The Path to College Education: Are Verbal Skills More Important than Math Skills?*

Esteban M. Aucejo†

London School of Economics and Political Science

Jonathan James‡

California Polytechnic State University

March 26, 2016

Abstract

The aim of this paper is to study the differential roles of math and verbal skills for educational outcomes. By estimating a multi-period factor model of skills, using a rich panel database that follows all students in England from elementary school to university, we find that verbal skills play a greater role in explaining university enrollment than math skills. In addition, we use our framework to study the timing of skill development during compulsory schooling. Results show that 40% of skills measured at the end of compulsory education are developed between the ages of 7 and 16, which indicates some scope for overcoming initial skill disadvantages. Finally, we study the gender gaps in college enrollment and STEM field enrollment, showing that verbal skills and comparative advantage in skills are key determinants of these gaps.

*We thank Peter Arcidiacono, Ghazala Azmat, Caroline Hoxby, Monica Langella, Alan Manning, Guy Michaels, Steve Pischke, Tyler Ransom, and Matteo Sandi 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, and University of Rochester. All errors are our own.

†Department of Economics, London School of Economics and Political Science. 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 earnings and employment prospects of less-educated workers have experienced a sharp decline since the early 1980's (Acemoglu and Autor, 2011). 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 are an important determinant of educational attainment (Heckman et al., 2006; Cawley et al., 2001; Cameron and Heckman, 2001). However, skills are multiple in nature (Cunha and Heckman, 2007). Therefore, more attention should be given to understanding exactly which types of skills have the greatest influence on post-secondary educational outcomes and how these specific skills evolve over the schooling career.

The aim of this paper is threefold. First, we study whether math and verbal skills, developed during compulsory education, differentially impact university enrollment and other university outcomes. Unpacking skills in the math-verbal dimension offers a deeper understanding of the education decision process and compliments previous work that has focused on the role of socio-emotional skills (Almlund et al., 2011). Second, we study the rate at which math and verbal skills are shaped over the formative years of compulsory education. In this regard, we investigate the malleability of skills and the extent to which shocks to skill formation (e.g. having a better teacher in later schooling years) can help to overcome initial skill disadvantages. Third, we analyze whether differences in math and verbal skills between males and females explain the well established gender gaps in college enrollment and in STEM (Science, Technology, Engineering, and Math) major choice. Evidence from many developed countries has shown that males are less likely to attend university than females, but males are more likely to enroll in STEM fields. However, little is known regarding the role 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 for more than 500,000 students. For each student we observe a rich set of demographic characteristics, the elementary school attended, as well as university records that show whether the student progresses to university, their institution and field of study, and whether they graduated.

Most important, for each student we observe the results from more than 30 subject-specific exams taken over the period of compulsory education.[1] We use these performance measures to estimate a multi-period factor model that precisely recovers the latent skills (i.e. math and verbal) for each individual at different points in the schooling career. Using the estimates of the latent skills, we study how these skills influence educational outcomes.

First, our results indicate that verbal skill has a substantially stronger influence on university enrollment and graduation than math skill. Specifically, a one standard deviation increase in math skill improves the probability of university enrollment by 9.6 percentage points, while a similar increase in verbal skill leads to an improvement of 18.7 percentage points. The relative importance of verbal skill is robust to different model specifications, subsamples, and accounts for multiple sources of endogeneity, including parental inputs and student motivation.[2] We provide evidence that this result is *not* driven by conflating factors like socio-emotional skills and family background characteristics. Furthermore, we confirm similar patterns with data from the United States.

Second, we provide a novel estimate that describes the timing of skills development. We find that about 40% of math and verbal skills measured at the end of compulsory education are developed between the ages of 7 and 16. This finding suggests there is scope during compulsory education to overcome initial skill disadvantages. Third, we show that while gender differences in math skill are small, females have a large advantage in verbal skill, which are on average 39% of a standard deviation higher than the verbal skill of males. Combined with our main result, that verbal skill has a disproportionate influence on college enrollment, we show that the gender difference in verbal skill is the main driver of the gender gap in college enrollment. In fact, after controlling for math and verbal skills, females are slightly *less* likely to attend college than males. Females' advantage in verbal skill also has implications for the gender gap in STEM major choice, as we show that comparative advantage in math heavily influences the decision to major in STEM. Since females and males have similar distributions of math skill, the male disadvantage in verbal translates to a comparative advantage in math, leading to an increase in male representation in STEM.

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

[2] This result is consistent with one of the findings in Chetty et al. (2014). The authors show that a high quality English teacher corresponds to a greater increase in students' college quality (almost twice) than a high quality math teacher.

