\theta_t^k|\cdot)$ defined in Eq.(2) are:

$$f(\theta_t^k|\theta_{t-1}^m, \theta_{t-1}^v, x_{it}, Q_{it}, \pi) = \frac{1}{\sqrt{2\pi\varsigma_{tk}^2}} \exp\left(-\frac{\left(\theta_t^k - \hat{\theta_t^k}\right)^2}{2\varsigma_{tk}^2}\right)$$

Where, $\hat{\theta_t^k} =$

$$\begin{cases} x_{i1}'\psi_1^k + \delta_1^k Q_{i,1} + \pi_i^1 \qquad \text{for } t = 1 \\ x_{it}'\psi_t^k + \pi^k + \frac{r_t^k}{\rho_t^k} \ln\left[\delta_t^k \left(\alpha_t^k \left(e^{\theta_{i,t-1}^m}\right)^{\gamma_t^k} + \left(1-\alpha_t^k\right)\left(e^{\theta_{i,t-1}^v}\right)^{\gamma_t^k}\right)^{\rho_t^k/\gamma_t^k} + \left(1-\delta_t^k\right)(Q_{it})^{\rho_t^k}\right] \\ \qquad \text{for } t \in \{2,3,4\} \end{cases}$$

We assume that the multi-dimensional vector of unobservables π is drawn from a latent class distribution, such that $\pi_i = \{\pi_u^1, \pi_u^m, \pi_u^v\}$ with probability ϕ_u for $u = 1, \ldots, U$.[43] Because of the non-linearities of the CES production function, the integral in Eq. (4) does not have a closed-form solution and must be simulated. We approximate it with $R = 10$ simulated draws for each individual using a Metropolis-Hastings algorithm.[44] Let $\hat{\theta}_{ir}$ be the rth draw of θ for individual i from $h(\theta|w_i, a_i, \hat{\Psi}_A)$. In the second stage we maximize:

$$LL_{\substack{\text{second}\\\text{stage}}} = \sum_{i=1}^{n} \frac{1}{R} \sum_{r=1}^{R} \ln\left[\sum_{u=1}^{U} \phi_u f(\theta_{ir,4}^v|\theta_{ir,3}^m, \theta_{ir,3}^v, x_{i4}, Q_{i4}, \pi_u) \cdots f(\theta_{ir,1}^m|x_{i1}, Q_{i1}, \pi_u)\right]$$

One of the features of our two-step estimator is that if the production functions have no unobserved heterogeneity, π, then the second stage log-likelihood becomes additively separable, such that maximizing this likelihood can be done by maximizing each production function separately. However, our goal is to recover the structural parameters of a dynamic system, and failing to account for correlation in the TFP across skills and time may produce biased estimates of the production function. We address this problem in two ways. First, we include a rich set of controls, x_{it}, that

[43] In estimation we allow $U = 5$ and normalize $\pi_u = \{0, 0, 0\}$ for $u = 1$ since the covariates in TFP contain a constant.

[44] Because the integral is outside of the log operator, simulation does not produce bias. However, we use 10 draws to reduce simulation error which affects the standard errors.

represent the observed component of TFP (i.e. race, mother tongue, free school meal eligibility, special education needs, and 23,000 neighborhood fixed effects in each period).[45] Second, following Cunha et al. (2010), we allow for correlation in the unobserved component of TFP through π. This accommodates contemporaneous correlation in the unobserved production function error because $\text{Cov}(\pi^m, \pi^v) \neq 0$. Furthermore it allows for time-series correlation in the unobserved component of TFP through the presence of π^k in each of the equations for skill k. Finally, our estimation accounts for endogeneity of the initial conditions by allowing for correlation in π^1 and $\{\pi^m, \pi^v\}$.

Identification of the unobserved heterogeneity is achieved through the fact that the production function only depends on a finite number of lags. Thus, we can identify π^m through the covariance of the KS2 math and the KS4 math, as KS2 math has no direct effect on KS4 math once we condition on KS3 math. Likewise, identification of π^v is achieved through the covariance of KS2 verbal and KS4 verbal. Identification of the $\text{Cov}(\pi^m, \pi^v)$ is achieved through the covariance of KS2 math and KS4 verbal as well as KS2 verbal and KS4 math. The covariance between these variables conditional on KS3 math and verbal can be attributed to the covariance of π^m and π^v. Similarly, identification of $\text{Cov}(\pi^1, \pi^m)$ and $\text{Cov}(\pi^1, \pi^v)$ comes from the covariance of KS3 math and KS4 math with KS1 math as well as the covariance of KS3 verbal and KS4 verbal with KS1 verbal. Finally, π^1 is identified from the covariance of the KS1 math and verbal factors.

4 Adult Outcomes

Our next objective is to relate adult outcomes to the stock of skills at the end of compulsory education. By doing so, we will be able to analyze the differential role of math and verbal skills in explaining post-secondary education outcomes. In particular, we study university enrollment, university quality, major field of study, and university graduation. We characterize the level of human capital that is responsible for the production of these outcomes as a flexible function of the stock of skills at the end of compulsory education, $\theta^k_{i,4}$.

After completing Key Stage 4, students can finish their formal education or continue their

[45] The fact that we control for neighborhood fixed effects is important to properly identify the effect of school quality on test scores.

studies. We are interested in understanding how the latent skills derived from the factor model influence these decisions. There are S outcomes for each individual, with realization y^*_{is} for $s = 1, 2, \ldots, S$ that follows

$$y^*_{is} = z'_i\omega_s + \begin{bmatrix}\theta^m_{i,4} & (\theta^m_{i,4})^2 & \theta^v_{i,4} & (\theta^v_{i,4})^2 & \theta^m_{i,4}\theta^v_{i,4}\end{bmatrix}\beta_s + \varphi_s(\pi_i) + \varepsilon_{is} \qquad \text{(outcome equation)}$$

In the outcome equation, β_s captures the main effects of math and verbal skills for outcome s (e.g. college enrollment). These coefficients could support multiple economic interpretations. However, we interpret them as capturing the differential role that skills may have on the "psychic cost" related to schooling progression. Specifically, we argue that individuals lacking these skills may have a more challenging time moving into more advanced educational levels. A growing literature supports this view by establishing the relevance of psychic costs in explaining why many students do not continue their schooling, even though it is financially convenient for them to do so. In particular, Cunha et al. (2005, 2006a,b); Cunha and Heckman (2008a); Heckman et al. (2006) establish that these costs are related to cognitive and/or socio-emotional skills. Therefore, to the extent that verbal skills have a larger impact on educational attainment than math skills, this result could reflect their larger role in shaping psychic costs associated with educational decisions.

We allow for non-linearities in the effect of skills by including interaction and squared terms. The parameters ω represent the influence of our main control variables, which include observed covariates, z_i, that contain race, mother tongue, free school meal eligibility, special education needs, number of GCSE exams taken, excused and unexcused absences, eight neighborhood characteristic variables, and school Ofsted score in Key Stage 4 or school fixed effects depending the specification.[46] The last two terms in the outcome equation, $\varphi_s(\pi_i)$ and ε_{is}, represent the unobserved component. The first term, $\varphi_s(\pi_i)$, is equal to φ_{su} if individual i's unobserved persistent TFP

[46]The neighborhood characteristic variables correspond to those that conform to the different domains of the index of multiple deprivation. This data is calculated by the Department for Communities and Local Government and covers seven areas: Income Deprivation, Employment Deprivation, Health Deprivation and Disability, Education Skills and Training Deprivation, Barriers to Housing and Services, Living Environment Deprivation, and Crime. We also use an additional domain which is the IDACI index (i.e. income deprivation affecting children index). IDACI is a subcategory of the overall Income Deprivation index. We do not include neighborhood fixed effects in adult outcomes specifications due to the fact that many neighborhoods do not have students attending college, a specific field of study or type of university (e.g. selective). Therefore, it is not easy to handle a large set of fixed effects when many of them cannot be identified.

shocks belong to class u, i.e., $\pi_i = \pi_u$. It operates as a class specific intercept for each of the outcome equations. The final term, ε_{is}, represents the remaining determinants of the outcome variable that cannot be explained by the other parts of the model and is assumed to be independent.

4.1 Estimation

Each of the outcomes that we aim to study are discrete, so y^*_{is} represents the underlying latent variable process. We assume that ε is distributed type-I extreme value, and we only observe the outcome $y_{is} = 1$ if $y^*_{is} > 0$ and zero otherwise. Conditional on z_i, θ, and π, each ε_{is} is assumed to be independent, so the likelihood of the observed outcomes is the product of the probabilities for each individual outcome yielding, $L(y_i|z_i, \theta, \pi) = \prod_{s=1}^{S} \Pr(y_{is}|z_i, \theta, \pi)$. In principle we intend to estimate:

$$LL_{\substack{\text{adult}\\\text{outcomes}}} = \sum_{i=1}^{n} \ln \left[\int_{\theta,\pi} L(y_i|z_i, \theta, \pi) p(\theta, \pi|x_i, Q_i, w_i, a_i) d\theta d\pi \right] \tag{8}$$

Similar to Heckman et al. (2013) and the second stage estimation of the production function, we assume conditional independence of the unobservables π and θ with respect to the observed outcomes, i.e. $h(\theta|w_i, a_i, y_i, z_i) = h(\theta|w_i, a_i)$ and $p(\pi_i|\theta, x_i, Q_i, y_i, z_i) = p(\pi_i|\theta, x_i, Q_i)$.[47] As mentioned in the discussion of the production function, the primary purpose of these assumptions is that it facilitates validation of the model and provides more transparency over the identification of the effect of the factors on the outcomes. Under these assumptions, the objective function can be re-written as:[48]

$$LL_{\substack{\text{adult}\\\text{outcomes}}} = \sum_{i=1}^{n} \int_{\theta,\pi} \ln \left[L(y_i|z_i, \theta, \pi) \right] p(\pi|\theta, x_i, Q_i) h(\theta|w_i, a_i) d\theta d\pi \tag{9}$$

[47] This conditional independence assumption is reasonable given our data. This is in large part because we are using a very large measurement system that contains many continuous variables. For example, we find that 96% of the variation in the KS4 factors can be explained by the observed measures, which we calculate using the law of total variance as $\mathrm{E}(\mathrm{Var}(\theta|w, a))/\mathrm{Var}(\theta)$. Thus, we consider this remaining 4% as classical measurement error that cannot be informed by the observed outcomes or covariates and our estimator is primarily concerned with addressing attenuation bias.

[48] Given the conditional independence assumptions, this log-likelihood is derived in a manner similar to the derivation of Eq. (4).

Let Ψ_B denote the parameters of the production function. Conditional on θ_{ir}, the same draws used to estimate the production function, we draw π_{ir} with probability $p(\pi_{ir} = \pi_u | \theta_{ir}, x_i, Q_i, \hat{\Psi}_B)$.[49] Thus we simulate the integral in the likelihood above as:

$$LL_{\substack{\text{adult}\\\text{outcomes}}} = \sum_{i=1}^{n} \frac{1}{R} \sum_{r=1}^{R} \ln \left[L(y_i | z_i, \theta_{ir}, \pi_{ir}) \right] \tag{10}$$

Since θ and π are treated as data, the outcome equations can be estimated by S standard logit models.[50]

5 Main Results

This section presents our main findings. First, we characterize the distribution of the recovered skills in the population. Second, we report the factor loadings and residual variance for a selection of the KS4 measurements to provide insight into the nature of skills. Third, we present the estimates of the production function parameters. Fourth, we show the effect of math and verbal skills on educational decisions. Finally, we provide robustness checks to address possible endogeneity concerns.

5.1 Factor Distribution

Our empirical strategy recovers nine correlated factors, a math and a verbal factor for each of the four Key Stages, in addition to a selection factor in KS4. Table 4 displays the mean, standard deviation and correlation matrix for the estimated factors. The top panel shows that the factor means increase over time, which is expected given that they are scaled based on the National Curriculum levels. We also find that while gender disparities in math skills are small at each stage of compulsory education, females' large advantage in verbal skills is persistent and increasing during the schooling career.

[49] $\Psi_B = \{\psi, r, \rho, \delta, \gamma, \alpha, \varsigma, \pi_1, \ldots, \pi_U, \phi_1, \ldots, \phi_U\}$.

[50] The log-likelihood in Eq. (10) is only valid under the conditional independence assumptions. For robustness we estimate the parameters of the model without imposing the conditional independence assumption using the log-likelihood in Eq. (8), which takes the form: $\sum_{i=1}^{n} \ln[\frac{1}{R} \sum_{r=1}^{R} L(y_i|z_i, \theta_{ir}, \pi_{ir})]$. Because the integration is inside of the log function, we use a larger number of draws ($R = 25$) to reduce simulation bias (Lee, 1995). We get very similar results with this alternative specification, which we discuss in more detail in Footnote 68.

Table 4: Factor Moments By Gender

	Math				Verbal				KS4 Selection
	KS1	KS2	KS3	KS4	KS1	KS2	KS3	KS4	
Factor Means									
Female	2.277	4.493	6.068	7.117	2.134	4.565	5.838	7.622	4.312
	(0.003)	(0.003)	(0.006)	(0.009)	(0.002)	(0.003)	(0.004)	(0.008)	(0.003)
Male	2.304	4.536	6.107	7.097	1.966	4.363	5.491	7.068	4.341
	(0.003)	(0.002)	(0.004)	(0.007)	(0.002)	(0.002)	(0.003)	(0.007)	(0.003)
Difference	-0.027	-0.043	-0.039	0.021	0.168	0.203	0.347	0.554	-0.029
	(0.002)	(0.002)	(0.004)	(0.005)	(0.001)	(0.002)	(0.003)	(0.004)	(0.002)
Factor Standard Deviation									
Female	0.557	0.671	1.139	1.698	0.499	0.653	0.818	1.399	0.756
	(0.002)	(0.001)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.003)	(0.002)
Male	0.612	0.704	1.176	1.712	0.495	0.686	0.862	1.503	0.708
	(0.001)	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.004)	(0.002)
Factor Correlation: Lower Diagonal Female, Upper Diagonal Male									
KS1 Math	1	0.779	0.776	0.654	0.847	0.744	0.672	0.588	0.286
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
KS2 Math	0.777	1	0.942	0.796	0.703	0.844	0.749	0.677	0.373
	(0.001)		(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
KS3 Math	0.775	0.941	1	0.924	0.708	0.840	0.839	0.798	0.522
	(0.001)	(0.000)		(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
KS4 Math	0.664	0.801	0.929	1	0.621	0.753	0.821	0.898	0.729
	(0.001)	(0.001)	(0.000)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
KS1 Verb	0.869	0.727	0.730	0.644	1	0.826	0.740	0.640	0.338
	(0.001)	(0.001)	(0.001)	(0.001)		(0.001)	(0.001)	(0.002)	(0.002)
KS2 Verb	0.756	0.859	0.852	0.764	0.842	1	0.906	0.793	0.416
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)		(0.001)	(0.001)	(0.002)
KS3 Verb	0.689	0.774	0.861	0.836	0.757	0.914	1	0.902	0.520
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)		(0.000)	(0.002)
KS4 Verb	0.611	0.705	0.824	0.916	0.663	0.807	0.904	1	0.748
	(0.002)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)		(0.002)
KS4 Selection	0.307	0.382	0.538	0.747	0.338	0.402	0.508	0.744	1
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	

Note: The last panel reports the correlation of the factors by gender. The lower diagonal corresponds to females and the upper diagonal to males. Bootstrapped standard errors are reported in parentheses.

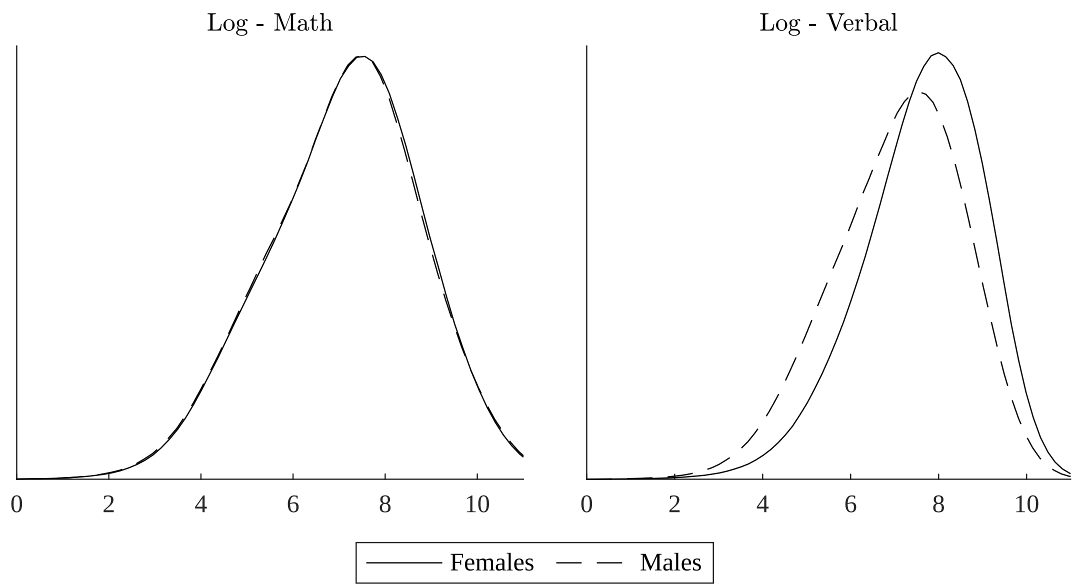


Figure 1: Distribution of Key Stage 4 Skills By Gender

The middle panel of Table 4 reports the standard deviations for the factors, showing that skill dispersion increases over time. These estimates suggest that the math factors have higher variation than the verbal factors at each stage of the schooling career, and the variance of the factors tends to be larger for males than females, with the exception of verbal skills in KS1 and the selection factor. To complete the characterization of the skill distributions, Figure 1 plots the density of the KS4 factors by gender. Despite the fact that we allow for a mixture of 10 normals, none of the distributions exhibit "separation" or bi-modality, however the Kolmogorov-Smirnov test rejects the normality assumption for each of the nine factors.[51]

Finally, the last panel of Table 4 shows the correlation matrix of the factors, where the *lower* diagonal corresponds to females and the *upper* diagonal to males. The correlation between contemporaneous math and verbal skills is around 0.85 at each Key Stage for both gender groups. However, as expected, factors that are more distant in time have a declining correlation, where the coefficient of correlation for females between the KS1 and KS2 math (verbal) factors is 0.777 (0.842) while the correlation between KS1 and KS4 is around 0.664 (0.663). Towards the end of

[51] Moreover, our estimates show that we are not drawing from a single component mixture. In particular, the probability for each of the 10 mixtures is as follows for females [0.093, 0.109, 0.102, 0.084, 0.145, 0.086, 0.063, 0.185, 0.032, 0.101], while for males is [0.085, 0.156, 0.097, 0.078, 0.105, 0.082, 0.081, 0.160, 0.034, 0.123].

compulsory education, skills in adjacent periods are highly correlated, where female KS3 and KS4 math (verbal) factors have a correlation of 0.929 (0.904), similar correlations can be found among males. Finally, the KS4 selection factor correlates similarly with both math and verbal skills.[52]

5.2 Factor Loadings

The factor model relies on multiple Key Stage test scores to identify the latent skills. To assess how skills load on each of the different tests and to determine the importance of measurement error, Table 5 displays, for each gender, the loadings on the math and verbal factors and the residual 'noise' for a subset of KS4 measurements. These results show that performance in statistics mainly relies on math skills while social science relies heavily on verbal skills. However, geography, design and technology and applied business load on both skills. Columns (3) and (6) of Table 5 report the proportion of the total variance of a given measurement that can be interpreted as 'noise' (i.e. $Var(\eta_{ij})/[Var(\lambda_j'\theta_i) + Var(\eta_{ij})]$) for females and males, respectively. Estimates show that measurement error is pervasive in some of our measures. For example, 39.1% of the variability of design and technology scores corresponds to 'noise' when considering female estimates, while for geography it is substantially smaller (14.5%). Finally, there does not appear to be any consequential differences in the loadings between males and females, suggesting that measurements tend to provide similar information regardless of gender. A full list of the estimated factor loadings and residual 'noise' for the 70 measurements are reported in Table 17 for females and Table 18 for males in Appendix E.

5.3 Skill Production Functions

The skill production function estimates characterize the process of skill formation. However, it is important to emphasize that while the CES production technologies are well known, their mathematical simplicity can be deceitful. La Grandville (1989); La Grandville and Solow (2009); Klump and de La Grandville (2000); Klump et al. (2007a,b); Temple (2012), among others have argued

[52]Table 19 in Appendix F shows a transition matrix that illustrates the share of students that are able to climb up the skill distribution between KS1 and KS4. For example, among those students that were in the bottom quartile of the math skill distribution in KS1, 63% of them remained in that same quartile in KS4, and only 2% were able to reach the top quartile in KS4.

Table 5: Factor Loadings: Selected Measurements only loading in Key Stage 4

Measurement	Females			Males		
	KS4 Math	KS4 Verbal	Percent Noise	KS4 Math	KS4 Verbal	Percent Noise
Math	1	0	0.097	1	0	0.097
English	0	1	0.111	0	1	0.106
Design and Technology: Resistant Materials Technology	0.497	0.427	0.391	0.362	0.455	0.478
Geography	0.498	0.827	0.145	0.428	0.794	0.176
Physics	1.111	0	0.077	1.064	0	0.095
Chemistry	1.095	0	0.085	1.041	0	0.102
Social Science	0.146	0.996	0.303	0.096	1.013	0.333
English Literature	0	1.092	0.176	0	1.117	0.164
Statistics	0.994	-0.040	0.174	0.978	0.026	0.175
Applied Business	0.498	0.724	0.298	0.476	0.711	0.330

Note: Values of 0 or 1 denote normalizations. Appendices 17 and 18 report the loadings and percent noise as well as standard errors for all the measurements used in the identification of the KS1-KS4 factors. Percent noise refers to $Var(\eta_{ij})/[Var(\lambda_j'\theta_i) + Var(\eta_{ij})]$.

that comparisons of CES technologies require caution when production technology parameters differ. For example, according to Temple (2012) comparisons of the following production functions, $AF(I_1; I_2)$ and $BG(I_1; I_2)$, are not straightforward. While, the parameters A and B have the same interpretation (i.e. TFP parameters), the production technologies differ, making comparison of the magnitudes of A and B not very informative.[53] In a similar vein, La Grandville (1989) claims that if the elasticity of substitution varies then both the TFP and the input share parameters cannot be analyzed independently of the elasticity. Therefore, while in the following subsection we describe (mainly within period) the nested CES parameters, we also provide a more cohesive characterization of our key estimates by reporting overall marginal effects of each of the production function inputs.

[53] Temple (2012) also discusses issues that arise when normalizing CES production functions due to differences in the scale of the inputs. However, in our case all the inputs are on the same scale, given that 1 log-unit increase in any of our inputs is interpreted as a 100% increase.

5.3.1 Parameters of the Nested CES Technology of Skill Formation

Table 6 displays selected parameter estimates by gender for the nested-CES production functions in Eq.(1) for each Key Stage.[54] Appendix H reports for completeness estimates corresponding to regular two-input CES and nested CES production functions that do not account for unobserved heterogeneity or covariates in the TFP term.[55]

Math Production Function We begin by describing the parameters of the inner-nest (Ω_{it}^k). The share parameters (α) indicate that self-productivity effects play a large role in the production of math skills at each Key Stage. Furthermore, cross effects (i.e. verbal skills) are only relevant in the production of math skills in later Key Stages. In the case of KS2, our estimates indicate that (α) is essentially one. Therefore, the data are indicating that the production function resembles a two-input CES (i.e. math skills and school quality), where verbal skills do not seem to play any role in the production of math skills for this specific period. Due to this fact, it is not possible to recover a meaningful elasticity of substitution between math and verbal skills in KS2.[56] It is important to emphasize that the absence of cross effects in KS2 emerges once we account for unobserved heterogeneity and background characteristics in the TFP, highlighting their relevance in our model.[57] Regarding the elasticities of substitution between math and verbal skills ($1/(1-\gamma)$) in KS3 and KS4, our estimates indicate that γ's are in almost all cases not statistically different from zero, suggesting that the inner nest approximates a Cobb-Douglas production function. The share parameters of the outer nest (δ) indicate that while lagged skills play a predominant role in the production of future skills, school quality is also important. The elasticities of substitution corresponding to skills-school quality ($1/(1-\rho)$) show some degree of complementarities at earlier stages of the schooling career (i.e. KS2). This implies that high-skilled 11 year old students in

[54] Appendix G reports the TFP and unobserved heterogeneity parameters.

[55] In particular, we estimated models that correspond to regular CES production functions (i.e. not nested), where the production function is characterized as follows: $\Theta_{i,t}^k = A_{it}^k[\delta_t^k \left(\Theta_{i,t-1}^m\right)^{\rho_t^k} + (1-\delta_t^k)\left(\Theta_{i,t-1}^v\right)^{\rho_t^k}]^{r_t^k/\rho_t^k}$. The main difference with respect to the nested CES models is that we include school quality in the TFP instead of being an input. Moreover, given that school quality enters linearly in these specifications, we model it through school fixed effects.

[56] Dashes in the table indicate that the complementarity parameter between math and verbal skills is not meaningful given that the share parameter of the inner nest is close to 1.

[57] Tables 22 and 23 in Appendix H show production function estimates by gender when we do not control for unobserved heterogeneity or covariates in the TFP.

Table 6: Production Function: 3-Input Nested-CES(m,v,Q)

	Math			Verbal		
	KS2	KS3	KS4	KS2	KS3	KS4
Panel A: Female						
Math Coefficient (α)	0.980	0.923	0.926	0.013	0.000	0.001
	(0.002)	(0.003)	(0.003)	(0.005)	(0.000)	(0.000)
Verbal Coefficient ($1-\alpha$)	0.020	0.077	0.074	0.987	1.000	0.999
	(0.002)	(0.003)	(0.003)	(0.005)	(0.000)	(0.000)
Complementarity Parmaeter Math/verbal (γ)	–	-0.180	0.143	–	–	–
		(0.136)	(0.037)			
Math/Verbal Aggregator Coefficeint (δ)	0.893	0.915	0.879	0.939	0.868	0.982
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
School Coefficient($1-\delta$)	0.107	0.085	0.121	0.061	0.132	0.018
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
Complementarity Parameter Skill/School (ρ)	-0.282	0.295	0.285	0.055	0.436	-0.153
	(0.046)	(0.013)	(0.009)	(0.064)	(0.019)	(0.030)
Return to Scale (r)	0.735	1.458	1.075	0.866	1.025	1.095
	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)
Variance of shocks ν (ς^2)	0.114	0.083	0.105	0.074	0.073	0.119
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)
Panel B: Male						
Math Coefficient (α)	0.992	0.933	0.976	0.062	0.000	0.001
	(0.001)	(0.004)	(0.003)	(0.005)	(0.000)	(0.001)
Verbal Coefficient ($1-\alpha$)	0.008	0.067	0.024	0.938	1.000	0.999
	(0.001)	(0.004)	(0.003)	(0.005)	(0.000)	(0.001)
Complementarity Parmaeter Math/verbal (γ)	–	0.117	-0.072	0.034	–	–
		(0.295)	(0.070)	(0.100)		
Math/Verbal Aggregator Coefficeint (δ)	0.893	0.901	0.858	0.940	0.851	0.954
	(0.003)	(0.002)	(0.004)	(0.002)	(0.002)	(0.004)
School Coefficient($1-\delta$)	0.107	0.099	0.142	0.060	0.149	0.046
	(0.003)	(0.002)	(0.004)	(0.002)	(0.002)	(0.004)
Complementarity Parameter Skill/School (ρ)	-0.390	0.423	0.308	0.133	0.581	0.107
	(0.046)	(0.012)	(0.009)	(0.046)	(0.012)	(0.023)
Return to Scale (r)	0.738	1.460	1.056	0.894	1.020	1.117
	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)
Variance of shocks ν (ς^2)	0.132	0.087	0.128	0.091	0.087	0.144
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)

Note: Panel A corresponds to estimates using the female sample, while panel B corresponds to estimates based on the male sample. Models account for covariates in TFP, and unobserved heterogeneity. Covariates in TFP include: race, mother tongue, free school meal eligibility, special education needs, and neighborhood fixed effects. Dashes in the table indicate that the complementarity parameter is not recovered because the share parameter of the inner nest is close to 1. Therefore, these models should essentially be interpreted as a two-input CES. Bootstrapped standard errors are reported in parentheses.

math would benefit more from better schools than low-skilled students. To conclude, the returns to scale parameters (r) are not necessarily equal to 1, indicating that relaxing the assumption of constant returns may not be trivial.

Verbal Production Function The parameters corresponding to the inner-nest indicate that self-productivity effects $(1-\alpha)$ play an important role in the production of verbal skills. However, verbal skills do not rely on cross-effects as α is essentially zero at all stages of the schooling career with the exception of the KS2 production function for males.[58] Due to this fact, the elasticity of substitution between math and verbal skills $(1/(1-\gamma))$ is not meaningful. Regarding the outer-nest, we do not find complementarities between skills and school quality at earlier stages of the schooling career (as it is in the case of math). However, the share parameters corresponding to the outer nest $(1-\delta)$ are relatively large and different from zero in most Key Stages suggesting that schools do play a relevant role in the formation of skills. Finally, the returns to scale parameters are close to one (with the exception of KS2). This finding indicates that the assumption of constant returns to scale at later stages of the schooling career is sensible for the production of verbal skills.

5.3.2 Marginal Effects

To provide a more cohesive characterization of the overall effects of each of the nested CES inputs, this section analyses the marginal effects of the inputs under three different specifications of the CES production function.[59] Our aim is twofold. First, we attempt to better characterize the overall role of self-productivity, cross-effects and school quality in the production of skills. Second, we intend to assess the sensitivity of our main findings. Table 7 displays the marginal effects by gender for the three models. Panels A and B correspond to nested CES production functions, where the main difference between them is the inclusion of covariates and unobserved heterogeneity in the TFP term. Panel C reports marginal effects of a regular two-input (i.e. math and verbal skills) CES production function, where the TFP accounts for background characteristics, unobserved

[58] Similar to math skills, a comparison between model specifications shows that the inclusion of covariates and unobserved heterogeneity substantially reduces the size of the share parameters. See Appendix H: Tables 22 and 23 with estimates that correspond to alternative specifications.

[59] Marginal effects of the CES have a closed analytical form which are derived in Appendix D.

heterogeneity, and school fixed effects.[60] Therefore, the main difference between panels A and C is how we account for school effects.[61] While the regular two-input CES model specifies school quality as fixed effects in the TFP, the nested CES models account for school quality from the Ofsted report cards as an additional input in the production function. Given that marginal effects of CES production functions depend on the ratio of inputs, we evaluate them at the somewhat "typical" student (i.e. performing *at expected level* in each skill, and attending a *satisfactory* school for the three-input CES).[62]

Five main results emerge from Table 7. First, irrespective of how we model school inputs, as an input in the nested CES production function (Panel A) or as school fixed effects included in the TFP (Panel C), the skill marginal effects are very similar. Second, a comparison between Panels A and B shows that allowing for a rich specification of the TFP leads to large drops in skill marginal effects, particularly when considering cross effects.[63] Third, self-productivity effects play an important role in the formation of skills where the level effects are larger in secondary school (i.e. KS3 and KS4) than in elementary school (i.e. KS2). For example, Panel A shows that increasing KS1 female (male) math skills by one log unit would increase KS2 math skills by 0.643 (0.653) log units, while a similar increase in KS3 skills would lead to an increase of 0.940 (0.967) log units in KS4.[64] Fourth, the relevance of cross-effects heavily depends on the type of skill. Panels A and C show that math skills have almost no impact in the production of verbal skills, while verbal skills have a positive effect on the production of math skills, particularly in secondary school. Finally, Panel A shows that school quality, included as an input in the nested CES production function, plays an important role in the production of skills. For example, moving a female student from a

[60] The two-input CES takes the form $\Theta^k_{i,t} = A^k_{it}[\delta^k_t \left(\Theta^m_{i,t-1}\right)^{\rho^k_t} + (1-\delta^k_t)\left(\Theta^v_{i,t-1}\right)^{\rho^k_t}]^{r^k_t/\rho^k_t}$, where A^k_{it} includes school fixed effects in addition to race, mother tongue, free school meal eligibility, special education needs, and neighborhood fixed effects.

[61] Appendix H shows the estimates of the CES parameters that lead to the marginal effects that are derived in Panels B and C of Table 7, while Table 6 shows the parameter estimates that lead to the marginal effects reported in Panel A.

[62] Performing *at expected level* corresponds to a specific classification determined by the National Curriculum Standards of achievement for pupils in compulsory education. In particular, it implies assigning skill log levels of 2, 4, and 5 in KS1, KS2, and KS3, respectively. See Section 2 for more details.

[63] The zeros in Table 7 are given by the fact that the input elasticity in the production function is essentially zero.

[64] The scale has a specific interpretation that corresponds to the National Curriculum levels, one log unit could imply (depending on the Key Stage and the lagged level of the student) moving a student from below to meeting expectations in performance.

Table 7: Marginal Effects (elasticities) of Skills

	Math			Verbal		
	KS2	KS3	KS4	KS2	KS3	KS4
Female						
Panel A: 3-Input Nested-CES(m,v,Q), With Covariates in TFP†, With Unobserved Heterogeneity						
Math ($\theta^m_{i,t-1}$)	0.643 (0.003)	1.280 (0.003)	0.940 (0.002)	$0^\natural$	$0^\natural$	$0^\natural$
Verbal ($\theta^v_{i,t-1}$)	$0^\natural$	0.107 (0.004)	0.076 (0.003)	0.803 (0.005)	0.964 (0.004)	1.063 (0.003)
School (Q_{it})	0.079 (0.002)	0.071 (0.001)	0.060 (0.001)	0.052 (0.002)	0.061 (0.001)	0.030 (0.001)
Panel B: 3-Input Nested-CES(m,v,Q), No Covariates in TFP, No Unobserved Heterogeneity						
Math ($\theta^m_{i,t-1}$)	0.847 (0.005)	1.366 (0.003)	1.179 (0.003)	0.125 (0.008)	$0^\natural$	0.189 (0.004)
Verbal ($\theta^v_{i,t-1}$)	0.138 (0.007)	0.284 (0.003)	0.312 (0.004)	0.981 (0.011)	1.161 (0.002)	1.295 (0.004)
School (Q_{it})	0.079 (0.002)	0.077 (0.002)	0.107 (0.002)	0.054 (0.003)	0.061 (0.001)	0.079 (0.002)
Panel C: 2-Input CES(m,v), With Covariates in TFP‡, With Unobserved Heterogeneity						
Math ($\theta^m_{i,t-1}$)	0.658 (0.004)	1.271 (0.003)	0.939 (0.001)	0.018 (0.002)	$0^\natural$	$0^\natural$
Verbal ($\theta^v_{i,t-1}$)	0.017 (0.001)	0.095 (0.004)	0.073 (0.003)	0.790 (0.004)	0.929 (0.003)	1.060 (0.004)
Male						
Panel A: 3-Input Nested-CES(m,v,Q), With Covariates in TFP†, With Unobserved Heterogeneity						
Math ($\theta^m_{i,t-1}$)	0.653 (0.003)	1.300 (0.004)	0.967 (0.002)	0.052 (0.004)	$0^\natural$	$0^\natural$
Verbal ($\theta^v_{i,t-1}$)	$0^\natural$	0.094 (0.005)	0.024 (0.003)	0.788 (0.006)	0.967 (0.003)	1.077 (0.003)
School (Q_{it})	0.079 (0.002)	0.066 (0.001)	0.065 (0.002)	0.054 (0.002)	0.053 (0.002)	0.038 (0.001)
Panel B: 3-Input Nested-CES(m,v,Q), No Covariates in TFP, No Unobserved Heterogeneity						
Math ($\theta^m_{i,t-1}$)	0.931 (0.003)	1.382 (0.003)	1.085 (0.003)	0.203 (0.005)	$0^\natural$	0.119 (0.003)
Verbal ($\theta^v_{i,t-1}$)	0.026 (0.001)	0.254 (0.004)	0.368 (0.005)	0.953 (0.007)	1.174 (0.002)	1.412 (0.004)
School (Q_{it})	0.074 (0.003)	0.070 (0.001)	0.102 (0.002)	0.041 (0.002)	0.049 (0.001)	0.083 (0.002)
Panel C: 2-Input CES(m,v), With Covariates in TFP‡, With Unobserved Heterogeneity						
Math ($\theta^m_{i,t-1}$)	0.659 (0.004)	1.278 (0.007)	0.956 (0.002)	0.052 (0.004)	$0^\natural$	$0^\natural$
Verbal ($\theta^v_{i,t-1}$)	$0^\natural$	0.074 (0.009)	0.021 (0.003)	0.768 (0.006)	0.906 (0.003)	1.066 (0.002)

Note: The top panels correspond to estimates using the female sample, while the bottom panels correspond to estimates based on the male sample. Bootstrapped standard errors are reported in parentheses.

† Controls include race, mother tongue, free school meal eligibility, special education needs, and neighborhood fixed effects.

‡ Controls include race, mother tongue, free school meal eligibility, special education needs, neighborhood fixed effects and **school fixed effects**.

♮ Denotes that the effect is zero given that the input elasticity estimate from the production function is essentially zero.

satisfactory school to a *good* school would increase log math skills in KS2 (KS4) by 0.079 (0.06) log points.

5.3.3 Comparing the Production Function of Females and Males

To conclude, the estimates in Table 6 and the marginal effects in Table 7 do not seem to show important gender differences in the production of skills, indicating that female advantage in verbal skills is likely driven by initial conditions rather than by differences in the production of these skills. However, given the large number of parameters and the complexity of the nested CES functions it is difficult to form a definitive conclusion. To provide a concise answer to this matter, we perform a simple simulation where we apply the female nested CES production function parameters to the male skill production functions.[65] This exercise will not only provide an overall measure of gender differences in the production of skills, but it will also help to determine whether the large female advantage in verbal skills is driven by disparities in initial conditions or in the production of skills. Table 8 shows the results of two simulations. Column (1) displays the actual KS4 skills of males, which serves as the benchmark. Column (2) reports counterfactual KS4 skills of males when KS1 skills and KS2–KS4 TFP terms, which are a function of background characteristics, are equalized between gender groups. Finally, Column (3) shows KS4 skills of males in the counterfactual scenario where they share the same CES production functions as females during KS2–KS4, though we hold constant their initial skills and the TFP terms.[66] Results indicate that equalizing the nested CES parameters between females and males does not lead to substantial changes in the level of skills of males. For example, assigning to males the female nested CES production function parameters in the KS2–KS4 period would increase (decrease) their math (verbal) skills by 0.065 (0.113) log points. While these differences are not negligible, their size does not suggest large differences in how the CES parameters impact the skills of males and females. Moreover, they cannot explain the gender disparities in verbal skills. In fact, the current gender differences in the production of these skills contribute to close the gap. However, if instead we equalize the initial skills and the

[65]We recover counterfactual KS4 skills of males by combining their initial conditions with the female production functions.

[66]This implies that these results are focused on the main components of the production functions.

Table 8: Male Key Stage 4 Skills with Female Production Function

	Baseline	Equalize Initial KS 1 Skills and KS2–KS4 TFP	Equalize KS2–KS4 CES Parameters
	(1)	(2)	(3)
KS4 Math	7.134	7.041	7.199
	(0.007)	(0.047)	(0.051)
KS4 Verbal	7.101	7.767	6.988
	(0.007)	(0.045)	(0.045)

Note: This table presents two counterfactual simulation results. Column (1) reports actual level of skills for males, which serves as a benchmark. Column (2) reports counterfactual KS4 skills of males after equalizing initial skills (i.e. KS1 skills) and TFP terms (i.e. males are given the female characteristics). Finally, in Column (3) males are assigned the female nested CES production function parameters at the different stages of the schooling career (i.e between KS2 and KS4), while holding fixed the initial skills (i.e. KS1 skills) and the TFP term.

TFP terms, then we can fully explain the gap in verbal skills in KS4. In summary, these findings suggest that differences in the nested CES parameters are not driving the large gender disparities in verbal skills and that initial skills and the TFP components are the main drivers of the gap in KS4 skills.

5.4 The Role of Skills in Overall University Enrollment

In this section, we study how skills impact educational outcomes. Table 9 reports the results by gender for different logistic models that analyze the effect of KS4 math and verbal skills on university enrollment. We focus on KS4 skills because this is the period when students begin making educational decisions. All specifications include a rich set of baseline controls (i.e. race, mother tongue, eight covariates of neighborhood characteristics, free school meal eligibility, special education needs, number of GCSE exams taken in KS4, number of excused and unexcused absences in KS4, KS4 school quality determined by Ofsted reports, and unobserved heterogeneity derived from the production functions).[67] Columns (1) and (2) of Table 9 display the average marginal

[67] The eight covariates of neighborhood characteristics correspond to Income Deprivation, Employment Deprivation, Health Deprivation and Disability, Education Skills and Training Deprivation, Barriers to Housing and Services, Living Environment Deprivation, Crime and Income Deprivation Affecting Children. We do not include neighborhood fixed

effect of each skill on college enrollment when included in separate specifications, which indicate that both skills have a large impact on educational attainment when considered separately. Column (3) of Table 9 shows the average marginal effects for each skill when both skills are included in the same specification. These results show that the partial effect of a one log-unit increase in verbal skills is approximately 3 times larger than an analogous increase in math.[68] Specifications in Columns (4)-(6) provide further robustness checks to this finding. Column (4) adds higher order factor terms (i.e. squared term of each factor, and their interaction), Column (5) additionally controls for Key Stage 1 math and verbal skills to further control for differences in background characteristics, and Column (6) includes KS4 school fixed effects instead of the school quality measures obtained from the Ofsted reports. Each of these specifications further confirm the finding that verbal skills play a larger role than math skills in explaining college enrollment.[69] Finally, while males and females show a similar pattern of a higher importance of verbal skills relative to math skills in explaining college enrollment, the ratio of the math-verbal skills coefficients for males tend to be larger than for females, suggesting differential responses to skill endowments.[70] To conclude, while this analysis is informative in terms of assessing the relevance of each skill in predicting adult outcomes, it does not provide any evidence on the cost to improve one skill relative to the other, making it more difficult to suggest policy recommendations on which skill should be targeted. Section 6 provides a set of counterfactual simulations that will shed light on this issue.

To further analyze the role of KS4 skills in adult outcomes, we explore the extent to which non-linear effects impact college enrollment decisions. The aim of this exercise is to determine

effects given that many neighborhoods do not have students attending college or enrolled in a given field of study, making the identification of fixed effects more problematic given the sparsity of the data.

[68]The results in Table 9 are estimated under the conditional independence assumption in Section 4. We also estimated the outcome equations without imposing the conditional independence assumption and obtained results very similar to the ones in the table. For example, for females, the equivalent Column (4) marginal effects in Table 9 without imposing conditional independence is 0.049 for math and 0.172 for verbal.

[69]If the actual scores in the main math and verbal (i.e. English) tests were used instead of the recovered latent skills, the differences in the coefficients between the math and verbal scores would be substantially smaller (see Table 3). This result suggests that addressing problems of measurement error as discussed in Section 2 is highly relevant.

[70]Our main analysis does not include any features of the supply-side of the university market, which could exaggerate the importance of verbal skills relative to math skills. To consider the supply-side, we conduct a program-by-program analysis of degree prerequisites that shows that nearly half of all degrees granted in England, weighted by total population enrolled, require some math or science preparation prior to enrollment, which suggests that the imbalance of math and verbal is not necessarily a byproduct of university course/majors offerings. Appendix J provides the details of this analysis.

Table 9: Logistic Regression: University Enrollment and KS4 Skills

	(1)	(2)	(3)	(4)	(5)	(6)
	Average Marginal Effect of 1 log unit increase in Skills					
Females						
KS4 Math	0.142 (0.001)	–	0.053 (0.001)	0.054 (0.001)	0.057 (0.001)	0.054 (0.001)
KS4 Verbal	–	0.198 (0.001)	0.149 (0.001)	0.147 (0.001)	0.150 (0.001)	0.148 (0.001)
Males						
KS4 Math	0.131 (0.001)	–	0.055 (0.001)	0.058 (0.001)	0.061 (0.001)	0.058 (0.001)
KS4 Verbal	–	0.173 (0.001)	0.128 (0.001)	0.126 (0.001)	0.127 (0.001)	0.125 (0.001)
Controls For:						
Baseline Controls	X	X	X	X	X	X
Higher Order Factor Terms	–	–	–	X	X	X
Key Stage 1 Factors	–	–	–	–	X	X
School Fixed Effects	–	–	–	–	–	X

Note: Results correspond to logistic regressions. Baseline controls include race, mother tongue, Key Stage 4 school quality or school fixed effects depending on the specification, free school lunch, special education needs, number of GCSE exams taken in KS4, number of excused and unexcused absences in KS4, neighborhood characteristics, and unobserved persistent TFP shock. Key Stage 1 factors correspond to math and verbal skills in KS1. Higher order factor terms involve squared terms of Key Stage 4 skills and their interactions. Bootstrapped standard errors at the school level reported in parentheses.

whether increases in verbal and math skills are particularly important in specific areas of the skill distributions. Figure 2 shows the predicted share of college enrollment at different points of the skill distributions if all students in the population were assigned that given value of math (verbal) skills, while holding fixed the population distribution of the other covariates impacting college enrollment, including their verbal (math) skills.[71] Two key messages emerge from this figure. First, if verbal skills are sufficiently low, then a very low share of students would enroll in college. However, this is not the case for math skills, where giving students low levels of math skills would still lead to a relatively high share of students attending college (i.e. 30%). Second, increases in math skills when their level is low enough (i.e. log-skill between 4 and 7) have almost no impact on improving the probability of enrolling in college. However, this is not the case with verbal skills, where improvements, even at low levels, increase the probability of enrolling in college.[72] In summary, Figure 2 indicates that verbal skills play an important role in explaining college enrollment and that increasing math skills among low ability students has a very small impact on educational attainment. Therefore, these findings suggest that targeting the verbal skills of the students that lag behind (instead of their math skills) could be more efficient if the goal is to improve educational attainment. This result is consistent with the fact that among students with low probability of attending college, their probability of choosing a STEM field is substantially smaller. For example, our data show that among those college enrollees that have a *below* average population probability of attending college, only 23% of them enrolled in STEM fields. While among those that have an *above* average population probability, 31% enrolled in STEM fields.