Furthermore we document that even for males and females with the same bundle of skills, the role played by comparative advantage is much stronger for males, which also contributes to the gender gap in STEM enrollment.

Our finding on verbal skill 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 skill has 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). Regressions with skill and curriculum effects that control for years of education may mute one of the main channels in which verbal skills influence labor market outcomes, through increased 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.

This paper also contributes from a methodological perspective by offering a tractable method to estimate a high-dimensional, correlated factor model (with nine factors in total) that contains more than 140,000 parameters. We demonstrate how the model can be easily estimated with a simple expectation-maximization algorithm (Dempster et al., 1977; Ruud, 1991). For our context, this is implemented by iterative single-equation least squares estimation. This approach is ideal given our big data setting, which contains close to 15 million data points (30 measures for each 500,000 individuals).

The rest of the paper is organized as follows. Section 2 describes the data and the institutional setting of education in England. Section 3 describes the empirical model and our estimation approach. Section 4 shows the main results regarding the importance of verbal skills for college enrollment. Section 5 reinforces our main findings by providing additional supportive evidence. Section 6 discusses practical implications of the main findings for empirical research. Section 7 analyzes the evolution of skills during the schooling years. 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. We also perform simple regressions to summarize the data and show preliminary evidence of the main data patterns we wish to investigate in our formal 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.[3] Table 1 summarizes the stages of the English compulsory education system. Students enter school at age 4, the Foundation Stage, then proceed to Key Stage 1 (KS1), spanning ages 5 and 6, and Key Stage 2 (KS2, ages 7 to 11).[4] 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-levels performance.

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 infor-

[3] Recently, a series of reforms regarding the assessment of students have been implemented. However, these reforms were not in place for the years that we are analyzing.

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

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

mation on approximately 500,000 students, which only excludes students in independent (private) schools because these schools are not covered in the census.[5] Our database links information from the census of all state school children in England with information from the Higher Education Statistics Agency (HESA).[6] HESA collects information on all students in public founded 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 math ability at each of the four Key Stages for each student.[7]

Table 2 presents summary statistics of the key variables in our data. The top panel shows information on student background characteristics. The Income Deprivation Affecting Children

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

[6]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.

[7]This paper does use the A-level information because only those students who are college bound will continue to A-level, producing selection bias issues.

Index (IDACI) is an index of poverty in the neighborhood[8] where the student lives. It measures the proportion of children in a local area under the age of 16 that live in a low-income household, where higher scores correspond to more impoverished areas. The data also identifies 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 15% of the students in our sample are eligible for FSM, where differences between genders are small. On the contrary, the indicator for SEN shows that only 16.9% of female students are included in this category compared to 27.9% of male students. In the sample, 94.7% of the students have a mother who speaks English, and more than 89% of students are white. The second panel of Table 2 shows overall performance in national assessment exams in math and English at each stage of the schooling career.[9] The data show that females largely outperform males in verbal exams at each KS, while in math males seem to perform better until KS3. The last row of this panel indicates that students take 8.17 GCSE subject-exams on average, which includes the compulsory math and English exam, where females tend to take more exams than males. The third panel shows average authorized and unauthorized absence rates (i.e. average proportion of sessions absent) in KS4. Students are absent for 9.4% of the sessions, where most absences are authorized and differences between genders are small. Finally, the bottom panel provides an overview of post-secondary education outcomes. Around 36.3% of the students in our sample enrolled in university, with females being 7 percentage points more likely than males to enroll. 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).[10] Approximately 7% of

[8]Neighborhood denotes a lower layer super output area (LSOA), about 1,500 people, which is roughly equivalent to a postcode area.

[9]Test scores have been standardized to have mean zero and standard deviation 1. While we have information on more test scores in KS4, Table 2 only presents overall performance in math and verbal.

[10]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.

students enroll in these institutions, with females being overrepresented.

2.3 Preliminary Evidence

Before moving forward, this section uses simple regressions to describe the aspects of the data we intend to study more deeply with our main analysis. Table 3 shows the results of a linear probability model of university enrollment with separate regressions controlling for the math and verbal test scores at each Key Stage.[11] These results offer three main insights. First, test scores are highly predictive of university enrollment. The regression including the Key Stage 4 tests explains almost 36% of the college enrollment decision, as measured by the R-squared. Second, while tests are highly predictive of university enrollment, verbal scores appear to have a larger effect relative to math scores. The magnitude of the difference varies across the Key Stages. The difference is most pronounced in Key Stage 2, where the coefficient on the verbal score is 39% larger than the coefficient on math. Third, Table 3 shows that controlling for test scores in these regressions has a large impact on the coefficient for the female indicator. This suggests that skill differences between genders may play an important role in explaining the gender gap in certain educational outcomes.