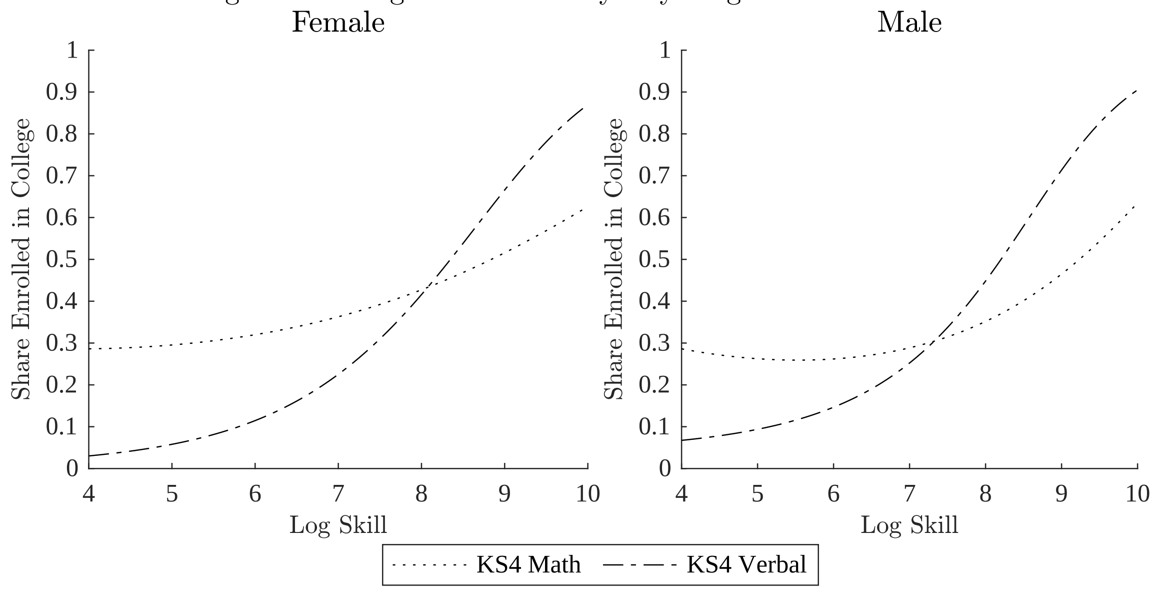
5.4.1 Endogeneity Concerns

Given that our findings point towards a more relevant role of verbal skills in explaining educational attainment than previously discussed in the literature, it is important to show that this result is not driven by endogeneity bias. Despite the fact that we control for a large number of background covariates, unobserved heterogeneity, number of GCSE exams taken, KS4 absences, and school

[71] These figures are derived from the model in Column (4) of Table 9, where squared terms of the factors and their interactions are included as additional covariates.

[72] We perform a similar analysis but focusing on unconditional STEM enrollment. Figure 4 in Appendix I shows important non-linearities, suggesting that in order to enroll in STEM fields a minimum level of math skills is necessary.

Figure 2: College Enrollment By Key Stage 4 Skill and Gender



Note: Shares refer to average probability of college enrollment when all students are assigned a given level of math (verbal) skills while holding fixed the rest of their characteristics, including their verbal (math) skills.

quality or school fixed effects and neighborhood characteristics, it is still possible that some unobserved characteristic could disproportionately correlate with, for example, verbal skills. In this regard, we performed several robustness checks to assess the extent to which confounding factors may be driving this result. To preserve space, we briefly describe these strategies in this subsection and provide detailed explanations and complete results in Appendix J.

Role of Motivation A possible concern is that students at KS4 (age 16) may have already made their educational decisions (e.g. whether to attend university and field of study if attending), and therefore the effort that they may exert is a function of these decisions. For example, students who have already planned to obtain a degree in history may not spend much time studying math. One way to address this concern is to look at the relative importance of skills using earlier measures (e.g. KS1, age 7), when effort and motivation at school are less likely to be determined by decisions that will be made 11 years later. The top panel of Appendix Table 24 repeats the analysis in Table 9 but instead including skills from earlier Key Stages. Results show that verbal skills uniformly have a larger effect on university enrollment than math skills at each stage of the schooling career. Moreover, the magnitude of the differential effect is sizable across the board, which further substantiates our main findings.[73]

Decomposition of KS4 Skills We propose a statistical decomposition of KS4 skills into the component that is predicted by early skills and a component that represents the prediction error/residual. In particular, we express θ_{i4} as:

$$\theta_{i4} = \underbrace{\left(\theta_{i4} - \hat{\theta}_{i4|3,2,1}\right)}_{\text{KS4 Residual}} + \underbrace{\left(\hat{\theta}_{i4|3,2,1} - \hat{\theta}_{i4|2,1}\right)}_{\text{KS3 Residual}} + \underbrace{\left(\hat{\theta}_{i4|2,1} - \hat{\theta}_{i4|1}\right)}_{\text{KS2 Residual}} + \hat{\theta}_{i4|1} \tag{11}$$

where $\hat{\theta}_{i4|3,2,1} = \mathrm{E}(\theta_{i4}|\theta_{i3}, \theta_{i2}, \theta_{i1}), \hat{\theta}_{i4|2,1} = \mathrm{E}(\theta_{i4}|\theta_{i2}, \theta_{i1})$, and $\hat{\theta}_{i4|1} = \mathrm{E}(\theta_{i4}|\theta_{i1})$. The benefit of this approach is to study the relative contribution to university enrollment of math and verbal skill "residuals" after conditioning on prior skill levels. We include all these "KS residual" terms in a college enrollment regression, which will allow us to recover the marginal effect of the new information received (in terms of predicting KS4 skills) at different school stages. Appendix Table 25 shows that, at each Key Stage, performing better than expected in verbal has a much larger impact on college enrollment than performing better than expected in math. This finding shows that even after conditioning on different KS skills (that could be considered as playing the role of sufficient statistics) our result on the relevance of verbal skills still holds. Therefore, it suggests that endogeneity in parental inputs that benefit one skill over the other is not likely to be driving our findings. A more detailed description of this analysis can be found in appendix J.[74]

Interrelation between Test Scores and Externalizing Behavior, Family Background Characteristics, and IQ It is possible that our verbal factor is capturing other types of skills

[73] As an additional robustness check, we repeated the analysis but this time constrained the sample in two different ways. First, we focused on students whose mother's native language is English. Second, we further constrained this group to only include white students. Table 24 in Appendix J shows that in all cases verbal skills are more important than math skills in explaining college enrollment.

[74] Our analysis does not include any features of the supply-side of the university market, which could exaggerate the importance of verbal skills relative to math skills. To consider the supply-side, we conduct a program-by-program analysis of degree prerequisites that shows that nearly half of all degrees granted in England, weighted by total population enrolled, require some math or science preparation prior to enrollment, which suggests that the imbalance of math and verbal is not necessarily a byproduct of university course/majors offerings. Appendix J provides the details of this analysis.

that affect schooling outcomes, which are not present in the math factor. For example, if externalizing behavior/socio-emotional skills are more related to English test scores than math test scores, then we could be confounding the larger effect of verbal skills with the role played by externalizing behavior. While we control for special education needs, number of absences, number of GCSE exams taken and, and unobserved heterogeneity in our main estimates in Table 9, we further investigate this issue using a database that contains richer measures of externalizing behavior. We use the Avon Longitudinal Study of Parents and Children (ALSPAC) which is a large scale longitudinal study of children born in Avon (United Kingdom) during the early 1990s. Although this database cannot be linked to one of our main databases (i.e. HESA), it is useful for further analysis because it contains rich information on student background characteristics, and the individuals in the sample are UK students that are similar in age to students in our main database. The data contains proxies for externalizing behavior obtained from the Strengths and Difficulties Questionnaire (SDQ), and student performance on KS2 math and English tests. To study if externalizing behavior/socio-emotional skills have a higher correlation with verbal skills than math skills, we perform a regression analysis where the dependent variables are performance in KS2 math or verbal (English) exams and the independent variables are the SDQ measures. Appendix Table 26 shows that, while these proxies for externalizing behavior are highly predictive of math and English test scores, they do not favor one skill over the other, i.e. all components of the SDQ questionnaire have similar effects on both tests. We provide a detailed description of this analysis in Appendix J, where we also extend the analysis by considering detailed family background characteristics and measures of student IQ. Results show in all cases that these variables have a very similar impact on KS2 math and English scores.

5.5 The Effect of Skills in University Selectiveness, Field of Study, and Graduation

To further explore how math and verbal skills affect educational outcomes, Table 10 shows logistic regression results by gender that focus on selectiveness of the university attended, field of study, and graduation.[75]

University Selectiveness In Column (1), we analyze the effect of math and verbal skills on university enrollment when those attending the most selective institutions are removed from the sample (i.e. 20% of total enrollees). The purpose of this analysis is to distinguish whether our previous findings are driven by the bottom 80% of enrollees or the top 20%. Removing the top 20% of enrollees produced estimates that were nearly identical to the full sample in Table 9. This further confirms that the larger effect of verbal skills on university enrollment is likely driven by those students who are at the extensive margin of the university enrollment decision not the intensive margin of school selectivity. To provide additional insight, Column (2) of Table 10 shows the effect of skills on the probability of enrolling in a selective institution (i.e. Russell group). This analysis only looks at college enrollees and shows that math and verbal skills have a much similar impact on the probability of attending a selective institution. Moreover, depending on gender the relative importance in the size of skills' effects is changed. While among females verbal skills continue to be more important than math skills to enroll in selective institutions, among males we find that math skills become as important as verbal skills. In summary, this last result suggests that selective universities are enrolling students from the top of both skill distributions.

STEM Enrollment Column (3) and (4) display results for enrollment in STEM fields unconditional and conditional on university enrollment, respectively.[76] As expected, Column (3) shows that math skills have a positive effect on enrolling in STEM fields, while verbal skills have an effect close to zero. Similarly, results in Column (4), where the sample is restricted to college enrollees, show large positive effects for math and negative effects for verbal.[77] However, this column highlights interesting disparities between males and females, suggesting that males are much more responsive to skills than females. The larger responsiveness of males to skills occurs in both directions. For example, a 1 log unit skill increase in math leads to a 16.2 percentage point increase in STEM

[75] All these specifications include the same controls as in Column (4) of Table 9.

[76] The following majors are considered as STEM: biology sciences, physical sciences, math sciences, engineering, computer sciences, technologies, and combined sciences.

[77] The negative coefficient is driven by the fact that the sample includes only those who enrolled in college, therefore the alternative option is enrollment in non-STEM.

Table 10: Logistic Regression Other Outcomes

	Enroll in Non-Selective Institution	Enroll in Russell	Enroll in STEM	Enroll in STEM	Graduation Overall	Graduation in STEM	Graduation in Non-STEM
	conditional not enrolling Russell	conditional enrollment	unconditional	conditional enrollment	unconditional	conditional enrollment STEM	conditional enrollment Non-STEM
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Average Marginal Effect of 1 log-Skill Increase						
Females							
KS4 Math	0.051	0.069	0.052	0.107	0.039	0.020	0.004
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.004)	(0.002)
KS4 Verbal	0.146	0.115	-0.007	-0.089	0.122	0.060	0.061
	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)	(0.005)	(0.003)
Base Prob.	35%	19%	9%	22%	29%	76%	72%
Obs.	229339	98302	248479	98302	248479	21879	76423
Males							
KS4 Math	0.052	0.093	0.075	0.162	0.032	0.013	-0.006
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.003)	(0.003)
KS4 Verbal	0.125	0.090	-0.008	-0.170	0.108	0.089	0.085
	(0.001)	(0.003)	(0.001)	(0.004)	(0.001)	(0.005)	(0.004)
Base Prob.	28%	20%	13%	39%	22%	64%	68%
Obs.	233802	82528	250257	82528	250257	32327	50201

Note: Results correspond to logistic regressions, we report average marginal effects. Baseline controls include race, mother tongue, Key Stage 4 school quality, free school lunch, special education needs, number of GCSE exams taken in KS4, number of excused and unexcused absences in KS4, neighborhood characteristics, unobserved persistent TFP shock, and higher order factor terms (i.e. squared terms and their interactions). Bootstrapped standard errors at the school level.

enrollment probabilities for males, while a similar increase for females will only increase enrollment by 10.7 percentage points. Looking at the effect of verbal skills, a 1 log unit skill increase will lead to a 17 percentage point *drop* in STEM enrollment for males, while it only produces a 8.9 percentage point *drop* in STEM enrollment for females. This result suggests that males are more sensitive to skill comparative advantage than females when deciding to enroll in scientific fields.[78]

Graduation Finally, we study overall graduation outcomes and graduation by field of study. Consistent with the findings in Table 10, Column (5) shows a larger role for verbal skills when analyzing overall graduation. However, it is important to highlight that this effect is not entirely driven by the larger effect of verbal skills on enrollment. As it can be inferred from the results in Columns (6) and (7) that examine graduation outcomes in STEM and Non-STEM conditional on enrollment in those fields, marginal changes in verbal skills have a large effect on graduation in STEM. Specifically, conditional on enrolling in STEM, increasing verbal skills by 1 log unit increases graduation rates for females by 6 percentage points, while a similar increase in math skills only increases graduation rates by 2 percentage points. While, this last result is likely to be driven by the presence of less variation in math skills conditional on STEM enrollment, it is consistent with our previous finding that points toward a key role of verbal skills in explaining educational attainment.

6 Simulations: The Role of Schools, Initial Skills, and TFP

In this section, we perform a set of counterfactual simulations to provide a deeper understanding of the role of school quality, initial skills, and TFP in explaining gaps in adult outcomes.

Simulation 1 Three exercises are performed to understand how the different inputs of the skill production functions contribute to explain gaps in adult outcomes. In particular, we compare two specific groups of the student population: 1) students that consistently attended low quality schools (disadvantaged students) and, 2) students that consistently attended high quality schools

[78]More detailed analysis on the gender gap in college enrollment and STEM majors is discussed in Section 8.

(advantaged students).[79] Table 11 shows baseline characteristics and gaps between these groups. For example, students in the advantaged group are 31 to 35 percentage points more likely to attend college than disadvantaged students, depending on the gender. Differences in KS4 skills are also substantial with gaps of at least of 1.3 log points. We explore three different mechanisms that could contribute to explain these gaps. First, we study the role of schools. In particular, we equalized school quality by assigning disadvantaged students into "outstanding" schools at each stage of compulsory education.[80] Table 11 shows that this counterfactual would lead to a 20% (approx. 7 percentage points) drop in the college enrollment gap, and a 30% decline (approx.) in the math and verbal skill gaps. These effects are sizable suggesting that schools could play an important role in closing gaps in the population. The second exercise focuses on the role of initial skills. We equalized KS1 math and verbal skills between groups, while keeping school quality and TFP constant. Table 11 indicates that the college enrollment gap would drop by 15% (approx. 5 percentage points), while differences in skills would decline by 20%. Finally, the last exercise involves equalizing the TFP component (i.e. background characteristics). This counterfactual suggests a large decrease in the college enrollment gap, around 60% (approx. 20 percentage points), as well as sizable declines in skill gaps.[81] Overall, this analysis indicates that school quality is a greater contributor to gaps in adult outcomes than differences in initial skill levels, suggesting that policies aiming to improve school quality could help overcome initial skill disadvantages. However, initial conditions as a whole (i.e. initial skills and background characteristics) play a fundamental role in explaining gaps in the population as well.

Simulation 2 Next, we explore counterfactual outcomes in the hypothetical situation where all schools in England were of "outstanding" quality (i.e. Ofsted highest scale value). This implies assigning a high quality school at each stage of the schooling career to all students in the popula-

[79]The group of disadvantaged students is defined by those that attended *inadequate* or *satisfactory* schools at each stage of their schooling career, while the group of advantaged students is constituted of those that attended *outstanding* schools at each stage of their schooling career.

[80]We assume that all the remaining covariates that are included in the TFP (such as background characteristics and unobserved heterogeneity) remain constant.

[81]We would like to emphasize that in these simulations the only channel that leads to an increase in college enrollment is an increase in skills, we are not considering any direct effect that schools may have on college enrollment (e.g. helping students with their college applications). Finally, as a caveat we should also mention that these simulations do not allow parents to react to changes in school quality or in initial skills.

Table 11: Simulation (1) - The Effect of School Quality, Initial Skills, TFP on Adult Outcomes

	School Quality (Q) KS2-KS4	KS1 Math	KS1 Verbal	KS4 Math	KS4 Verbal	College Enroll-ment
Panel A: Female						
Only Attended Outstanding School ($obs. = 3721$)	4.000	2.503	2.341	8.189	8.464	0.624
	(0.000)	(0.014)	(0.013)	(0.037)	(0.031)	(0.009)
Only Attended Inadequate or Satisfactory School ($obs. = 14360$)	1.860	2.153	2.018	6.538	7.135	0.273
	(0.004)	(0.005)	(0.005)	(0.020)	(0.019)	(0.004)
Gap	-2.140	-0.350	-0.323	-1.651	-1.329	-0.351
	(0.004)	(0.016)	(0.014)	(0.042)	(0.036)	(0.010)
Counterfactual Gap						
Equalizing KS2 – KS4 School Quality				-1.147	-0.967	-0.279
				(0.043)	(0.036)	(0.011)
Equalizing KS1 Math and KS1 Verbal Skills				-1.329	-1.056	-0.297
				(0.034)	(0.029)	(0.010)
Equalizing KS2 – KS4 Total Factor Product				-0.821	-0.618	-0.128
				(0.016)	(0.015)	(0.003)
Panel B: Male						
Only Attended Outstanding School ($obs. = 3395$)	4.000	2.542	2.167	8.119	7.977	0.537
	(0.000)	(0.013)	(0.013)	(0.039)	(0.033)	(0.010)
Only Attended Inadequate or Satisfactory School ($obs. = 15508$)	1.856	2.179	1.856	6.560	6.575	0.227
	(0.004)	(0.008)	(0.005)	(0.024)	(0.022)	(0.004)
Gap	-2.144	-0.363	-0.310	-1.559	-1.402	-0.310
	(0.004)	(0.015)	(0.013)	(0.047)	(0.037)	(0.011)
Counterfactual Gap						
Equalizing KS2 – KS4 School Quality				-1.049	-0.986	-0.242
				(0.051)	(0.041)	(0.011)
Equalizing KS1 Math and KS1 Verbal Skills				-1.232	-1.124	-0.262
				(0.037)	(0.030)	(0.009)
Equalizing KS2 – KS4 Total Factor Product				-0.832	-0.663	-0.120
				(0.013)	(0.011)	(0.003)

Note: Bootstrapped standard errors in parentheses.

tion. It is important to emphasize that evidence on the impact of schools on adult outcomes once conditioning on neighborhood characteristics is a result that is not often discussed in the literature. Therefore, our goal is to further analyze the scope of a key policy instrument (i.e. improving school quality) that has been at the center of the educational debate for the last two decades. In addition, this exercise will shed light on the differential degree of malleability of math and verbal skills in response to a given treatment. Column (1) of Table 12 reports the average counterfactual increase in school quality (i.e. treatment) experienced by females and males.[82] Columns (2) and (3) show the effect of a such policy on the accumulation of math and verbal skills, indicating that improving school quality would increase female (male) average KS4 math and verbal skills by 0.284 (0.293) and 0.207 (0.245) log points, respectively.[83] Column (4) shows that such changes in skills would translate into a 11% (13.5%) increase in female (male) college attendance (i.e. 4.5 percentage points). The fact that math skills have increased slightly more in log levels than verbal skills in response to the same intervention could suggest a differential degree of malleability of skills. However, it is important to highlight that due to the larger impact of verbal skills on college enrollment, approximately 65% of the total counterfactual increase in college enrollment corresponds to improvements in verbal skills as displayed in the last row of each panel. In summary, these findings quantify the likely role that school quality plays in improving educational outcomes and its likely importance as a policy instrument.[84]

7 Further Exploration of the Role of Math and Verbal skills in Educational Attainment: Evidence from the U.S.

Our data only covers students in the English education system. It is possible that the relative importance of verbal skills is strictly a phenomenon among this population. To address this critique,

[82] Ofsted evaluations take values between 1 (inadequate) to 4 (outstanding). Notice that this policy would only have an effect on those students that are attending lower quality schools (i.e. inadequate, satisfactory or good schools) which approximately represent 90% of the total student population.

[83] These effects approximately represent 15% of KS4 skills standard deviations.

[84] School accountability programs have been implemented in many countries with the aim to improve student achievement. For example, No Child Left Behind (which has been implemented by the US federal government) punishes schools if students do not show adequate progress.

Table 12: Simulation (2) - The Effect of School Quality on Adult Outcomes

	School Quality (Q) KS2-KS4	KS4 Math	KS4 Verbal	College Enrollment
	(1)	(2)	(3)	(4)
Females				
Baseline - Full Population	2.803	7.160	7.657	0.402
	(0.005)	(0.009)	(0.008)	(0.002)
Counterfactual - Attend Outstanding School KS2-KS4	4.000	7.445	7.864	0.446
	(0.000)	(0.010)	(0.009)	(0.002)
Difference	1.197	0.284	0.207	0.045
	(0.005)	(0.004)	(0.004)	(0.001)
Approximate Contribution to Change in Enrollment	–	33.7%	66.3%	–
Males				
Baseline - Full Population	2.772	7.134	7.101	0.334
	(0.004)	(0.007)	(0.007)	(0.002)
Counterfactual - Attend Outstanding School KS2-KS4	4.000	7.427	7.346	0.380
	(0.000)	(0.011)	(0.011)	(0.002)
Difference	1.228	0.293	0.245	0.045
	(0.004)	(0.006)	(0.006)	(0.001)
Approximate Contribution to Change in Enrollment	–	35.3%	64.7%	–

Note: Baseline refers to the whole population of students. Approximate contribution to change in enrollment refers to how changes in math and verbal skills due to treatment have affected changes in college enrollment. Bootstrapped standard errors in parentheses.

we analyze university enrollment decisions using data from the United States and show that similar patterns to our main findings persist in these data as well. Finally, we conclude this section by discussing the implications of our main findings when interpreting results from Mincer specifications that intend to quantify the return to math and verbal skills.

7.1 Is the Larger Effect of Verbal Skills on University Enrollment a Specific Phenomenon of the UK?

To assess the possibility that our result on the larger role of verbal skills in college enrollment is a consequence of the particular institutional features in the UK, we examine US data for similar patterns on skills and college enrollment. One shortcoming of such a comparison is that few datasets contain such extensive and repeated measures of subject-specific performance as we have available in the UK data. Nonetheless, we make use of subject-specific aptitude and high school transcript data available in the National Longitudinal Survey of Youth of 1997 (NLSY97). The NLSY97 is a nationally representative sample of youths from the United States who were 13 to 17 years old when they were first surveyed in 1997. It collects extensive information on family background characteristics, educational experiences, and labor market outcomes through time. In addition, for a subset of respondents, the data also contains performance on the Armed Services Vocational Aptitude Battery (ASVAB), which involves 12 subject-specific tests, including tests that assess math and verbal skills. Moreover, for a subset of survey respondents, high school transcript records were collected, providing detailed information on course taking and grades from their high school career.

Columns (1) to (4) of Table 13 show the results from a linear probability model that studies the effect of the scores on two of the subject-specific ASVAB tests on college enrollment. We use the score on Paragraph Comprehension (PC) as a proxy for verbal skill and the score on Mathematical Knowledge (MK) as a proxy for math skill. The exam was administered to most participants in 1997. Given that the respondents took the test at different ages, each of the subject-specific scores are normalized by the age-specific mean and age-specific standard deviation of the respondent to make the scores comparable across test takers. In addition, the regressions control for gender, race,

Table 13: Linear Probability Model: University Enrollment (NLSY97)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ASVAB test specific scores†							
Paragraph Comprehension (PC)	0.0954*** (0.0091)	0.2143*** (0.0099)	–	0.1279*** (0.0152)	0.1347*** (0.0094)	0.1478*** (0.0157)	–
Math Knowledge (MK)	0.1662*** (0.0089)	–	0.2164*** (0.0098)	0.1258*** (0.0152)	–	–	–
Arithmetic Reasoning (AR)	–	–	–	–	0.1076*** (0.0093)	0.0948*** (0.0163)	–
Age at ASVAB (age)	-0.0034 (0.0043)	-0.0032 (0.0046)	-0.0046 (0.0045)	-0.0037 (0.0044)	–	-0.0034 (0.0046)	–
(PC) × (age-13)	–	-0.0018 (0.0041)	–	-0.0162** (0.0064)	–	-0.0066 (0.0066)	–
(MK) × (age-13)	–	–	0.0081** (0.0039)	0.0200*** (0.0062)	–	–	–
(AR) × (age-13)	–	–	–	–	–	0.0063 (0.0065)	–
Subject Course Grades‡							
Ninth Grade English	–	–	–	–	–	–	0.1026*** (0.0137)
Algebra 1	–	–	–	–	–	–	0.0709*** (0.0133)
R-squared	0.257	0.201	0.239	0.259	0.223	0.222	0.100
Obs.	5037	5044	5037	5037	5044	5044	2004

† Individual test scores are standardized by age when the test was administered.
‡ Course grades are in grade points, e.g. A is 4.0, B is 3.0, etc.
Note: This analysis only includes the NLSY97 cross-sectional sample of 6,748 respondents designed to be representative of people living in the United States. College enrollment is defined as having a highest grade completed of 13 or more at the most recent interview. Those whose age at last interview was less than 21 were dropped. Round 1 observation weights were used for all regressions. All specifications include controls for gender, race, and ethnicity.

and ethnicity. The results in Column (1) appear to refute our main findings. Math skills appear to be substantially more important than verbal skills for college enrollment. The implicit assumption in this model is that once we normalize the scores by age of test taker, the marginal effect on college enrollment of a one standard deviation increase in test score for a 13 year old is the same as the effect of a one standard deviation increase for a 16 year old. However, given the characteristics of the education system in the US, the distribution of scores for 16 year olds is less likely a reflection of differences in true ability and more likely due to differences in course taking (e.g., not all students have taken geometry by age 16).

To address this potential problem, we analyze the impact of test scores across test taker age. Apart from possible cohort effects, test taker age should be independent of college enrollment. In fact, in Column (1) we include in the regression a control for age at ASVAB test. Since skills are independent of age by construction, the near-zero coefficient confirms this assertion. Column (2), in addition to the base controls, only includes the verbal skill in the regression and the verbal skill interacted with age of test taker.[85] In this regression, age of test taker appears independent of college enrollment. Column (3) performs a similar regression, only looking at the math skill. The coefficient on the skill interacted with age is positive and significant. This suggests that math rank is more predictive of college enrollment for 16 year olds than it is for 13 year olds, which raises concerns of reverse causality given the proximity of the older test takers to the college decision.

The regression results in Column (4) include both the math and verbal skills and their interactions with age of test taker. Both of the interaction terms are economically and statistically significant. These results show that, for the youngest respondents, performance on the verbal measure is slightly more predictive of college enrollment than performance on the math measure. However, for older respondents, the importance of verbal skill declines and the importance of math skill rises. This result raises questions about the appropriateness of using MK as a measure of math skill, especially given that its stated aim is “measuring knowledge of high school mathematics principles” when a large proportion of the respondents were not even exposed to most of the material. This test may simply be measuring differences in exposure to mathematics principles (due

[85] For the interaction, age is subtracted by 13 (the age of the youngest members in the sample), so the coefficient on the non-interacted term is the effect on the skill for a 13 year old.

to the characteristics of the US education system) that are correlated with college enrollment and not actual differences in skill. All the same, this regression demonstrates that the entirety of the disparity in Column (1) is driven by the older respondents, at least those in the ninth grade and higher.[86]

To address the possibility that MK is contaminated with other factors that influence college enrollment beyond math ability, we explore two other measures of math ability. The first is the Arithmetic Reasoning (AR) subject test of the ASVAB. This alternative measure of math aims to test the respondents "ability to solve arithmetic word problems." Differences in these scores may reflect deeper math skills rather than curriculum exposure. Column (5) compares this measure of math to the verbal component and finds that verbal is significantly more important than math, with magnitudes closer to those found in our main regressions in Table 9. Column (6) offers further analysis by interacting these skills with age of test taker. Unlike MK, the effect of AR on college enrollment is roughly constant for all test takers with results very similar to those in Column (5).

Finally, we study the impact of subject-specific grades on college enrollment for a subset of the respondents in the high school transcript data. Course grades possibly offer more breadth as skill measures than the ASVAB subject scores. Whereas the ASVAB MK section only contains 15 questions and PC only contains 10 questions, course grades represent a measure taken over a long period of time, which is directly linked to school performance. This aspect makes course grades most similar to the type of data used in the UK analysis. In this analysis, we focus only on ninth graders who were enrolled in Algebra I and ninth grade English concurrently and earned credit (i.e. did not fail) in both courses, and we compare the impact of grades in these courses on college enrollment.[87] This analysis leaves out students who took Algebra I in the eighth grade (high-achieving students)

[86]Given that older students are more likely to be exposed to different sets of math courses than younger students, results based on these students are less likely to be contaminated by an "exposure effect" and, therefore, are more reliable.

[87]We exclude those who failed or did not receive credit because many of these students were either expelled, suspended, or dropped out. Because we cannot easily distinguish between these students and those who simply had unsatisfactory performance, we look only at those with a D grade or better in both classes. We focus exclusively on ninth graders because for many states in the US, Algebra I is required for graduation and taken in ninth grade. Math course taking beyond ninth grade is highly tailored to student preferences and ability. Secondly ninth grade English curriculum appears to be more uniform across schools. For example, in tenth grade, some schools offer courses classified as "English Survey, Basic, Grade 10" and others offer "Literature."

and students taking Pre-Algebra in the ninth grade (low-achieving students).[88] Column (7) of Table 13 shows the impact of a student's ninth grade English grade and Algebra I grade, measured in grade points, on college enrollment. These results show that moving a student up one letter grade in ninth grade English increases the probability of college enrollment by 10 percentage points. In comparison, a similar increase in the Algebra I grade only increases the probability of enrollment by 7 percentage points. In summary, these findings using US data suggest that the key role of verbal skills in explaining college enrollment and graduation does not appear to be unique to the UK education system.

7.2 Undermining the Returns to Verbal Skills: The Role of Intermediate Outcomes

There are two main channels through which skills can impact labor market outcomes. The first is a direct return to different skills in the labor market. The second is an indirect effect in which skills impact a worker's total years of schooling. Previous work on the causal effect of skills on labor market outcomes has predominantly focused on the effect of skills *net* of total years of schooling, i.e. controlling for level of schooling (Altonji et al., 2012; Rose and Betts, 2004; Levine and Zimmerman, 1995).[89] However, if different types of skills have a differential effect on schooling attainment, as our results suggest, then simultaneously controlling for skills and level of education may understate the overall importance of certain skills, shutting off the second channel stated above.[90] Given that our UK data do not have wages, we use the individuals in the NLSY97 that were analyzed in Table 13 to study the multiple channels in which skills affect wages. Using both subject-specific ASVAB scores and high school course grades as skill measures, we conduct simple OLS regressions on log wages with and without a control for college degree to study how the inclusion of this control impacts the implied returns to different skills.

The first column in Table 14 shows the impact of the scores in paragraph comprehension (PC)

[88] While some of the transcript data contains information on eighth grade, most do not. In many cases it is not even possible to determine if a student took Algebra I in the eighth grade.

[89] Altonji et al. (2012) provides some discussion about this specific point.

[90] Neal and Johnson (1996) also discusses the implications of controlling for educational attainment in a wage regression.

Table 14: Linear Regression Model: Log Wage (NLSY97)

	(1)	(2)	(3)	(4)	(5)
College Graduate	–	–	0.3387*** (0.0224)	–	0.2941*** (0.0323)
ASVAB test specific scores†					
Paragraph Comprehension (PC)	0.0358** (0.0149)	0.0550** (0.0250)	0.0114 (0.0234)	–	–
Arithmetic Reasoning (AR)	0.1369*** (0.0159)	0.0866*** (0.0276)	0.0624** (0.0254)	–	–
Age at ASVAB (age)	0.0171** (0.0082)	0.0142* (0.0083)	-0.0008 (0.0078)	–	–
(PC) × (age-13)	–	-0.0094 (0.0105)	-0.0038 (0.0098)	–	–
(AR) × (age-13)	–	0.0245** (0.0111)	0.0166 (0.0103)	–	–
Subject Course Grades‡					
Ninth Grade English	–	–	–	0.0440** (0.0217)	0.0018 (0.0214)
Algebra 1	–	–	–	0.0787*** (0.0187)	0.0536*** (0.0184)
R-squared	0.136	0.138	0.217	0.093	0.160
Obs.	2448	2448	2448	1060	1060

† Individual test scores are standardized by age when the test was administered.
‡ Course grades are in grade points, e.g. A is 4.0, B is 3.0, etc.
Notes: This analysis only includes the NLSY97 cross-sectional sample of 6,748 respondents designed to be representative of people living in the United States. Wages are only analyzed for respondents in their most recent survey year if they were a full-time (35+ hours per week), full-year (50+ weeks per year) worker. The top 1% ($120+/hr) of wages and those less than $5/hr were removed. College graduate is defined as having a Bachelor's degree or more at their most recent interview. Those whose age at last interview was less than 25 were not included in this analysis. Round 1 observation weights were used for all regressions. All specifications include controls for gender, ethnicity, race, actual full-time experience, and experience squared.

and arithmetic reasoning (AR) on wages, where the scores have been standardized by age of test taker. In Column 2, we interact these scores with the age of test taker. These results show that, for 13 year olds a one standard deviation increase in PC will increase future earnings by 5.5%, while a one standard deviation increase in AR will increase future earnings by 8.6%. Column 3 shows what happens to the return to skills when an indicator is included for level of schooling. Unsurprisingly, given that PC has a larger impact on college enrollment than AR, adding a control for college completion significantly reduces the coefficient on the verbal skill to 0.011, which is not statistically different from zero, and represents an 80% reduction in the return to verbal skills. In contrast, including the control for level of education only reduces the coefficient on AR by 30%. This large decline on the verbal skill coefficient when level of education is included in the regression has also been documented in the context of a different question by Fredriksson et al. (2015). More specifically, Table 1 of their paper shows, using Swedish administrative data, that the labor market returns to verbal skills suffer a substantially larger drop (i.e. from 0.0253 to 0.0031) than math skills (i.e. from 0.0373 to 0.0216) once controls for educational attainment are included in the econometric specification. Shutting off the channel in which verbal skill affects wages through level of education, as is done in Column 3 of Table 14, may significantly understate the importance of verbal skills for labor market outcomes.

Columns 4 and 5 in Table 14 perform a similar analysis using course grades as an alternative measure of skill. As before, we only include ninth graders who enrolled in and earned credit in Algebra I and ninth grade English. These results are consistent with our findings for the ASVAB subject tests, where the return to the verbal skills goes from positive and statistically significant to statistically insignificant once the control for schooling level is added. To conclude, the fact that controlling for schooling largely undermines the role of verbal skills on wages may partially explain why the economics literature (Levine and Zimmerman, 1995; Joensen and Nielsen, 2009; Cortes et al., 2015; Dougherty et al., 2015) and policymakers (e.g. the "Algebra-for-All" movement, Loveless (2008)) have mainly prioritized their attention to math skills over verbal skills (Long et al., 2012).

8 Understanding the Gender Gap in College Enrollment and STEM Majors

Evidence from many countries (OECD, 2012) has shown that males are less likely to attend university than females, but they are more likely to enroll in STEM fields. In particular, the gender gap in college enrollment is 6.6 percentage points in our sample, while the STEM gap conditional on enrollment is -16.9 percentage points. Despite the fact that these empirical regularities have been extensively described in the literature, there is not much evidence regarding the role of math and verbal skills in explaining these gaps.[91] Section 5.1 shows that while there are small gender differences in math skills, females have a large advantage in verbal skills. This fact, combined with our findings on the importance of verbal skills in explaining college enrollment, suggests a possible mechanism that could explain the gender gap in educational attainment.

Gender Gap in College Enrollment To analyze the extent to which differences in verbal skills could explain the gender gap in college enrollment, we rely on the results of the logistic college enrollment regression model presented in Table 9 to perform a set of simple decomposition exercises.[92] First, Table 15 shows that giving males the female coefficients from the enrollment probability model cannot explain the gap. In fact, males seem to show larger "preferences" for college than females.[93] In a similar vein, Table 15 also indicates that equalizing the mean math skill between gender groups has almost no impact on the gap, which is expected given that gender differences in math skills are close to zero. Finally, equalizing the mean verbal skills between males and females (i.e. shifting up the male distribution of verbal skills to have the same mean as female verbal skills) provides a completely different picture, where the gap is fully explained (i.e. females becoming 0.9 percentage points less likely to attend college than males). In summary, this simple decomposition exercise strongly indicates that gender differences in verbal skills play a strong role in the gender gap in college enrollment.

[91] See for example, Buchmann and DiPrete (2006); Becker et al. (2010); Autor and Wasserman (2013); Osikominu and Pfeifer (2018); Aucejo and James (2018).

[92] In particular, we use the coefficients from Column (4) in Table 9.

[93] This result is consistent with the findings in Becker et al. (2010).

Table 15: Gender Gap in College Enrollment

	Female	Male	Difference
Baseline College Enrollment	0.396 (0.002)	0.330 (0.002)	0.066 (0.002)
Males w/ Coefficeints of Females	–	0.325 (0.002)	0.071 (0.001)
Equalize Mean Math of Males to Female Mean	–	0.331 (0.002)	0.065 (0.001)
Equalize Mean Verbal of Males to Female Mean	–	0.405 (0.002)	-0.009 (0.002)

Note: Baseline college enrollment corresponds to the actual enrollment rates. Decomposition exercises in rows 2-4 rely on the the results of the logistic regression model presented in Table 9 Column (4). Bootstrapped standard errors in parentheses.

Gender Gap in STEM Conditional on College Enrollment Table 16 focuses on the STEM gap. First, it shows that males have slightly higher average math skills than females once conditioning on college enrollment, while females still hold a large advantage in verbal skills. Our estimates of the KS4 skills also indicate that males display a comparative advantage in math, while females display a comparative advantage in verbal. More specifically, the average difference between math and verbal skills among females is -0.292 log points, while among males it is 0.177 log points.[94] Given this fact, we are interested in determining the extent to which comparative advantage in math could contribute to explain the gender gap in STEM. To isolate its effect, we first consider a simple linear specification model:[95]

$$y_i^* = z_i'\omega + \theta_{i,4}^m\beta^m + \theta_{i,4}^v\beta^v + \varphi(\pi_i) + \varepsilon_i \tag{12}$$

Let $\delta = \beta^m + \beta^v$ be the effect of a joint one unit increase in skills on the outcome y^* (e.g., if both skills are increased by 1 then y^* will increase by δ). Re-writing δ in terms of β^m and substituting into Eq. (12) provides an analogous specification form that makes easier to interpret the role of comparative advantage:

$$y_i^* = z_i'\omega + \theta_{i,4}^m\delta + (\theta_{i,4}^m - \theta_{i,4}^v)\gamma + \varphi(\pi_i) + \varepsilon_i \tag{13}$$

[94] 67% of males show a percentile rank in math skills that is larger than the analogous rank in verbal skills, while only 33% of females present such an advantage.

[95] We rely on a linear model because it simplifies the interpretation of the parameters.

Table 16: Gender Gap in STEM Enrollment Conditional on College Attendance

	Female	Male	Difference
Baseline STEM Major Probability	0.223	0.392	-0.169
	(0.001)	(0.002)	(0.002)
Mean Math Conditional on College Enrollment	8.368	8.501	-0.133
	(0.007)	(0.005)	(0.008)
Mean Verbal Conditional on College Enrollment	8.660	8.324	0.336
	(0.004)	(0.004)	(0.005)
Mean Math Comparative Advantage $(\theta_m - \theta_v)$ Conditional on College Enrollment	-0.292	0.177	-0.469
	(0.004)	(0.003)	(0.004)
δ	0.124	-0.018	0.142
	(0.012)	(0.011)	(0.015)
γ	0.501	0.720	-0.219
	(0.020)	(0.017)	(0.024)
STEM Major Probability: Equalize Comparative Advantage	0.264	_	-0.127
	(0.003)		(0.002)
STEM Major Probability: Equalize Comparative Advantage and γ	0.277	_	-0.115
	(0.002)		(0.002)

Note: See Eq .(13) to interpret δ and γ. *Equalize comparative advantage* reports counterfactual female STEM enrollment and the STEM enrollment gap after equalizing the term $\theta^m_{i,4} - \theta^v_{i,4}$ for males and females. *Equalize comparative advantage and* γ reports counterfactual female STEM enrollment and the STEM enrollment gap after equalizing for both gender groups all the channels in which comparative advantage could operate. Baseline controls include race, mother tongue, Key Stage 4 school quality, free school lunch, special education needs, number of GCSE exams taken in KS4, number of excused and unexcused absences in KS4, neighborhood characteristics, and unobserved persistent TFP shock. Bootstrapped standard errors in parentheses.

While $\gamma = -\beta^v$ from Eq. (12), its interpretation in this specification is the partial effect of increasing the individual's comparative advantage holding fixed their math skill level. Thus, Eq. (13) specifies that enrolling in a STEM major depends on the level of math skills, a comparative advantage term, and some baseline controls. Table 16 shows key coefficients (δ, γ) for males and females. While STEM enrollment for females is relatively more responsive to their overall skills advantage (i.e. δ, level of math and verbal skills), males are more responsive to their comparative advantage (i.e. γ) than females. The fact that males have a large math comparative advantage, and are more responsive to it, suggests a channel to explain the STEM gap. In this regard, we performed two simple decomposition exercises. First, we equalized the comparative advantage term (i.e. $\theta^m_{i,4} - \theta^v_{i,4}$) of males and females while holding fixed their math level. Results in Table 16 show that the gender gap in STEM enrollment would drop by 4.2 percentage points, corresponding to a 25% decrease in the gap. Second, we analyzed the overall drop in the gender STEM gap when shutting down all the channels in which comparative advantage can operate. This implies equalizing $\theta_{i,4}^m - \theta_{i,4}^v$ and the differential gender responses to it (γ). The last row of Table 16 shows that the STEM gap would drop by an additional 1.2 percentage points, therefore explaining 32% of the STEM gap. Overall, we interpret these findings as indicating that comparative advantage is an important predictor of the gender gap in STEM.

To conclude, our results show that gender differences in verbal skills are a key factor in explaining the gender gap in college enrollment. However, the female advantage in verbal skills also translates into differential comparative advantages between gender groups which has important implications in shaping the gender STEM gap (conditional on college enrollment).

9 Conclusion

This paper estimates a dynamic factor model of skill formation with the aim of understanding how math and verbal skills develop during compulsory education, and to study their impact on adult outcomes. In addition, we further contribute to the literature on skill formation by proposing a novel estimation approach of nested CES production functions with three inputs.

We find that the production of math and verbal skills is inherently different. In particular, cross-effects are only present in the production of math skills. We also show that school quality plays an important role in the formation of skills, affecting adult outcomes. Counterfactual simulations show that improving school quality could increase college enrollment by 10%.

The analysis on adult outcomes indicates that the effect of verbal skills on university enrollment is at least twice as large as the effect of math skills. In this regard, we find that targeting verbal skills rather than math skills could be more efficient if policymakers aim to improve overall educational attainment. We also document that females hold a large advantage in verbal skills relative to males, which explains the gender gap in college enrollment. However, males' comparative advantage in math skills jointly with gender differences in response to comparative advantage contribute to increased male representation in STEM majors.

Finally, our finding on the predominant effect of verbal skills on college enrollment suggests that the role of these skills in explaining labor market outcomes could have been artificially undermined in previous studies. We show that log wage specifications that simultaneously control for skills and educational attainment tend to largely diminish the effect of verbal skills on wages due to the inclusion of an endogenous variable (i.e., educational attainment). By including educational attainment, previous work has inadvertently shut off one of the main mechanisms through which verbal skills affect wages.

To conclude, while the many curriculum based policy proposals designed to increase college attendance focus on enhancing math skills (e.g., the Algebra-for-All movement), our findings suggest broadening the scope of these types of policies to improve verbal skills as well.

References

D. Acemoglu. Why do new technologies complement skills? directed technical change and wage inequality. *The Quarterly Journal of Economics*, 113(4):1055–1089, 1998.

F. Agostinelli and M. Wiswall. Estimating the technology of children's skill formation. Technical report, National Bureau of Economic Research, 2016.

J. Altonji, E. Blom, and C. Meghir. Heterogeneity in human capital investments: High school curriculum, college major, and careers. *NBER working Paper Series 17985*, pages 1–35, 2012.

O. Attanasio, C. Meghir, and E. Nix. Human capital development and parental investment in india. Technical report, National Bureau of Economic Research, 2015.

O. Attanasio, C. Meghir, E. Nix, and F. Salvati. Human capital growth and poverty: Evidence from ethiopia and peru. *Review of economic dynamics*, 25:234–259, 2017.

E. Aucejo and J. James. Catching up to girls: Understanding the gender imbalance in educational attainment within race. 2018.

D. Autor and M. Wasserman. Wayward sons: The emerging gender gap in labor markets and education. *Third Way Report*, 2013.

Z. Bakk, F. B. Tekle, and J. K. Vermunt. Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology*, 43(1):272–311, 2013.

G. S. Becker, W. H. Hubbard, and K. M. Murphy. Explaining the worldwide boom in higher education of women. *Journal of Human Capital*, 4(3):203–241, 2010.

J. R. Betts. Does school quality matter? evidence from the national longitudinal survey of youth. *The Review of Economics and Statistics*, pages 231–250, 1995.

C. Buchmann and T. A. DiPrete. The growing female advantage in college completion: The role of family background and academic achievement. *American sociological review*, 71(4):515–541, 2006.

S. Cameron and J. Heckman. The dynamics of educational attainment for black, hispanic, and white males. *Journal of Political Economy*, 109(3):pp. 455–499, 2001.

G. Castex and E. K. Dechter. The changing roles of education and ability in wage determination. *Journal of Labor Economics*, 32(4):pp. 685–710, 2014.

J. Cawley, J. Heckman, and E. Vytlacil. Three observations on wages and measured cognitive ability. *Labour Economics*, 8(4):pp. 419–42, 2001.

R. Chetty, J. N. Friedman, and J. E. Rockoff. Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *The American Economic Review*, 104(9):2633–2679, 2014.

K. E. Cortes, J. S. Goodman, and T. Nomi. Intensive math instruction and educational attainment long-run impacts of double-dose algebra. *Journal of Human Resources*, 50(1):108–158, 2015.

F. Cunha and J. Heckman. The technology of skill formation. Technical report, National Bureau of Economic Research, 2007.

F. Cunha and J. Heckman. A new framework for the analysis of inequality. *Macroeconomic Dynamics*, 12(S2):315–354, 2008a.

F. Cunha and J. J. Heckman. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of human resources*, 43(4):738–782, 2008b.

F. Cunha, J. Heckman, and S. Navarro. Separating uncertainty from heterogeneity in life cycle earnings. *oxford Economic papers*, 57(2):191–261, 2005.

F. Cunha, J. J. Heckman, and S. Navarro. Counterfactual analysis of inequality and social mobility. *Mobility and inequality: Frontiers of research in sociology and economics*, pages 290–348, 2006a.

F. Cunha, J. J. Heckman, and S. Navarro. The evolution of earnings risk in the us economy. In *9th World Congress of the Econometric Society, London*, 2006b.

F. Cunha, J. J. Heckman, and S. M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

S. Dougherty, J. Goodman, D. Hill, E. Litke, and L. Page. Early math coursework and college readiness: Evidence from targeted middle school math acceleration. *NBER working Paper Series*, pages 1–50, 2015.

P. Fredriksson, L. Hensvik, and O. Nordstrom Skans. Mismatch of talent: Evidence on match quality, entry wages, and job mobility. *IZA Discussion Paper Series No. 9585*, pages 1–54, 2015.

E. A. Hanushek. *Economic outcomes and school quality*, volume 4. International Institute for Educational Planning Paris, 2005.

J. Heckman, R. Pinto, and P. Savelyev. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *The American Economic Review*, 103(6):1–35, 2013.

J. J. Heckman, J. Stixrud, and S. Urzua. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–482, 2006.

G. Hobbs and A. Vignoles. *Is free school meal status a valid proxy for socio-economic status (in schools research)?* Centre for the Economics of Education, London School of Economics and Political Science, 2007.

J. James. Mm algorithm for general mixed multinomial logit models. *Journal of Applied Econometrics*, 32(4):841–857, 2017.

J. S. Joensen and H. S. Nielsen. Is there a causal effect of high school math on labor market outcomes? *Journal of Human Resources*, 44(1):171–198, 2009.

R. Klump and O. de La Grandville. Economic growth and the elasticity of substitution: Two theorems and some suggestions. *American Economic Review*, 90(1):282–291, 2000.

R. Klump, P. McAdam, and A. Willman. Factor substitution and factor-augmenting technical progress in the united states: a normalized supply-side system approach. *The Review of Economics and Statistics*, 89(1):183–192, 2007a.

R. Klump, P. McAdam, and A. Willman. The long-term success of the neoclassical growth model. *Oxford Review of Economic Policy*, 23(1):94–114, 2007b.

P. Krusell, L. E. Ohanian, J.-V. Ríos-Rull, and G. L. Violante. Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica*, 68(5):1029–1053, 2000.

O. La Grandville. In quest of the slutsky diamond. *The American Economic Review*, pages 468–481, 1989.

O. d. La Grandville and R. Solow. Capital-labour substitution and economic growth. *Economic Growth: A Unified Approach*, pages 389–416, 2009.

L.-F. Lee. Asymptotic bias in simulated maximum likelihood estimation of discrete choice models. *Econometric Theory*, 11(3):437–483, 1995.

P. B. Levine and D. J. Zimmerman. The benefit of additional high-school math and science classes for young men and women. *Journal of Business & Economic Statistics*, 13(2):137–149, 1995.

M. C. Long, D. Conger, and P. Iatarola. Effects of high school course-taking on secondary and postsecondary success. *American Educational Research Journal*, 49(2):285–322, 2012.

T. Loveless. *The Misplaced Math Student: Lost in Eighth Grade Algebra*. Washington, DC: Brown Center on Education Policy, Brookings Institution, 2008.

G. S. Maddala. *Introduction to econometrics*. Macmillan New York, second edition, 1992.

D. A. Neal and W. R. Johnson. The role of premarket factors in black-white wage differences. *Journal of political Economy*, 104(5):869–895, 1996.

OECD. Education at a glance 2012: Highlights. Technical report, OECD, 2012.

A. Osikominu and G. Pfeifer. Perceived wages and the gender gap in stem fields. 2018.

M. Pandey. Human capital aggregation and relative wages across countries. *Journal of Macroeconomics*, 30(4):1587–1601, 2008.

H. Rose and J. R. Betts. The effect of high school courses on earnings. *Review of Economics and Statistics*, 86(2):497–513, 2004.

K. Sato. A two-level constant-elasticity-of-substitution production function. *The Review of Economic Studies*, 34(2):201–218, 1967.

M. B. Stewart. On least squares estimation when the dependent variable is grouped. *The Review of Economic Studies*, 50(4):737–753, 1983.

J. Temple. The calibration of ces production functions. *Journal of Macroeconomics*, 34(2):294–303, 2012.

P. E. Todd and K. I. Wolpin. On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113:3–33, 2003.

Appendix

A Parameter Bias With Multiple Variables Measured With Error

Consider the data generating process $Y = \beta_1 x_1 + \beta_2 x_2 + v$. Only noisy measures of the variables x_1 and x_2 are observed:

$$X_1 = x_1 + u_1$$
$$X_2 = x_2 + u_2$$

Without loss of generality, normalize $\operatorname{Var}(X_1) = \operatorname{Var}(X_2) = 1$ and let $\lambda_1 = \operatorname{Var}(u_1)/\operatorname{Var}(X_1)$ and $\lambda_2 = \operatorname{Var}(u_2)/\operatorname{Var}(X_2)$ denote the fraction of the variance of the observed variables measured with error.

Under the assumption of classical measurement error, Maddala (1992) gives formulas for the probability limits of b_1 and b_2, the OLS estimates of $\widehat{Y} = b_1 X_1 + b_2 X_2$ that are a function of β_1, β_2, λ_1, λ_3, and ρ, where $\rho = \operatorname{Cov}(X_1, X_2)$:

$$\operatorname{plim} b_1 = \beta_1 - \frac{\beta_1\lambda_1 - \rho\beta_2\lambda_2}{1-\rho^2}$$
$$\operatorname{plim} b_2 = \beta_2 - \frac{\beta_2\lambda_2 - \rho\beta_1\lambda_1}{1-\rho^2}$$

These equations suggest a rather surprising result. If $\lambda_1 = \lambda_2$, the level of measurement error of both of the variables is the same, then regardless of the true ratio of the parameters, β_1 and β_2, the ratio of the OLS estimates converges to 1 as x_1 and x_2 become more correlated. To see this, let $\operatorname{Cov}(x_1, x_2) = \rho^*$. Since $\operatorname{Cov}(X_1, X_2) = \operatorname{Cov}(x_1, x_2)$, then $\rho = \rho^*\sqrt{(1-\lambda_1)(1-\lambda_2)}$. Setting $\lambda_1 = \lambda_2$ and taking the ratio of the probability limits we have

$$\begin{aligned}\frac{\operatorname{plim} b_1}{\operatorname{plim} b_2} =& \frac{\beta_1(1-\rho^2) - \lambda(\beta_1 - \rho\beta_2)}{\beta_2(1-\rho^2) - \lambda(\beta_2 - \rho\beta_1)} \\ =& \frac{\beta_1\left[1-(\rho^*)^2(1-\lambda)^2\right] - \lambda(\beta_1 - \rho^*(1-\lambda)\beta_2)}{\beta_2\left[1-(\rho^*)^2(1-\lambda)^2\right] - \lambda(\beta_2 - \rho^*(1-\lambda)\beta_1)}\end{aligned}$$

Taking the limit as ρ^* approaches 1:

$$\lim_{\rho^* \to 1} \frac{\operatorname{plim} b_1}{\operatorname{plim} b_2} = 1$$

B Interval Censored Measurements

This section provides two examples of the different types of variables that are used in the measurement system. In particular, Figure 3 illustrates a common issue facing researchers when analyzing educational outputs. The plot on the left shows a histogram of the Key Stage 2 Reading Test, which we use as a measure to identify $\theta^v_{i,2}$. This variable is approximately continuous on the scale of 0 to 50. The plot on the right shows the histogram of another measure we use to identify $\theta^v_{i,2}$, which is the Key Stage 2 English Teacher Assessment. Unlike the previous plot of the reading test,

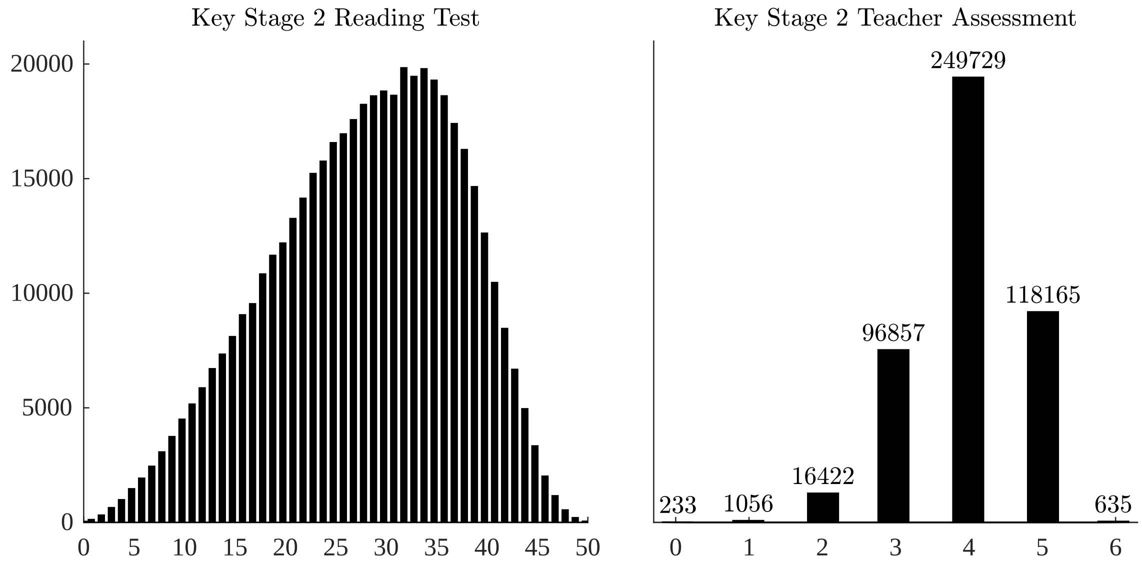


Figure 3: Histograms of Select Key Stage 2 Measurements

the teacher assessment is coarsely distributed across seven mass points. Notice that this type of data is very common in other data sets, for example any data with letter grades A–F.

C Estimation Algorithm for Factor Model

This section outlines the iterative algorithm used in the first stage to estimate the non-parametric distribution of the factors as well as all of the parameters of the measurement system and selection equation. Estimating these parameters requires maximizing Eq. (7). Our estimation approach uses the MM algorithm proposed in James (2017) and incorporates the results in Stewart (1983) to address the interval coded data. Both of the these methods are based on the ideas of the expectation-maximization algorithm in Dempster et al. (1977). To simplify the description of the algorithm, we define the following likelihood elements

$$
\begin{aligned}
f(\theta) &= \sum_{d=1}^{D} \kappa_d K_d(\theta) && \text{PDF of the factors} \\
L_{\mathcal{J}_{cont}}(w_i|\theta) &= \prod_{j \in \mathcal{J}_{cont}} \Pr(w_{ij}|\theta) && \text{Likelihood of observed continuous measures} \\
L_{\mathcal{J}_{cens}}(w_i|\theta) &= \prod_{j \in \mathcal{J}_{cens}} \Pr(w_{ij}|\theta) && \text{Likelihood of observed censored measures} \\
L_{sel}(a_i|\theta) &= \prod_{c=1}^{C} \Pr(a_{ic}|\theta) && \text{Likelihood of observed selection outcomes}
\end{aligned}
$$

The likelihood in Eq. (7) can be written as

$$LL(\Psi) = \sum_{i=1}^{n} \ln \left[\int_{\theta} L_{\mathcal{J}_{cont}}(w_i|\theta) L_{\mathcal{J}_{cens}}(w_i|\theta) L_{sel}(a_i|\theta) f(\theta) d\theta \right]$$

Where $\Psi \in \{\mu, \lambda, \sigma, \tau, \kappa, \xi, \Delta\}$ includes all of the first stage parameters.

The algorithm begins with an initial value of the parameter vector $\Psi^{(0)}$. Given these parameters, using the likelihood function above, the posterior distribution for each individual given their data is constructed using

$$h(\theta|w_i, a_i, \Psi^{(0)}) = \frac{L_{\mathcal{J}_{cont}}(w_i|\theta) L_{\mathcal{J}_{cens}}(w_i|\theta) L_{sel}(a_i|\theta) f(\theta)}{\int_{\theta'} L_{\mathcal{J}_{cont}}(w_i|\theta') L_{\mathcal{J}_{cens}}(w_i|\theta') L_{sel}(a_i|\theta') f(\theta') d\theta'}$$

Using these distribution functions, Dempster et al. (1977) shows that the original likelihood can be bound below around $\Psi^{(0)}$ by the function:

$$Q(\Psi|\Psi^{(0)}) = \sum_{i=1}^{n} \int_{\theta} \ln \left[L_{\mathcal{J}_{cont}}(w_i|\theta) L_{\mathcal{J}_{cens}}(w_i|\theta) L_{sel}(a_i|\theta) f(\theta) \right] h(\theta|w_i, a_i, \Psi^{(0)}) d\theta \tag{14}$$

By finding $\Psi^{(1)}$ that maximizes this function above produces a new set of values that are guaranteed to improve the likelihood, i.e., $LL(\Psi^{(1)}) > LL(\Psi^{(0)})$ replacing $\Psi^{(0)}$ with $\Psi^{(1)}$ and repeating the process produces a sequence of estimates that converges to the maximum of the likelihood function.

The integral in Eq. (14) does not have a closed form and must be simulated by drawing from $h(\theta|w_i, a_i, \Psi^{(m)})$, where m denotes the mth iteration of the algorithm. To do this, we take R draws of θ, each labeled $\theta_{ir}^{(m)}$ from $f(\theta|\Psi^{(m)})$ and then compute the weight

$$w_{ir}^{(m)} = \frac{L_{\mathcal{J}_{cont}}(w_i|\theta_{ir}^{(m)}) L_{\mathcal{J}_{cens}}(w_i|\theta_{ir}^{(m)}) L_{sel}(a_i|\theta_{ir}^{(m)})}{\sum_{r'=1}^{R} L_{\mathcal{J}_{cont}}(w_i|\theta_{ir'}^{(m)}) L_{\mathcal{J}_{cens}}(w_i|\theta_{ir'}^{(m)}) L_{sel}(a_i|\theta_{ir'}^{(m)})}$$

The lower bound function at the mth iteration becomes

$$Q(\Psi|\Psi^{(m)}) = \sum_{i=1}^{n} \sum_{r=1}^{R} w_{ir}^{(m)} \ln \left[L_{\mathcal{J}_{cont}}(w_i|\theta_{ir}^{(m)}) L_{\mathcal{J}_{cens}}(w_i|\theta_{ir}^{(m)}) L_{sel}(a_i|\theta_{ir}^{(m)}) f(\theta_{ir}^{(m)}) \right] \tag{15}$$

The lower bound function is a complete data likelihood that treats θ as observed and includes the weights. So the parameters of the model that maximize this function has a familiar form. For example the loadings for the continuous measures can be found with weighted OLS. The parameters for the censored variables can be found using the closed form solutions in Stewart (1983), and the parameters for the selection equations can be found by applying the closed form solutions in James (2017). The parameter updates for the mixture distribution are found by computing

$$\hat{\kappa}_{ir} = w_{ir}^{(m)} \frac{\kappa_d^{(m)} normpdf(\theta_{ir}^{(m)}, \xi_d^{(m)}, \Delta_d^{(m)})}{\sum_{d'=1}^{D} \kappa_{d'}^{(m)} normpdf(\theta_{ir}^{(m)}, \xi_{d'}^{(m)}, \Delta_{'d}^{(m)})}$$

Then we have the following updating equations for all d

$$\begin{aligned}
\kappa_d^{(m+1)} &= \frac{\sum_{i=1}^n \sum_{r=1}^R \hat{\kappa}_{ir}}{n} \\
\xi_d^{(m+1)} &= \frac{\sum_{i=1}^n \sum_{r=1}^R \hat{\kappa}_{ir}\theta_{ir}^{(m)}}{\sum_{i=1}^n \sum_{r=1}^R \hat{\kappa}_{ir}} \\
\Delta_d^{(m+1)} &= \frac{\sum_{i=1}^n \sum_{r=1}^R \hat{\kappa}_{ir}\theta_{ir}^{(m)}(\theta_{ir}^{(m)})'}{\sum_{i=1}^n \sum_{r=1}^R \hat{\kappa}_{ir}} - \xi_d^{(m+1)}(\xi_d^{(m+1)})'
\end{aligned}$$

We use $R = 4,500$ and define convergence when the first, second, third, and fourth uncentered moments of the factor distribution change by less than $1/10$ of one percent.

D Marginal Effects

This section derives the closed form expressions for the marginal effects for the 3-input CES production function characterized in Eq. (1), which takes the form

$$y = e^A \left[\delta \left(\alpha x_1^\gamma + (1-\alpha)x_2^\gamma\right)^{\rho/\gamma} + (1-\delta)x_3^\rho\right]^{r/\rho} e^\varepsilon$$

We are interested in expressions for $\partial \ln y / \partial \ln x_k$ for $k \in \{1, 2, 3\}$. We begin writing the log of output.

$$\ln y = A + \frac{r}{\rho} \ln \left[\delta \left(\alpha x_1^\gamma + (1-\alpha)x_2^\gamma\right)^{\rho/\gamma} + (1-\delta)x_3^\rho\right] + \varepsilon$$

$$\begin{aligned}
\frac{\partial \ln y}{\partial x_1} &= \frac{r}{\rho}\frac{\rho}{\gamma}\frac{\gamma}{1}\left(\frac{\delta\left(\alpha x_1^\gamma + (1-\alpha)x_2^\gamma\right)^{\rho/\gamma}}{\delta\left(\alpha x_1^\gamma + (1-\alpha)x_2^\gamma\right)^{\rho/\gamma} + (1-\delta)x_3^\rho}\right)\left(\frac{\alpha x_1^\gamma}{\alpha x_1^\gamma + (1-\alpha)x_2^\gamma}\right)\frac{1}{x_1} \\
\frac{\partial \ln y}{\partial x_1 / x_1} &= r\left(\frac{\delta\left(\alpha x_1^\gamma + (1-\alpha)x_2^\gamma\right)^{\rho/\gamma}}{\delta\left(\alpha x_1^\gamma + (1-\alpha)x_2^\gamma\right)^{\rho/\gamma} + (1-\delta)x_3^\rho}\right)\left(\frac{\alpha x_1^\gamma}{\alpha x_1^\gamma + (1-\alpha)x_2^\gamma}\right) \\
\frac{\partial \ln y}{\partial \ln x_1} &= r\left(\frac{\delta\left(\alpha x_1^\gamma + (1-\alpha)x_2^\gamma\right)^{\rho/\gamma}}{\delta\left(\alpha x_1^\gamma + (1-\alpha)x_2^\gamma\right)^{\rho/\gamma} + (1-\delta)x_3^\rho}\right)\left(\frac{\alpha x_1^\gamma}{\alpha x_1^\gamma + (1-\alpha)x_2^\gamma}\right)
\end{aligned}$$

Repeating similar steps for the other inputs

$$\begin{aligned}
\frac{\partial \ln y}{\partial \ln x_2} &= r\left(\frac{\delta\left(\alpha x_1^\gamma + (1-\alpha)x_2^\gamma\right)^{\rho/\gamma}}{\delta\left(\alpha x_1^\gamma + (1-\alpha)x_2^\gamma\right)^{\rho/\gamma} + (1-\delta)x_3^\rho}\right)\left(\frac{(1-\alpha)x_2^\gamma}{\alpha x_1^\gamma + (1-\alpha)x_2^\gamma}\right) \\
\frac{\partial \ln y}{\partial \ln x_3} &= r\left(\frac{(1-\delta)x_3^\rho}{\delta\left(\alpha x_1^\gamma + (1-\alpha)x_2^\gamma\right)^{\rho/\gamma} + (1-\delta)x_3^\rho}\right)
\end{aligned}$$

E Measurement System and Normalizations

Table 17: Measurement System and Normalizations: Female

No.	Description	Data Type	No. of Intervals	Intercept	Loadings: KS1 Math	KS1 Verbal	KS2 Math	KS2 Verbal	KS3 Math	KS3 Verbal	KS4 Math	KS4 Verbal	KS4 Motive	Percent Noise
1	Math Test	Intervaled	6	0	1	0	0	0	0	0	0	0	0	0.181 (0.001)
2	Math Using and Applying TA	Intervaled	4	0.414 (0.003)	0.875 (0.001)	0	0	0	0	0	0	0	0	0.111 (0.001)
3	Math Number and Algebra TA	Intervaled	4	0.545 (0.002)	0.863 (0.001)	0	0	0	0	0	0	0	0	0.086 (0.001)
4	Math Shapes and Measure TA	Intervaled	4	0.545 (0.003)	0.844 (0.001)	0	0	0	0	0	0	0	0	0.123 (0.002)
5	Writing Test	Intervaled	6	0	0	1	0	0	0	0	0	0	0	0.148 (0.001)
6	Writing TA	Intervaled	4	0.255 (0.002)	0	1.023 (0.001)	0	0	0	0	0	0	0	0.083 (0.001)
7	Reading TA	Intervaled	4	0.263 (0.004)	0	1.139 (0.002)	0	0	0	0	0	0	0	0.170 (0.001)
8	Listening TA	Intervaled	4	0.719 (0.005)	0	0.892 (0.002)	0	0	0	0	0	0	0	0.307 (0.002)
9	Math Test Paper A	Continuous	–	-26.573 (0.051)	0	0	11.308 (0.011)	0	0	0	0	0	0	0.123 (0.001)
10	Math Test Paper B	Continuous	–	-26.545 (0.070)	0	0	11.517 (0.014)	0	0	0	0	0	0	0.145 (0.001)
11	Math Arithmetic Test	Continuous	–	-17.005 (0.049)	0	0	6.604 (0.011)	0	0	0	0	0	0	0.189 (0.001)
12	Math TA	Intervaled	6	0	0	0	1	0	0	0	0	0	0	0.156 (0.004)
13	Reading Test	Continuous	–	-30.746 (0.086)	0	0	0	12.953 (0.019)	0	0	0	0	0	0.218 (0.001)
14	Writing Test	Continuous	–	-0.164 (0.078)	0	0	0	5.153 (0.016)	0	0	0	0	0	0.416 (0.002)
15	Spelling Test	Continuous	–	-6.757 (0.033)	0	0	0	2.945 (0.007)	0	0	0	0	0	0.402 (0.001)
16	English TA	Intervaled	6	0	0	0	0	1	0	0	0	0	0	0.132 (0.003)
17	Math Test Paper 1	Continuous	–	0.235 (0.004)	0	0	0	0	0.341 (0.001)	0	0	0	0	0.402 (0.002)
18	Math Test Paper 2	Continuous	–	0.194 (0.005)	0	0	0	0	0.352 (0.001)	0	0	0	0	0.408 (0.002)
19	Math Arithmetic Test	Continuous	–	-0.848 (0.006)	0	0	0	0	0.365 (0.001)	0	0	0	0	0.404 (0.001)
20	Math TA	Intervaled	7	0	0	0	0	0	1	0	0	0	0	0.108 (0.001)
21	Writing Test (Longer)	Continuous	–	-18.261 (0.091)	0	0	0	0	0	5.538 (0.016)	0	0	0	0.349 (0.001)
22	Reading Test	Continuous	–	-20.373 (0.059)	0	0	0	0	0	6.383 (0.009)	0	0	0	0.253 (0.001)
23	Writing Test (Shorter)	Continuous	–	-11.361 (0.053)	0	0	0	0	0	3.774 (0.009)	0	0	0	0.337 (0.001)
24	Reading Test (Shakespeare)	Continuous	–	-11.642 (0.039)	0	0	0	0	0	3.268 (0.007)	0	0	0	0.448 (0.002)
25	English TA	Intervaled	7	0	0	0	0	0	0	1	0	0	0	0.241 (0.001)
26	Math	Intervaled	8	0	0	0	0	0	0	0	1	0	0	0.097 (0.000)
27	English	Intervaled	8	0	0	0	0	0	0	0	0	1	0	0.111 (0.001)
28	Design and Technology: Graphic Products	Intervaled	8	-0.146 (0.071)	0	0	0	0	0	0	0.491 (0.012)	0.515 (0.013)	0	0.371 (0.003)

(*Continued on next page*)

Note: TA denotes teacher assessment. Any values of 0, 1, or -1 denote normalizations. Standard errors in parentheses.

Table 17: Measurement System and Normalizations: Female

No.	Description	Data Type	No. of Intervals	Intercept	Loadings: KS1 Math	KS1 Verbal	KS2 Math	KS2 Verbal	KS3 Math	KS3 Verbal	KS4 Math	KS4 Verbal	KS4 Motive	Percent Noise
29	Design and Technology: Resistant Materials Technology	Intervaled	8	0.704 (0.064)	0	0	0	0	0	0	0.497 (0.016)	0.427 (0.019)	0	0.391 (0.007)
30	Design and Technology: Textiles Technology	Intervaled	8	0.523 (0.051)	0	0	0	0	0	0	0.397 (0.013)	0.568 (0.017)	0	0.351 (0.005)
31	Art and Design	Intervaled	8	2.140 (0.022)	0	0	0	0	0	0	0.209 (0.007)	0.552 (0.006)	0	0.475 (0.003)
32	History	Intervaled	8	-4.775 (0.028)	0	0	0	0	0	0	0	1.516 (0.003)	0	0.154 (0.001)
33	Geography	Intervaled	8	-2.891 (0.040)	0	0	0	0	0	0	0.498 (0.006)	0.827 (0.009)	0	0.145 (0.001)
34	French	Intervaled	8	-1.459 (0.038)	0	0	0	0	0	0	0.384 (0.008)	0.727 (0.011)	0	0.213 (0.002)
35	German	Intervaled	8	-0.969 (0.056)	0	0	0	0	0	0	0.399 (0.009)	0.653 (0.012)	0	0.229 (0.003)
36	Business Studies	Intervaled	8	-2.102 (0.061)	0	0	0	0	0	0	0.390 (0.011)	0.833 (0.014)	0	0.214 (0.004)
37	Religious Studies	Intervaled	8	-3.041 (0.036)	0	0	0	0	0	0	0	1.373 (0.004)	0	0.212 (0.002)
38	Short Religious Studies	Intervaled	8	-3.037 (0.038)	0	0	0	0	0	0	0	1.329 (0.005)	0	0.277 (0.002)
39	Physical Education	Intervaled	8	0.071 (0.038)	0	0	0	0	0	0	0.436 (0.011)	0.545 (0.013)	0	0.372 (0.003)
40	Physics	Intervaled	8	-1.582 (0.049)	0	0	0	0	0	0	1.111 (0.005)	0	0	0.077 (0.002)
41	Chemistry	Intervaled	8	-1.324 (0.084)	0	0	0	0	0	0	1.095 (0.009)	0	0	0.085 (0.003)
42	Biology	Intervaled	8	-1.505 (0.058)	0	0	0	0	0	0	0.853 (0.011)	0.277 (0.013)	0	0.097 (0.003)
43	Drama	Intervaled	8	0.858 (0.042)	0	0	0	0	0	0	-0.003 (0.009)	0.899 (0.011)	0	0.404 (0.004)
44	Information Technology	Intervaled	8	-0.567 (0.088)	0	0	0	0	0	0	0.564 (0.018)	0.481 (0.024)	0	0.345 (0.004)
45	Short Information Technology	Intervaled	8	-1.005 (0.075)	0	0	0	0	0	0	0.585 (0.019)	0.441 (0.023)	0	0.411 (0.004)
46	Spanish	Intervaled	8	-2.263 (0.093)	0	0	0	0	0	0	0.343 (0.012)	0.868 (0.016)	0	0.245 (0.005)
47	Music	Intervaled	8	-0.916 (0.067)	0	0	0	0	0	0	0.255 (0.015)	0.817 (0.020)	0	0.366 (0.005)
48	Social Science	Intervaled	8	-1.412 (0.071)	0	0	0	0	0	0	0.146 (0.012)	0.996 (0.017)	0	0.303 (0.004)
49	Design and Technology: Electronic Products	Intervaled	8	0.688 (0.372)	0	0	0	0	0	0	0.548 (0.105)	0.370 (0.129)	0	0.371 (0.023)
50	Design and Technology: System and Control	Intervaled	8	0.213 (0.510)	0	0	0	0	0	0	0.661 (0.122)	0.258 (0.166)	0	0.330 (0.033)
51	English Literature	Intervaled	8	-0.777 (0.012)	0	0	0	0	0	0	0	1.092 (0.001)	0	0.176 (0.001)
52	Design and Technology: Food Technology	Intervaled	8	-0.107 (0.043)	0	0	0	0	0	0	0.352 (0.008)	0.693 (0.010)	0	0.294 (0.002)
53	Science	Intervaled	8	-0.547 (0.033)	0	0	0	0	0	0	0.455 (0.011)	0.544 (0.013)	0	0.199 (0.003)
54	Statistics	Intervaled	8	-0.120 (0.063)	0	0	0	0	0	0	0.994 (0.010)	-0.040 (0.014)	0	0.174 (0.004)
55	Medial, Film and Television Studies	Intervaled	8	-0.824 (0.034)	0	0	0	0	0	0	0.060 (0.013)	1.061 (0.014)	0	0.259 (0.003)
56	Fine Art	Intervaled	8	2.016 (0.040)	0	0	0	0	0	0	0.190 (0.014)	0.588 (0.015)	0	0.441 (0.005)
57	Office Technology	Intervaled	8	-0.205 (0.097)	0	0	0	0	0	0	0.534 (0.014)	0.496 (0.019)	0	0.293 (0.006)

(Continued on next page)

Note: TA denotes teacher assessment. Any values of 0, 1, or -1 denote normalizations. Standard errors in parentheses.

Table 17: Measurement System and Normalizations: Female

No.	Description	Data Type	No. of Intervals	Intercept	Loadings: KS1 Math	KS1 Verbal	KS2 Math	KS2 Verbal	KS3 Math	KS3 Verbal	KS4 Math	KS4 Verbal	KS4 Motive	Percent Noise
58	Home Economics: Child Development	Intervaled	8	-1.122 (0.051)	0	0	0	0	0	0	0.266 (0.015)	0.889 (0.017)	0	0.286 (0.005)
59	Italian	Intervaled	8	-0.790 (0.274)	0	0	0	0	0	0	0.212 (0.079)	0.878 (0.088)	0	0.382 (0.020)
60	Urdu	Intervaled	8	1.883 (0.215)	0	0	0	0	0	0	0.164 (0.057)	0.635 (0.066)	0	0.588 (0.017)
61	Additional Applied Science	Intervaled	8	0.085 (0.164)	0	0	0	0	0	0	0.436 (0.028)	0.523 (0.040)	0	0.192 (0.010)
62	Leisure and Tourism	Intervaled	8	-1.734 (0.107)	0	0	0	0	0	0	0.240 (0.031)	0.919 (0.032)	0	0.344 (0.006)
63	Applied ICT	Intervaled	8	-0.241 (0.136)	0	0	0	0	0	0	0.476 (0.026)	0.470 (0.032)	0	0.444 (0.010)
64	Applied Science	Intervaled	8	0.507 (0.055)	0	0	0	0	0	0	0.341 (0.015)	0.560 (0.015)	0	0.319 (0.005)
65	Health and Social Care	Intervaled	8	-0.974 (0.068)	0	0	0	0	0	0	0.207 (0.012)	0.899 (0.015)	0	0.350 (0.006)
66	Applied Business	Intervaled	8	-1.937 (0.129)	0	0	0	0	0	0	0.498 (0.027)	0.724 (0.032)	0	0.298 (0.007)
67	Double Science	Intervaled	8	-0.138 (0.014)	0	0	0	0	0	0	0.995 (0.002)	0	0	0.135 (0.001)
	Selection Equations													
68	Took any science course	Binary	–	-3.198 (0.040)	0	0	0	0	0.047 (0.009)	0.543 (0.009)	0	0	0.753 (0.006)	–
69	Took any college preperation science	Binary	–	-5.494 (0.063)	0	0	0	0	0.313 (0.008)	0.539 (0.011)	0	0	0.448 (0.009)	–
70	Took any advanced science	Binary	–	-13.970 (0.038)	0	0	0	0	0.957 (0.007)	0.461 (0.010)	0	0	0.403 (0.005)	–
71	Took any relgion course	Binary	–	-3.465 (0.057)	0	0	0	0	-0.146 (0.010)	0.442 (0.017)	0	0	0.610 (0.009)	–
72	Took Religious Studies	Binary	–	-5.213 (0.063)	0	0	0	0	-0.137 (0.013)	0.500 (0.016)	0	0	0.488 (0.011)	–
73	Took Design and Technology: Graphic Products	Binary	–	-4.016 (0.045)	0	0	0	0	0.100 (0.005)	0.108 (0.009)	0	0	0.131 (0.007)	–
74	Took Design and Technology: Resistant Materials Technology	Binary	–	-2.396 (0.032)	0	0	0	0	0.321 (0.006)	-0.529 (0.009)	0	0	0.101 (0.007)	–
75	Took Design and Technology: Textiles Technology	Binary	–	-2.577 (0.040)	0	0	0	0	0.030 (0.009)	-0.076 (0.013)	0	0	0.248 (0.009)	–
76	Took Art and Design	Binary	–	-1.979 (0.041)	0	0	0	0	-0.081 (0.007)	0.184 (0.011)	0	0	0.185 (0.006)	–
77	Took History	Binary	–	-7.152 (0.028)	0	0	0	0	-0.294 (0.011)	1.051 (0.012)	0	0	0.439 (0.007)	–
78	Took Geography	Binary	–	-5.029 (0.040)	0	0	0	0	0.190 (0.009)	0.068 (0.014)	0	0	0.541 (0.009)	–
79	Took French	Binary	–	-7.410 (0.041)	0	0	0	0	0.066 (0.008)	0.633 (0.008)	0	0	0.561 (0.009)	–
80	Took German	Binary	–	-7.871 (0.032)	0	0	0	0	0.268 (0.010)	0.438 (0.012)	0	0	0.359 (0.008)	–
81	Took Business Studies	Binary	–	-3.634 (0.041)	0	0	0	0	0.284 (0.008)	-0.139 (0.011)	0	0	0.130 (0.007)	–
82	Took Physical Education	Binary	–	-2.436 (0.047)	0	0	0	0	0.326 (0.011)	-0.326 (0.015)	0	0	0.241 (0.008)	–
83	Took Drama	Binary	–	-3.519 (0.045)	0	0	0	0	-0.355 (0.011)	0.609 (0.013)	0	0	0.147 (0.010)	–
84	Took Information Technology	Binary	–	-3.940 (0.050)	0	0	0	0	0.141 (0.006)	-0.013 (0.011)	0	0	0.259 (0.009)	–

(*Continued on next page*)

Note: TA denotes teacher assessment. Any values of 0, 1, or -1 denote normalizations. Standard errors in parentheses.

Table 17: Measurement System and Normalizations: Female

No.	Description	Data Type	No. of Inter-vals	Intercept	Loadings KS1 Math	KS1 Verbal	KS2 Math	KS2 Verbal	KS3 Math	KS3 Verbal	KS4 Math	KS4 Verbal	KS4 Motive	Percent Noise
85	Took Short Information Technology	Binary	–	-3.918 (0.059)	0	0	0	0	-0.056 (0.010)	0.246 (0.014)	0	0	0.258 (0.011)	–
86	Took Spanish	Binary	–	-7.487 (0.047)	0	0	0	0	0.110 (0.007)	0.520 (0.014)	0	0	0.297 (0.006)	–
87	Took Music	Binary	–	-7.350 (0.039)	0	0	0	0	-0.010 (0.007)	0.629 (0.009)	0	0	0.269 (0.006)	–
88	Took Social Science	Binary	–	-2.545 (0.055)	0	0	0	0	-0.089 (0.012)	0.052 (0.016)	0	0	0.173 (0.010)	–
89	Took Design and Technology: Electronic Products	Binary	–	-7.998 (0.012)	0	0	0	0	0.536 (0.002)	-0.365 (0.003)	0	0	0.240 (0.001)	–
90	Took Design and Technology: System and Control	Binary	–	-10.137 (0.012)	0	0	0	0	0.645 (0.002)	-0.067 (0.003)	0	0	0.079 (0.001)	–
91	Took English Literature	Binary	–	-7.988 (0.054)	0	0	0	0	0.048 (0.009)	1.306 (0.012)	0	0	0.631 (0.007)	–
92	Took Design and Technology: Food Technology	Binary	–	-0.665 (0.050)	0	0	0	0	-0.086 (0.010)	-0.195 (0.016)	0	0	0.205 (0.007)	–
93	Took Statistics	Binary	–	-6.631 (0.059)	0	0	0	0	0.899 (0.013)	-0.308 (0.013)	0	0	0.149 (0.010)	–
94	Took Medial, Film and Television Studies	Binary	–	-2.235 (0.041)	0	0	0	0	-0.178 (0.012)	0.226 (0.013)	0	0	-0.039 (0.009)	–
95	Took Fine Art	Binary	–	-3.977 (0.042)	0	0	0	0	-0.094 (0.007)	0.239 (0.011)	0	0	0.205 (0.006)	–
96	Took Office Technology	Binary	–	-2.239 (0.041)	0	0	0	0	0.042 (0.011)	-0.213 (0.013)	0	0	0.112 (0.007)	–
97	Took Home Economics: Child Development	Binary	–	1.050 (0.043)	0	0	0	0	-0.095 (0.007)	-0.407 (0.014)	0	0	-0.073 (0.009)	–
98	Took Italian	Binary	–	-11.941 (0.016)	0	0	0	0	-0.034 (0.004)	0.816 (0.005)	0	0	0.435 (0.002)	–
99	Took Urdu	Binary	–	-5.342 (0.012)	0	0	0	0	-0.354 (0.005)	-0.488 (0.005)	0	0	1.175 (0.006)	–
100	Took Leisure and Tourism	Binary	–	-0.059 (0.035)	0	0	0	0	-0.037 (0.007)	-0.441 (0.008)	0	0	-0.153 (0.006)	–
101	Took Applied ICT	Binary	–	-3.969 (0.044)	0	0	0	0	0.158 (0.007)	-0.189 (0.007)	0	0	0.199 (0.008)	–
102	Took Health and Social Care	Binary	–	1.553 (0.059)	0	0	0	0	-0.124 (0.013)	-0.407 (0.017)	0	0	-0.162 (0.011)	–
103	Took Applied Business	Binary	–	-2.665 (0.029)	0	0	0	0	0.152 (0.006)	-0.300 (0.007)	0	0	0.004 (0.006)	–
104	Missing absence information	Binary	–	-2.018 (0.011)	0	0	0	0	0	0	0	0	-0.842 (0.002)	–
105	Any authorize absences	Binary	–	4.732 (0.025)	0	0	0	0	0	0	0	0	-0.367 (0.006)	–
106	Any unauthorized absences	Binary	–	5.072 (0.030)	0	0	0	0	0	0	0	0	-1.239 (0.007)	–
107	log authorize absences	Continuous	–	-0.860 (0.013)	0	0	0	0	0	0	0	0	-0.474 (0.004)	0.868 (0.001)
108	log unauthorized absences	Continuous	–	0	0	0	0	0	0	0	0	0	-1.000 (0.000)	0.653 (0.001)

Note: TA denotes teacher assessment. Any values of 0, 1, or -1 denote normalizations. Standard errors in parentheses.

Table 18: Measurement System and Normalizations: Male

No.	Description	Data Type	No. of Intervals	Intercept	Loadings: KS1 Math	KS1 Verbal	KS2 Math	KS2 Verbal	KS3 Math	KS3 Verbal	KS4 Math	KS4 Verbal	KS4 Motive	Percent Noise
1	Math Test	Intervaled	6	0	1	0	0	0	0	0	0	0	0	0.163 (0.001)
2	Math Using and Applying TA	Intervaled	4	0.374 (0.002)	0.886 (0.001)	0	0	0	0	0	0	0	0	0.096 (0.001)
3	Math Number and Algebra TA	Intervaled	4	0.521 (0.002)	0.869 (0.001)	0	0	0	0	0	0	0	0	0.074 (0.001)
4	Math Shapes and Measure TA	Intervaled	4	0.524 (0.002)	0.841 (0.001)	0	0	0	0	0	0	0	0	0.101 (0.001)
5	Writing Test	Intervaled	6	0	0	1	0	0	0	0	0	0	0	0.138 (0.001)
6	Writing TA	Intervaled	4	0.030 (0.002)	0	1.139 (0.001)	0	0	0	0	0	0	0	0.069 (0.001)
7	Reading TA	Intervaled	4	0.105 (0.005)	0	1.226 (0.002)	0	0	0	0	0	0	0	0.166 (0.001)
8	Listening TA	Intervaled	4	0.667 (0.003)	0	0.940 (0.001)	0	0	0	0	0	0	0	0.322 (0.002)
9	Math Test Paper A	Continuous	–	-26.236 (0.053)	0	0	11.316 (0.011)	0	0	0	0	0	0	0.116 (0.000)
10	Math Test Paper B	Continuous	–	-24.872 (0.061)	0	0	11.194 (0.013)	0	0	0	0	0	0	0.149 (0.001)
11	Math Arithmetic Test	Continuous	–	-16.373 (0.036)	0	0	6.541 (0.007)	0	0	0	0	0	0	0.174 (0.001)
12	Math TA	Intervaled	6	0	0	0	1	0	0	0	0	0	0	0.154 (0.004)
13	Reading Test	Continuous	–	-27.226 (0.074)	0	0	0	12.450 (0.016)	0	0	0	0	0	0.225 (0.001)
14	Writing Test	Continuous	–	-1.272 (0.078)	0	0	0	5.258 (0.017)	0	0	0	0	0	0.413 (0.002)
15	Spelling Test	Continuous	–	-7.116 (0.028)	0	0	0	3.054 (0.006)	0	0	0	0	0	0.392 (0.002)
16	English TA	Intervaled	6	0	0	0	0	1	0	0	0	0	0	0.139 (0.002)
17	Math Test Paper 1	Continuous	–	0.467 (0.004)	0	0	0	0	0.310 (0.001)	0	0	0	0	0.436 (0.001)
18	Math Test Paper 2	Continuous	–	0.161 (0.004)	0	0	0	0	0.348 (0.001)	0	0	0	0	0.399 (0.002)
19	Math Arithmetic Test	Continuous	–	-0.796 (0.003)	0	0	0	0	0.367 (0.001)	0	0	0	0	0.398 (0.001)
20	Math TA	Intervaled	7	0	0	0	0	0	1	0	0	0	0	0.105 (0.001)
21	Writing Test (Longer)	Continuous	–	-18.504 (0.072)	0	0	0	0	0	5.722 (0.013)	0	0	0	0.313 (0.002)
22	Reading Test	Continuous	–	-18.701 (0.044)	0	0	0	0	0	6.094 (0.008)	0	0	0	0.253 (0.001)
23	Writing Test (Shorter)	Continuous	–	-12.412 (0.044)	0	0	0	0	0	3.937 (0.007)	0	0	0	0.308 (0.001)
24	Reading Test (Shakespeare)	Continuous	–	-11.948 (0.045)	0	0	0	0	0	3.274 (0.008)	0	0	0	0.410 (0.002)
25	English TA	Intervaled	7	0	0	0	0	0	0	1	0	0	0	0.236 (0.001)
26	Math	Intervaled	8	0	0	0	0	0	0	0	1	0	0	0.097 (0.001)
27	English	Intervaled	8	0	0	0	0	0	0	0	0	1	0	0.106 (0.000)
28	Design and Technology: Graphic Products	Intervaled	8	-0.233 (0.052)	0	0	0	0	0	0	0.397 (0.013)	0.565 (0.014)	0	0.456 (0.005)

(*Continued on next page*)

Note: TA denotes teacher assessment. Values of 0, 1, or -1 denote normalizations. Standard errors in parentheses.

Table 18: Measurement System and Normalizations: Male

No.	Description	Data Type	No. of Intervals	Intercept	Loadings KS1 Math	KS1 Verbal	KS2 Math	KS2 Verbal	KS3 Math	KS3 Verbal	KS4 Math	KS4 Verbal	KS4 Motive	Percent Noise
29	Design and Technology: Resistant Materials Technology	Intervaled	8	1.277 (0.023)	0	0	0	0	0	0	0.362 (0.006)	0.455 (0.008)	0	0.478 (0.003)
30	Design and Technology: Textiles Technology	Intervaled	8	0.386 (0.233)	0	0	0	0	0	0	0.204 (0.069)	0.700 (0.087)	0	0.492 (0.022)
31	Art and Design	Intervaled	8	2.348 (0.032)	0	0	0	0	0	0	0.116 (0.013)	0.573 (0.014)	0	0.574 (0.004)
32	History	Intervaled	8	-3.480 (0.019)	0	0	0	0	0	0	0	1.400 (0.002)	0	0.174 (0.001)
33	Geography	Intervaled	8	-1.962 (0.023)	0	0	0	0	0	0	0.428 (0.007)	0.794 (0.008)	0	0.176 (0.001)
34	French	Intervaled	8	-1.354 (0.023)	0	0	0	0	0	0	0.362 (0.006)	0.701 (0.006)	0	0.256 (0.002)
35	German	Intervaled	8	-1.056 (0.032)	0	0	0	0	0	0	0.366 (0.011)	0.671 (0.012)	0	0.258 (0.004)
36	Business Studies	Intervaled	8	-1.860 (0.051)	0	0	0	0	0	0	0.352 (0.013)	0.841 (0.016)	0	0.236 (0.003)
37	Religious Studies	Intervaled	8	-2.690 (0.043)	0	0	0	0	0	0	0	1.319 (0.005)	0	0.222 (0.002)
38	Short Religious Studies	Intervaled	8	-2.527 (0.024)	0	0	0	0	0	0	0	1.250 (0.004)	0	0.299 (0.002)
39	Physical Education	Intervaled	8	1.253 (0.024)	0	0	0	0	0	0	0.349 (0.007)	0.527 (0.007)	0	0.342 (0.002)
40	Physics	Intervaled	8	-0.951 (0.052)	0	0	0	0	0	0	1.064 (0.005)	0	0	0.095 (0.002)
41	Chemistry	Intervaled	8	-0.862 (0.057)	0	0	0	0	0	0	1.041 (0.006)	0	0	0.102 (0.002)
42	Biology	Intervaled	8	-0.471 (0.046)	0	0	0	0	0	0	0.786 (0.012)	0.224 (0.014)	0	0.120 (0.003)
43	Drama	Intervaled	8	1.308 (0.050)	0	0	0	0	0	0	-0.018 (0.014)	0.852 (0.014)	0	0.444 (0.004)
44	Information Technology	Intervaled	8	-0.944 (0.071)	0	0	0	0	0	0	0.474 (0.011)	0.592 (0.016)	0	0.368 (0.004)
45	Short Information Technology	Intervaled	8	-0.936 (0.061)	0	0	0	0	0	0	0.549 (0.017)	0.422 (0.021)	0	0.463 (0.005)
46	Spanish	Intervaled	8	-1.840 (0.076)	0	0	0	0	0	0	0.345 (0.017)	0.787 (0.017)	0	0.290 (0.006)
47	Music	Intervaled	8	-0.233 (0.044)	0	0	0	0	0	0	0.194 (0.015)	0.817 (0.017)	0	0.428 (0.004)
48	Social Science	Intervaled	8	-1.281 (0.047)	0	0	0	0	0	0	0.096 (0.012)	1.013 (0.015)	0	0.333 (0.005)
49	Design and Technology: Electronic Products	Intervaled	8	0.024 (0.098)	0	0	0	0	0	0	0.580 (0.019)	0.378 (0.023)	0	0.437 (0.009)
50	Design and Technology: System and Control	Intervaled	8	0.028 (0.121)	0	0	0	0	0	0	0.543 (0.026)	0.380 (0.033)	0	0.432 (0.012)
51	English Literature	Intervaled	8	-0.995 (0.012)	0	0	0	0	0	0	0	1.117 (0.002)	0	0.164 (0.001)
52	Design and Technology: Food Technology	Intervaled	8	0.254 (0.053)	0	0	0	0	0	0	0.252 (0.016)	0.684 (0.018)	0	0.385 (0.006)
53	Science	Intervaled	8	0.202 (0.032)	0	0	0	0	0	0	0.478 (0.008)	0.450 (0.006)	0	0.250 (0.003)
54	Statistics	Intervaled	8	-0.666 (0.042)	0	0	0	0	0	0	0.978 (0.012)	0.026 (0.014)	0	0.175 (0.004)
55	Medial, Film and Television Studies	Intervaled	8	-0.641 (0.062)	0	0	0	0	0	0	-0.023 (0.013)	1.100 (0.016)	0	0.285 (0.005)
56	Fine Art	Intervaled	8	2.179 (0.040)	0	0	0	0	0	0	0.071 (0.022)	0.645 (0.023)	0	0.530 (0.006)
57	Office Technology	Intervaled	8	-0.403 (0.054)	0	0	0	0	0	0	0.486 (0.018)	0.553 (0.020)	0	0.300 (0.006)

(*Continued on next page*)

Note: TA denotes teacher assessment. Values of 0, 1, or -1 denote normalizations. Standard errors in parentheses.

Table 18: Measurement System and Normalizations: Male

No.	Description	Data Type	No. of Intervals	Intercept	Loadings: KS1 Math	KS1 Verbal	KS2 Math	KS2 Verbal	KS3 Math	KS3 Verbal	KS4 Math	KS4 Verbal	KS4 Motive	Percent Noise
58	Home Economics: Child Development	Intervaled	8	-0.560 (0.424)	0	0	0	0	0	0	0.145 (0.117)	0.817 (0.144)	0	0.445 (0.049)
59	Italian	Intervaled	8	0.380 (0.487)	0	0	0	0	0	0	0.190 (0.095)	0.756 (0.112)	0	0.466 (0.047)
60	Urdu	Intervaled	8	2.030 (0.222)	0	0	0	0	0	0	0.207 (0.072)	0.507 (0.089)	0	0.619 (0.019)
61	Additional Applied Science	Intervaled	8	0.294 (0.179)	0	0	0	0	0	0	0.473 (0.042)	0.462 (0.059)	0	0.211 (0.010)
62	Leisure and Tourism	Intervaled	8	-1.786 (0.079)	0	0	0	0	0	0	0.166 (0.022)	0.956 (0.025)	0	0.366 (0.009)
63	Applied ICT	Intervaled	8	-1.102 (0.100)	0	0	0	0	0	0	0.389 (0.020)	0.631 (0.022)	0	0.433 (0.008)
64	Applied Science	Intervaled	8	0.661 (0.051)	0	0	0	0	0	0	0.338 (0.013)	0.543 (0.015)	0	0.330 (0.008)
65	Health and Social Care	Intervaled	8	-0.905 (0.307)	0	0	0	0	0	0	0.093 (0.068)	0.927 (0.083)	0	0.413 (0.029)
66	Applied Business	Intervaled	8	-1.864 (0.119)	0	0	0	0	0	0	0.476 (0.023)	0.711 (0.028)	0	0.330 (0.008)
67	Double Science	Intervaled	8	0.270 (0.016)	0	0	0	0	0	0	0.943 (0.002)	0	0	0.168 (0.001)
	Selection Equations													
68	Took any science course	Binary	–	-3.425 (0.039)	0	0	0	0	0.189 (0.008)	0.381 (0.010)	0	0	0.860 (0.007)	–
69	Took any college preperation science	Binary	–	-5.998 (0.042)	0	0	0	0	0.395 (0.009)	0.536 (0.012)	0	0	0.523 (0.011)	–
70	Took any advanced science	Binary	–	-14.794 (0.032)	0	0	0	0	0.974 (0.007)	0.558 (0.011)	0	0	0.526 (0.007)	–
71	Took any relgion course	Binary	–	-4.221 (0.047)	0	0	0	0	-0.090 (0.008)	0.398 (0.014)	0	0	0.721 (0.011)	–
72	Took Religious Studies	Binary	–	-5.957 (0.043)	0	0	0	0	-0.121 (0.011)	0.491 (0.016)	0	0	0.599 (0.012)	–
73	Took Design and Technology: Graphic Products	Binary	–	-3.378 (0.038)	0	0	0	0	-0.031 (0.009)	0.165 (0.007)	0	0	0.193 (0.009)	–
74	Took Design and Technology: Resistant Materials Technology	Binary	–	0.130 (0.050)	0	0	0	0	0.116 (0.007)	-0.544 (0.012)	0	0	0.224 (0.010)	–
75	Took Design and Technology: Textiles Technology	Binary	–	-3.991 (0.015)	0	0	0	0	-0.366 (0.001)	0.063 (0.003)	0	0	0.100 (0.002)	–
76	Took Art and Design	Binary	–	-1.110 (0.040)	0	0	0	0	-0.195 (0.007)	0.060 (0.012)	0	0	0.173 (0.010)	–
77	Took History	Binary	–	-6.496 (0.035)	0	0	0	0	-0.282 (0.009)	0.931 (0.011)	0	0	0.523 (0.010)	–
78	Took Geography	Binary	–	-5.117 (0.041)	0	0	0	0	0.120 (0.006)	0.145 (0.012)	0	0	0.631 (0.009)	–
79	Took French	Binary	–	-8.060 (0.033)	0	0	0	0	0.124 (0.008)	0.631 (0.012)	0	0	0.571 (0.011)	–
80	Took German	Binary	–	-8.974 (0.029)	0	0	0	0	0.305 (0.009)	0.515 (0.012)	0	0	0.440 (0.007)	–
81	Took Business Studies	Binary	–	-4.847 (0.038)	0	0	0	0	0.281 (0.007)	0.020 (0.010)	0	0	0.278 (0.011)	–
82	Took Physical Education	Binary	–	-1.600 (0.041)	0	0	0	0	0.195 (0.007)	-0.178 (0.011)	0	0	0.140 (0.009)	–
83	Took Drama	Binary	–	-3.581 (0.033)	0	0	0	0	-0.430 (0.006)	0.555 (0.009)	0	0	0.250 (0.009)	–
84	Took Information Technology	Binary	–	-4.635 (0.056)	0	0	0	0	0.151 (0.008)	0.051 (0.013)	0	0	0.365 (0.010)	–

(*Continued on next page*)

Note: TA denotes teacher assessment. Values of 0, 1, or -1 denote normalizations. Standard errors in parentheses.

Table 18: Measurement System and Normalizations: Male

					Loadings									
No.	Description	Data Type	No. of Intervals	Intercept	KS1 Math	KS1 Verbal	KS2 Math	KS2 Verbal	KS3 Math	KS3 Verbal	KS4 Math	KS4 Verbal	KS4 Motive	Percent Noise
85	Took Short Information Technology	Binary	–	-3.432 (0.047)	0	0	0	0	-0.040 (0.009)	0.163 (0.010)	0	0	0.203 (0.010)	–
86	Took Spanish	Binary	–	-8.238 (0.037)	0	0	0	0	0.179 (0.007)	0.494 (0.010)	0	0	0.331 (0.007)	–
87	Took Music	Binary	–	-6.334 (0.040)	0	0	0	0	-0.078 (0.007)	0.620 (0.009)	0	0	0.226 (0.006)	–
88	Took Social Science	Binary	–	-3.155 (0.053)	0	0	0	0	-0.100 (0.008)	0.095 (0.016)	0	0	0.261 (0.008)	–
89	Took Design and Technology: Electronic Products	Binary	–	-4.798 (0.031)	0	0	0	0	0.445 (0.007)	-0.391 (0.008)	0	0	0.265 (0.006)	–
90	Took Design and Technology: System and Control	Binary	–	-6.805 (0.033)	0	0	0	0	0.477 (0.004)	-0.269 (0.007)	0	0	0.394 (0.005)	–
91	Took English Literature	Binary	–	-8.123 (0.054)	0	0	0	0	0.063 (0.011)	1.269 (0.013)	0	0	0.644 (0.010)	–
92	Took Design and Technology: Food Technology	Binary	–	-2.719 (0.033)	0	0	0	0	-0.207 (0.010)	0.060 (0.010)	0	0	0.297 (0.006)	–
93	Took Statistics	Binary	–	-7.074 (0.042)	0	0	0	0	0.856 (0.011)	-0.220 (0.013)	0	0	0.187 (0.007)	–
94	Took Medial, Film and Television Studies	Binary	–	-2.310 (0.045)	0	0	0	0	-0.190 (0.006)	0.181 (0.009)	0	0	0.068 (0.010)	–
95	Took Fine Art	Binary	–	-3.328 (0.032)	0	0	0	0	-0.188 (0.007)	0.151 (0.011)	0	0	0.215 (0.008)	–
96	Took Office Technology	Binary	–	-2.740 (0.041)	0	0	0	0	0.012 (0.006)	-0.127 (0.011)	0	0	0.154 (0.007)	–
97	Took Home Economics: Child Development	Binary	–	-5.543 (0.004)	0	0	0	0	-0.206 (0.001)	0.046 (0.001)	0	0	0.029 (0.001)	–
98	Took Italian	Binary	–	-11.725 (0.011)	0	0	0	0	0.074 (0.002)	0.726 (0.002)	0	0	0.286 (0.002)	–
99	Took Urdu	Binary	–	-4.813 (0.011)	0	0	0	0	-0.166 (0.002)	-0.618 (0.004)	0	0	0.849 (0.004)	–
100	Took Leisure and Tourism	Binary	–	-0.805 (0.022)	0	0	0	0	-0.186 (0.006)	-0.321 (0.006)	0	0	-0.044 (0.005)	–
101	Took Applied ICT	Binary	–	-4.324 (0.045)	0	0	0	0	0.166 (0.006)	-0.101 (0.008)	0	0	0.227 (0.008)	–
102	Took Health and Social Care	Binary	–	-3.520 (0.010)	0	0	0	0	-0.474 (0.003)	0.061 (0.003)	0	0	0.158 (0.002)	–
103	Took Applied Business	Binary	–	-3.025 (0.028)	0	0	0	0	0.144 (0.006)	-0.196 (0.008)	0	0	0.021 (0.007)	–
104	Missing absence information	Binary	–	-0.955 (0.022)	0	0	0	0	0	0	0	0	-0.994 (0.005)	–
105	Any authorize absences	Binary	–	4.077 (0.033)	0	0	0	0	0	0	0	0	-0.302 (0.008)	–
106	Any unauthorized absences	Binary	–	5.204 (0.035)	0	0	0	0	0	0	0	0	-1.262 (0.008)	–
107	log authorize absences	Continuous	–	-0.937 (0.014)	0	0	0	0	0	0	0	0	-0.487 (0.003)	0.881 (0.002)
108	log unauthorized absences	Continuous	–	0	0	0	0	0	0	0	0	0	-1.000 (0.000)	0.680 (0.002)

Note: TA denotes teacher assessment. Values of 0, 1, or -1 denote normalizations. Standard errors in parentheses.

F Transition Matrix: Key Stage 1 to Key Stage 4

Table 19: Factor Transition Matrix KS1 to KS4

	KS4 0-25th PCTL	KS4 25-50th PCTL	KS4 50-75th PCTL	KS4 75-100th PCTL
Math				
KS1 0-25th PCTL	63%	26%	9%	2%
KS1 25-50th PCTL	26%	40%	27%	7%
KS1 50-75th PCTL	9%	22%	37%	32%
KS1 75-100th PCTL	3%	12%	27%	58%
Verbal				
KS1 0-25th PCTL	61%	27%	10%	2%
KS1 25-50th PCTL	29%	39%	26%	7%
KS1 50-75th PCTL	8%	23%	39%	30%
KS1 75-100th PCTL	3%	10%	26%	61%

Note: Rows sum to 100 percent. This table shows how students transition in the skills distribution between Key Stage 1 and Key Stage 4. For example, the first row shows that among the students that started in the bottom quartile of the math skills distribution in Key Stage 1, 63% remained in the same quartile in Key Stage 4, while only 2% were able to transition to the top quartile.

G Additional Nested CES Production Function Parameters

Table 20: Nested CES Production Function: TFP Coefficients

	Math			Verbal		
	KS2	KS3	KS4	KS2	KS3	KS4
Female						
Intercept	2.991 (0.012)	-0.200 (0.021)	-2.090 (0.024)	2.866 (0.011)	1.443 (0.018)	-1.401 (0.022)
Asian	0.039 (0.006)	0.097 (0.008)	0.083 (0.006)	0.030 (0.006)	0.061 (0.005)	0.053 (0.005)
Black	-0.039 (0.006)	0.011 (0.005)	0.011 (0.007)	-0.003 (0.005)	0.022 (0.005)	0.026 (0.006)
Race Other	0.008 (0.006)	0.009 (0.007)	0.010 (0.005)	0.028 (0.004)	0.026 (0.005)	0.030 (0.005)
Race Missing	-0.002 (0.007)	-0.022 (0.005)	-0.002 (0.005)	-0.003 (0.005)	-0.011 (0.004)	-0.002 (0.005)
Mother Tounge Not English	0.022 (0.008)	0.049 (0.007)	0.032 (0.007)	0.015 (0.005)	0.043 (0.004)	0.042 (0.005)
Free School Meal Eligible	-0.040 (0.003)	-0.096 (0.002)	-0.026 (0.002)	-0.051 (0.002)	-0.068 (0.002)	-0.041 (0.002)
Special Education Needs	-0.303 (0.004)	-0.067 (0.004)	0.023 (0.003)	-0.308 (0.003)	-0.024 (0.004)	-0.067 (0.003)
FSM/SEN Missing	-0.068 (0.010)	-0.112 (0.014)	-0.033 (0.015)	-0.074 (0.006)	-0.078 (0.013)	-0.081 (0.017)
KS4 Selection Factor	–	–	0.717 (0.003)	–	–	0.655 (0.003)
Male						
Intercept	3.010 (0.011)	-0.181 (0.018)	-2.143 (0.017)	2.739 (0.010)	1.329 (0.017)	-2.091 (0.018)
Asian	0.033 (0.008)	0.076 (0.007)	0.084 (0.007)	0.021 (0.007)	0.053 (0.006)	0.069 (0.005)
Black	-0.071 (0.006)	-0.043 (0.006)	-0.021 (0.008)	-0.033 (0.006)	-0.012 (0.006)	0.006 (0.006)
Race Other	-0.004 (0.005)	-0.005 (0.006)	0.007 (0.004)	0.022 (0.005)	0.019 (0.005)	0.019 (0.005)
Race Missing	0.005 (0.005)	-0.013 (0.006)	-0.005 (0.005)	0.002 (0.004)	-0.004 (0.005)	-0.001 (0.004)
Mother Tounge Not English	0.017 (0.009)	0.039 (0.006)	0.032 (0.008)	-0.002 (0.007)	0.028 (0.006)	0.022 (0.006)
Free School Meal Eligible	-0.035 (0.003)	-0.086 (0.002)	-0.028 (0.002)	-0.050 (0.003)	-0.068 (0.002)	-0.040 (0.003)
Special Education Needs	-0.254 (0.003)	-0.076 (0.002)	0.030 (0.002)	-0.285 (0.003)	-0.055 (0.002)	-0.040 (0.002)
FSM/SEN Missing	-0.087 (0.007)	-0.146 (0.013)	0.018 (0.015)	-0.095 (0.006)	-0.124 (0.015)	-0.028 (0.015)
KS4 Selection Factor	–	–	0.747 (0.003)	–	–	0.752 (0.003)

Note: These coefficients correspond to the TFP ($A_i t^k$) parameters described in eq. (1), which is our main nested CES specification. Neighborhood fixed effects are not reported and unobserved heterogeneity parameters are reported in appendix G.

Table 21: Nested CES Production Function: Unobserved Heterogeneity

		Type Share	Math				Verbal			
			KS1	KS2	KS3	KS4	KS1	KS2	KS3	KS4
Female										
	Type 1	0.118	0.728	0.258	0.258	0.258	0.728	0.176	0.176	0.176
	Type 2	0.249	-0.189	-0.386	-0.386	-0.386	-0.189	-0.327	-0.327	-0.327
	Type 3	0.156	0.594	-0.138	-0.138	-0.138	0.594	-0.121	-0.121	-0.121
	Type 4	0.076	-0.966	-0.251	-0.251	-0.251	-0.966	-0.183	-0.183	-0.183
	Type 5	0.400	0	0	0	0	0	0	0	0
Male										
	Type 1	0.109	0.694	0.282	0.282	0.282	0.694	0.243	0.243	0.243
	Type 2	0.236	-0.162	-0.391	-0.391	-0.391	-0.162	-0.351	-0.351	-0.351
	Type 3	0.144	0.586	-0.139	-0.139	-0.139	0.586	-0.093	-0.093	-0.093
	Type 4	0.112	-0.914	-0.210	-0.210	-0.210	-0.914	-0.144	-0.144	-0.144
	Type 5	0.400	0	0	0	0	0	0	0	0

Note: This table reports the unobserved heterogeneity parameters corresponding to the TFP of the production function described in eq. (1), which is our main nested CES specification (see section 3.1). Type share denotes the probability of each of the five types.

H CES Production Functions Estimates: Robustness Checks

Table 22: CES Production Function: 3-Input Nested-CES(m,v,Q) and 2-Input CES(m,v),Female

	Math			Verbal		
	KS2	KS3	KS4	KS2	KS3	KS4
Panel A: 3-Input Nested-CES(m,v,Q), No Covariates in TFP, No Unobserved Het.						
Math Coefficient (α)	0.860	0.828	0.791	0.113	0.000	0.127
	(0.007)	(0.002)	(0.002)	(0.007)	(0.000)	(0.003)
Verbal Coefficient ($1-\alpha$)	0.140	0.172	0.209	0.887	1.000	0.873
	(0.007)	(0.002)	(0.002)	(0.007)	(0.000)	(0.003)
Complementarity Parmaeter Math/verbal (γ)	-3.209	-0.200	0.722	-2.356	–	0.620
	(0.133)	(0.040)	(0.027)	(0.354)		(0.031)
Math/Verbal Aggregator Coefficeint (δ)	0.926	0.905	0.765	0.954	0.849	0.930
	(0.002)	(0.002)	(0.006)	(0.003)	(0.003)	(0.005)
School Coefficient($1-\delta$)	0.074	0.095	0.235	0.046	0.151	0.070
	(0.002)	(0.002)	(0.006)	(0.003)	(0.003)	(0.005)
Complementarity Parameter Skill/School (ρ)	-0.290	0.406	0.486	0.051	0.615	0.116
	(0.050)	(0.019)	(0.013)	(0.096)	(0.019)	(0.027)
Return to Scale (r)	1.064	1.728	1.597	1.159	1.222	1.563
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
Variance of shocks ν (ς^2)	0.166	0.131	0.349	0.118	0.105	0.321
	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.003)
Panel B: 3-Input Nested-CES(m,v,Q), With Covariates in TFP, No Unobserved Het.						
Math Coefficient (α)	0.939	0.850	0.847	0.086	0.000	0.039
	(0.006)	(0.002)	(0.002)	(0.004)	(0.000)	(0.001)
Verbal Coefficient ($1-\alpha$)	0.061	0.150	0.153	0.914	1.000	0.961
	(0.006)	(0.002)	(0.002)	(0.004)	(0.000)	(0.001)
Complementarity Parmaeter Math/verbal (γ)	-4.194	-0.200	0.171	1.000	–	-0.544
	(0.158)	(0.040)	(0.019)	(0.087)		(0.048)
Math/Verbal Aggregator Coefficeint (δ)	0.935	0.903	0.832	0.975	0.847	0.977
	(0.003)	(0.002)	(0.004)	(0.002)	(0.003)	(0.004)
School Coefficient($1-\delta$)	0.065	0.097	0.168	0.025	0.153	0.023
	(0.003)	(0.002)	(0.004)	(0.002)	(0.003)	(0.004)
Complementarity Parameter Skill/School (ρ)	0.135	0.474	0.547	0.747	0.652	0.141
	(0.077)	(0.016)	(0.013)	(0.060)	(0.019)	(0.050)
Return to Scale (r)	0.925	1.684	1.243	1.006	1.191	1.219
	(0.004)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
Variance of shocks ν (ς^2)	0.134	0.112	0.124	0.093	0.091	0.129
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Panel C: 2-Input CES(m,v), With Covariates in TFP, With Unobserved Het.						
Math Coefficient (α)	0.974	0.930	0.928	0.022	0.000	0.002
	(0.002)	(0.003)	(0.003)	(0.002)	(0.000)	(0.001)
Verbal Coefficient ($1-\alpha$)	0.026	0.070	0.072	0.978	1.000	0.998
	(0.002)	(0.003)	(0.003)	(0.002)	(0.000)	(0.001)
Complementarity Parmaeter Math/verbal (γ)	-5.690	-0.211	0.294	1.000	–	–
	(0.159)	(0.167)	(0.038)	(0.000)		
Return to Scale (r)	0.676	1.366	1.012	0.809	0.929	1.062
	(0.005)	(0.004)	(0.003)	(0.005)	(0.003)	(0.004)
Variance of shocks ν (ς^2)	0.099	0.080	0.103	0.066	0.069	0.117
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)

Note: Panels A and B correspond to three-input nested CES production functions (i.e. math and verbal skills, and school quality), as described in Section 3.1. However, the TFP term of the model in Panel A does not include covariates or unobserved heterogeneity, while in Panel B it only includes background covariates (i.e. race, mother tongue, free school meal eligibility, special education needs, and neighborhood fixed effects). Panel C corresponds to the following production function, $\Theta^k_{i,t} = A^k_{it}[\delta^k_t \left(\Theta^m_{i,t-1}\right)^{\rho^k_t} + (1-\delta^k_t)\left(\Theta^v_{i,t-1}\right)^{\rho^k_t}]^{r^k_t/\rho^k_t}$, where A^k_{it} includes background covariates, unobserved heterogeneity, and school fixed effects. The main difference between this two-input CES production function and the nested CES in Section 3.1 is how we account for school effects. While the two-input CES model specifies school quality as fixed effects in the TFP, the nested CES accounts for school quality as an additional input in the production function

Table 23: CES Production Function: 3-Input Nested-CES(m,v,Q) and 2-Input CES(m,v), Male

	Math			Verbal		
	KS2	KS3	KS4	KS2	KS3	KS4
Panel A: 3-Input Nested-CES(m,v,Q), No Covariates in TFP, No Unobserved Het.						
Math Coefficient (α)	0.972	0.845	0.747	0.175	0.000	0.078
	(0.001)	(0.002)	(0.003)	(0.005)	(0.000)	(0.002)
Verbal Coefficient ($1-\alpha$)	0.028	0.155	0.253	0.825	1.000	0.922
	(0.001)	(0.002)	(0.003)	(0.005)	(0.000)	(0.002)
Complementarity Parmaeter Math/verbal (γ)	-4.528	-0.125	0.561	-0.367	–	0.656
	(0.070)	(0.043)	(0.013)	(0.057)		(0.019)
Math/Verbal Aggregator Coefficeint (δ)	0.929	0.896	0.821	0.966	0.844	0.954
	(0.002)	(0.002)	(0.005)	(0.002)	(0.002)	(0.003)
School Coefficient($1-\delta$)	0.071	0.104	0.179	0.034	0.156	0.046
	(0.002)	(0.002)	(0.005)	(0.002)	(0.002)	(0.003)
Complementarity Parameter Skill/School (ρ)	-0.171	0.502	0.379	0.460	0.748	-0.041
	(0.047)	(0.012)	(0.011)	(0.044)	(0.013)	(0.020)
Return to Scale (r)	1.031	1.706	1.555	1.197	1.222	1.614
	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
Variance of shocks ν (ς^2)	0.182	0.138	0.374	0.140	0.127	0.382
	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.003)
Panel B: 3-Input Nested-CES(m,v,Q), With Covariates in TFP, No Unobserved Het.						
Math Coefficient (α)	0.988	0.865	0.882	0.159	0.000	0.040
	(0.001)	(0.003)	(0.002)	(0.004)	(0.000)	(0.002)
Verbal Coefficient ($1-\alpha$)	0.012	0.135	0.118	0.841	1.000	0.960
	(0.001)	(0.003)	(0.002)	(0.004)	(0.000)	(0.002)
Complementarity Parmaeter Math/verbal (γ)	–	-0.113	0.074	-0.026	–	-0.230
		(0.051)	(0.022)	(0.045)		(0.054)
Math/Verbal Aggregator Coefficeint (δ)	0.930	0.887	0.816	0.975	0.830	0.912
	(0.003)	(0.002)	(0.005)	(0.001)	(0.002)	(0.005)
School Coefficient($1-\delta$)	0.070	0.113	0.184	0.025	0.170	0.088
	(0.003)	(0.002)	(0.005)	(0.001)	(0.002)	(0.005)
Complementarity Parameter Skill/School (ρ)	0.053	0.587	0.514	0.787	0.796	0.555
	(0.057)	(0.013)	(0.008)	(0.030)	(0.015)	(0.025)
Return to Scale (r)	0.895	1.663	1.218	1.032	1.188	1.251
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)
Variance of shocks ν (ς^2)	0.148	0.119	0.153	0.111	0.111	0.159
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Panel C: 2-Input CES(m,v), With Covariates in TFP, With Unobserved Het.						
Math Coefficient (α)	0.990	0.945	0.978	0.063	0.000	0.001
	(0.001)	(0.007)	(0.003)	(0.005)	(0.000)	(0.000)
Verbal Coefficient ($1-\alpha$)	0.010	0.055	0.022	0.937	1.000	0.999
	(0.001)	(0.007)	(0.003)	(0.005)	(0.000)	(0.000)
Complementarity Parmaeter Math/verbal (γ)	–	0.585	0.403	0.344	–	–
		(0.848)	(0.086)	(0.085)		
Return to Scale (r)	0.665	1.352	0.977	0.820	0.906	1.067
	(0.004)	(0.004)	(0.002)	(0.005)	(0.003)	(0.002)
Variance of shocks ν (ς^2)	0.114	0.083	0.126	0.080	0.082	0.141
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)

Note: Panels A and B correspond to three-input nested CES production functions (i.e. math and verbal skills, and school quality), as described in Section 3.1. However, the TFP term of the model in Panel A does not include covariates or unobserved heterogeneity, while in Panel B it only includes background covariates (i.e. race, mother tongue, free school meal eligibility, special education needs, and neighborhood fixed effects). Panel C corresponds to the following production function, $\Theta_{i,t}^k = A_{it}^k[\delta_t^k \left(\Theta_{i,t-1}^m\right)^{\rho_t^k} + (1-\delta_t^k)\left(\Theta_{i,t-1}^v\right)^{\rho_t^k}]^{r_t^k/\rho_t^k}$, where A_{it}^k includes background covariates, unobserved heterogeneity, and school fixed effects. The main difference between this two-input CES production function and the nested CES in Section 3.1 is how we account for school effects. While the two-input CES model specifies school quality as fixed effects in the TFP, the nested CES accounts for school quality as an additional input in the production function

I The Effect of Math Skills on Unconditional STEM Enrollment

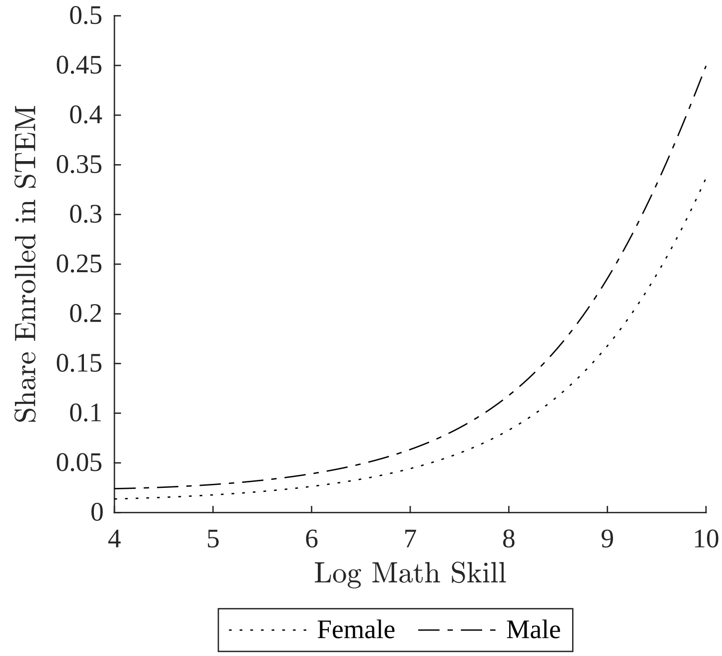


Figure 4: Unconditional Share Enrolled in STEM for Assigned KS4 Math Skills

J Endogeneity Concerns

Role of Motivation Table 24 repeats the analysis in Table 9 but using skills from earlier Key Stages. Table 24 shows that verbal skills uniformly have a larger effect on university enrollment than math skills at each stage of the schooling career and even after conditioning on different subsamples (i.e. white students, white students whose mother speaks English). Moreover, the magnitude of the differential effect is sizable across the board, which further substantiates our main findings. For example, the top panel of Table 24 indicates that among females the effect of KS1 verbal skills on university enrollment is three times larger than KS1 math skills.

Decomposition of KS4 Skills Since we have skill measures at each time period, we can statistically decompose the skills observed in KS4 as a function of the earlier skills. Let $\hat{\theta}_{i4|3,2,1} = \mathrm{E}(\theta_{i4}|\theta_{i3}, \theta_{i2}, \theta_{i1})$ denote the expected value of skills in KS4, conditional on everything that has happened to the student as of KS3. We can re-write θ_{i4} as:

$$\theta_{i4} = \underbrace{\left(\theta_{i4} - \hat{\theta}_{i4|3,2,1}\right)}_{\text{residual change in skill occurring in KS4}} + \hat{\theta}_{i4|3,2,1}$$

Writing the variable in this way is useful because it informs us about the portion of the skill that was determined in the preceding periods and the portion that was determined in the current period. In fact, we can further decompose this variable period-by-period

$$\theta_{i4} = \underbrace{\left(\theta_{i4} - \hat{\theta}_{i4|3,2,1}\right)}_{\text{KS4 Residual}} + \underbrace{\left(\hat{\theta}_{i4|3,2,1} - \hat{\theta}_{i4|2,1}\right)}_{\text{KS3 Residual}} + \underbrace{\left(\hat{\theta}_{i4|2,1} - \hat{\theta}_{i4|1}\right)}_{\text{KS2 Residual}} + \hat{\theta}_{i4|1} \tag{16}$$

where $\hat{\theta}_{i4|2,1} = \mathrm{E}(\theta_{i4}|\theta_{i2}, \theta_{i1})$, and $\hat{\theta}_{i4|1} = \mathrm{E}(\theta_{i4}|\theta_{i1})$

While the results in Table 9 show the total effect of skills when each of the four components in Eq. (16) are combined, an alternative approach would be to include each of these differences in a regression, which will allow us to recover the marginal effect of the new information received at each Key Stage. The benefit of this approach is that it is possible to study the relative contribution to university enrollment of math and verbal skills after conditioning on prior skills.

Table 25 shows the results of logistic regressions on college enrollment that includes the decomposition of KS4 math and verbal skills by gender as described in Eq. (16). This regression allows us to consider the following counterfactual: does performing better than expected in math, once we condition on earlier skills, have a larger impact on college enrollment than performing better than expected in verbal? These specifications also control for background characteristics and unobserved heterogeneity obtained from the production function estimation. The results in table 25 show that, at each Key Stage, performing better than expected in verbal has a much larger impact on college enrollment than performing better than expected in math. For example, a one log-unit increase in the predicted value of KS4 verbal skills occurring from an outcome in KS2, after conditioning on family background characteristics and KS1 information, leads to an increase in female (male) college enrollment of 15.6 (13) percentage points, while for math would lead to an increase of 5.5 (6.4) percentage points. These results strongly suggest that the conclusions from our earlier analysis are not driven by endogeneity in parental inputs that benefit one skill over the other.

Table 24: Logistic Regression: University Enrollment, Early Factors and Sub Samples

	Females				Males			
	KS1	KS2	KS3	KS4	KS1	KS2	KS3	KS4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Average Marginal Effect of 1 log unit increase in Skills							
	Full Sample							
Math	0.047	0.035	0.073	0.054	0.043	0.046	0.075	0.058
	(0.004)	(0.003)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Verbal	0.153	0.206	0.174	0.147	0.132	0.186	0.156	0.126
	(0.002)	(0.002)	(0.001)	(0.001)	(0.003)	(0.002)	(0.001)	(0.001)
Obs.	248479				250257			
	Mother Tongue English Only							
Math	0.045	0.033	0.073	0.054	0.041	0.046	0.075	0.057
	(0.004)	(0.003)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Verbal	0.162	0.213	0.178	0.150	0.140	0.192	0.157	0.127
	(0.003)	(0.002)	(0.001)	(0.001)	(0.004)	(0.002)	(0.001)	(0.001)
Obs.	235016				237109			
	White and Mother Tongue English Only							
Math	0.046	0.032	0.074	0.054	0.041	0.046	0.076	0.057
	(0.004)	(0.003)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Verbal	0.168	0.218	0.180	0.151	0.147	0.198	0.160	0.128
	(0.004)	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)	(0.002)	(0.001)
Obs.	214263				216057			

Note: Results correspond to logistic regressions where we analyze the average marginal effect of math and verbal skills at each Key Stage. The top panel includes the whole sample where the dependent variable is college enrollment, and the independent variables correspond to KS skills at different stages of the schooling career and baseline controls. The middle panel constrains the sample to students whose mother's native tongue is English, while the bottom panel further constrains this sample by only considering white students. All regressions include as the dependent variable university enrollment. Baseline controls include race, mother tongue, Key Stage 4 school quality, free school lunch, special education needs, neighborhood characteristics, and higher order factor terms (i.e. squared terms of skills and their interactions) and the unobserved persistent TFP shock. Bootstrapped standard errors at the school level.

Table 25: Logistic Regression: Effect of Residual Skills on University Enrollment

	Female (1)	Male (2)
Average Marginal Effect of 1 log unit increase in Skills		
$\theta_{i4}^{math} - \hat{\theta}_{4\|3,2,1}^{math}$	0.065	0.060
	(0.003)	(0.003)
$\hat{\theta}_{4\|3,2,1}^{math} - \hat{\theta}_{4\|2,1}^{math}$	0.056	0.063
	(0.002)	(0.002)
$\hat{\theta}_{4\|2,1}^{math} - \hat{\theta}_{4\|1}^{math}$	0.050	0.059
	(0.002)	(0.001)
$\hat{\theta}_{4\|1}^{math}$	0.046	0.043
	(0.002)	(0.001)
$\theta_{i4}^{verb} - \hat{\theta}_{4\|3,2,1}^{verb}$	0.146	0.127
	(0.003)	(0.002)
$\hat{\theta}_{4\|3,2,1}^{verb} - \hat{\theta}_{4\|2,1}^{verb}$	0.161	0.135
	(0.002)	(0.002)
$\hat{\theta}_{4\|2,1}^{verb} - \hat{\theta}_{4\|1}^{verb}$	0.153	0.126
	(0.002)	(0.001)
$\hat{\theta}_{4\|1}^{verb}$	0.145	0.131
	(0.002)	(0.001)

Note: Logistic regressions by gender. Dependent variable: college enrollment. Independent variables correspond to residuals from the KS4 skills decomposition. Specifications include controls for family background characteristics and unobserved heterogeneity obtained from the production function estimation.

Interrelation between Test Scores and Externalizing Behavior, Family Background Characteristics, and IQ It is possible that our verbal factor is capturing other types of skills that affect schooling outcomes, which are not present in the math factor. We further investigate this issue using a database that contains richer measures of externalizing behavior. We use the Avon Longitudinal Study of Parents and Children (ALSPAC) database which is a large scale longitudinal study of children born in Avon (United Kingdom) during the early 1990s. Although these data cannot be linked to one of our main databases (i.e. HESA), it is useful for further analysis because it has very rich information on student background characteristics, and the individuals in the sample, which also attend the UK educational system, are similar in age to students in our main database. This data contains proxies for externalizing behavior obtained from the Strengths and Difficulties Questionnaire (SDQ), which was completed by the student's teacher at age 7.[96] We have measures for emotional problems, conduct problems, hyperactivity/inattention, and peer relationship problems. Higher scores (scale of 0 to 10) indicate greater levels of severity. In addition, we have a measure for pro-social behavior that takes values from 0 to 10, where a higher value denotes more pro-social behavior.

This database also contains aggregate scores on math and English exams for each Key Stage.[97] To study if externalizing behavior/socio-emotional skills have a higher correlation with verbal skills than math skills, we perform a regression analysis where the dependent variables are performance in KS2 math or verbal (English) exams and the independent variables are the SDQ measures.[98] Table 26 shows regression outcomes where each coefficient corresponds to a separate regression (in each of them we control for gender).[99] Panel A of Table 26 shows that, while these proxies are highly predictive of math and verbal scores, they do not favor one skill over the other, i.e. all components of the SDQ questionnaire have similar effects on both exams. For example, a one-point increase in hyperactivity problems decreases the verbal test score by 0.169 of a standard deviation, which is very similar to the effect in math (0.165). Therefore, these results suggest that the larger effect of verbal skill on college enrollment is not likely to be driven by a larger correlation between verbal skill and externalizing behavior.[100]

Similarly, as we have discussed, family background characteristics might disproportionately impact verbal skill relative to math skill, and with inadequate controls we risk misattributing the effect of these characteristics on university enrollment to verbal skill. In our empirical model, we address this issue by following two strategies. First, we included the following controls for family background characteristics: free school meal eligibility, race, mother tongue, special education needs, eight measures of neighborhood characteristics, and effectiveness of school attended. Second,

[96] The SDQ is a behavioral screening questionnaire for children and adolescents ages 2 through 17 years old and was developed by the child psychiatrist Robert N. Goodman.

[97] We cannot reproduce the factor model analysis with this database because it only contains aggregate measures of performance in math and English rather than the detailed measures needed to identify the factor model. In addition, this database lacks college enrollment outcomes.

[98] The test scores on KS2 math and verbal have been standardized to have mean 0 and standard deviation 1

[99] We did not include all the measures in one regression because they are highly correlated, making the interpretation of the coefficients difficult due to multicolinearity.

[100] To resemble the structure of the factor model, this analysis considers math and English test scores as dependent variables. If instead, we were performing regressions where the SDQ questions would have been the dependent variable, and math and English test scores independent variables (i.e. controlling for math and English performance simultaneously), our results would have remained the same. Specifically, the coefficients on math and English in an OLS regression where the dependent variable is the average of the SDQ questions are -0.306 and -0.300, respectively (both coefficients are significant at the 1% level).

Table 26: Linear Regression Model: Key Stage 2 Test Scores and ALSPAC Data

	Verbal	Math
Panel A: Strengths and Difficulties Questionnaire (SDQ)		
Hyperactivity Problems (obs. = 5,434)	-0.169 (0.005)	-0.165 (0.005)
Emotional Problems (obs. = 5,464)	-0.092 (0.007)	-0.112 (0.006)
Conduct Problems (obs. = 5,460)	-0.163 (0.009)	-0.149 (0.009)
Peer Problems (obs. = 5,464)	-0.094 (0.007)	-0.103 (0.007)
Pro-social (obs. = 5,461)	0.084 (0.006)	0.081 (0.006)
Average SDQ (obs. = 5,424)	-0.247 (0.009)	-0.250 (0.009)
Panel B: Family Background Characteristics		
Parents Own House (obs. = 9,356)	0.549 (0.023)	0.534 (0.024)
Father Lives at Home (obs. = 7,985)	0.327 (0.032)	0.322 (0.032)
Mother College Degree (obs. = 10,232)	0.789 (0.028)	0.756 (0.029)
Father College Degree (obs. = 9,845)	0.753 (0.025)	0.727 (0.025)
Panel C: IQ Test		
WISC IQ Test (obs. = 6,427)	0.035 (0.001)	0.038 (0.001)

Notes: Each coefficient corresponds to a separate regression. The dependent variables, overall Key Stage 2 math and verbal test scores, have been standardized to have mean 0 and standard deviation 1. Average SDQ denotes the mean of the Strengths and Difficulties Questionnaire, where a higher value represents more severe externalizing behavior problems. All specifications include controls for gender.

we further examined our findings by conditioning on early Key Stage skills (i.e. KS1), which perhaps serve as sufficient statistics for any unobserved family background characteristic.[101] In order to perform a final check on this assumption, we make further use of the ALSPAC database, which provides more detailed information on parental background characteristics. Panel B of Table 26 shows OLS regressions of family background covariates such as parental education (i.e. parents' holding a college degree), and proxies for family composition (i.e. father living at home) and parental income (i.e. home ownership status) on KS2 math and verbal performance (note that each coefficient corresponds to a separate regression). Overall, the results seem to indicate that there is no differential effect of family background characteristics on math and verbal performance. For example, having a mother (father) with a college degree increases KS2 English and math performance by 0.789 (0.753) and 0.756 (0.727) of a standard deviation, respectively. In summary, these findings further suggest that our main results are not likely to be driven by differential effects of family characteristics on math and verbal test scores.

Finally, using the ALSPAC database, we also explore the correlation between IQ tests and KS2 test scores. Panel C of Table 26 shows the interrelation between the Wechsler Intelligence Scale for Children (WISC) IQ test and KS2 exams.[102] Results show that both math and verbal scores are highly and similarly correlated with the WISC score, suggesting that verbal test scores are not proxying students' IQ differentially than math test scores.

Can University Supply Explain the Relative Importance of Verbal Skills? It is possible to argue that the large effect of verbal skills on post-secondary enrollment could be a consequence of universities in the United Kingdom mainly offering programs that do not require an intensive use of math skills, e.g., humanities or social science. To study this possibility, for each enrolled student, we look at the subject specific A-level course requirements for the actual degree program in which they are enrolled. This information was extracted from the document "Informed Choices" created by the Institute of Career Guidance and the Russell Group universities. This publication provides information to all students considering A-level and equivalent options. We find that nearly half (44.5%) of students enrolled in university were required to obtain a qualification in the sciences, e.g., math or physics, before enrolling in college. The remaining 55.5% enrolled in programs with no science requirement, with 31% having a non-science requirement, and 24.5% having no requirement.

Using the same method, but broadening our definition to include programs that either require or recommend taking at least one science related A-level, we find that 62.3% of those in University are enrolled in a degree program with a recommended science qualification prior to enrollment. Overall, these simple statistics suggest that math skills are in fact required by universities and, therefore, our results are not likely to be driven by the type of majors that are offered in the higher education system in the UK.

[101] Similar assumptions have been made in the literature of teacher value-added (Todd and Wolpin, 2003).

[102] The Wechsler Intelligence Scale for Children (WISC) is an intelligence test for children between the ages of 6 and 16. The total IQ score represents a child's general intellectual ability. It also provides five primary index scores: verbal comprehension index, visual spatial index, fluid reasoning index, working memory index, and processing speed index. In this sample, the mean of the IQ score is 104 points and the standard deviation is 16.1. The raw correlations between the verbal and math test scores with the WISC index are 0.69 and 0.61, respectively.