While Table 3 offers tangential evidence on the relationship between skills and university enrollment, there are a number of shortcomings with this analysis. First, the use of test scores in these regressions only proxies for skills, which does not directly address our research question. It is not clear *a priori* which of the 70 observed measurements should be used to proxy for which skills. 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. Second, any method used to weight the scores to form aggregates would be arbitrary, with no method to guide which weighting scheme best captures skills. Third, any averaging of test scores will not fully correct for measurement error, which will lead to bias. Finally, this analysis provides no insight into the sources of variation in skills and how skills evolve over the schooling career. Simply studying correlations in test scores across time periods suffers from each of the

[11]Total performance in math and English in each Key Stage was obtained by adding the performance in the different subtests when there is more than one subtest score (e.g. adding test scores in reading and writing). The department of education in the United Kingdom reports total performance in this manner.

Table 2: Summary Statistics: Overall and by Gender

	All		Males		Females		All	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Min.	Max.
Background Characteristics								
IDACI Index	0.210	(0.174)	0.207	(0.173)	0.213	(0.176)	0.003	0.993
Free School Meal	0.150	(0.357)	0.146	(0.353)	0.154	(0.361)	0	1
Special Education Needs	0.224	(0.417)	0.279	(0.448)	0.169	(0.374)	0	1
Mother Tongue English	0.947	(0.224)	0.947	(0.223)	0.947	(0.226)	0	1
Race								
White	89.46%		89.61%		89.30%			
Asian	5.74%		5.69%		5.78%			
Black	2.41%		2.35%		2.47%			
Other	2.40%		2.34%		2.46%			
Standardized Key Stage (KS) Test Scores								
KS1 Math	0	(1.000)	0.014	(1.036)	-0.014	(0.962)	-3.234	3.089
KS1 Verbal	0	(1.000)	-0.148	(1.020)	0.148	(0.956)	-3.017	3.604
KS2 Math	0	(1.000)	0.057	(1.014)	-0.057	(0.982)	-3.194	1.802
KS2 Verbal	0	(1.000)	-0.108	(1.002)	0.108	(0.986)	-4.000	2.692
KS3 Math	0	(1.000)	0.035	(0.999)	-0.036	(0.999)	-3.204	2.148
KS3 Verbal	0	(1.000)	-0.176	(1.001)	0.177	(0.967)	-2.104	2.706
KS4 Math	0	(1.000)	-0.003	(1.003)	0.003	(0.997)	-2.582	1.900
KS4 Verbal	0	(1.000)	-0.175	(1.017)	0.176	(0.950)	-3.057	1.998
Total GCSE Exams	8.170	(2.070)	8.009	(2.157)	8.332	(1.963)	2	18
Average Proportion of Sessions Absent in KS4								
Authorized Absences	0.074	(0.073)	0.069	(0.069)	0.079	(0.077)	0	1
Unauthorized Absences	0.020	(0.059)	0.019	(0.056)	0.021	(0.072)	0	1
University Outcomes								
University Enrollment	0.363	(0.481)	0.330	(0.470)	0.400	(0.488)	0	1
University Graduation	0.253	(0.435)	0.219	(0.413)	0.287	(0.452)	0	1
University Enrollment STEM	0.118	(0.323)	0.140	(0.347)	0.096	(0.294)	0	1
University Graduation STEM	0.075	(0.263)	0.083	(0.275)	0.067	(0.250)	0	1
University Enrollment Top 24	0.071	(0.257)	0.066	(0.248)	0.077	(0.266)	0	1

Table 3: Linear Probability Model: University Enrollment

	Baseline	KS1	KS2	KS3	KS4
Constant	0.330	0.343	0.346	0.360	0.359
Female	0.066	0.039	0.046	0.020	0.007
Math Test Score	-	0.094	0.101	0.127	0.141
Verbal Test Score	-	0.098	0.140	0.159	0.165
R-Squared	0.005	0.136	0.214	0.295	0.357
Obs.	498,736	495,197	480,157	471,095	498,736

Notes: Math and verbal test scores have been standardized to have mean zero and standard deviation 1. Standard errors not shown. All coefficients are statistically significant at the 99% level.

issues described above.

The next section outlines the factor model that we use to address these issues. Factor analysis is a statistical method that condenses the covariance among available measures using low dimensional latent variables, which produces an output that is easy to interpret and directly addresses the problem of measurement error. This approach is ideal for our investigation because it allows us to directly measure skills and enables a comprehensive use of our rich data.

3 Empirical Model and Estimation

In our data we observe over 70 measures of student performance from age 5 to 16. We use this data to estimate a multi-period factor model that reduces these measures to a low-dimensional vector of skills that is easy to interpret.[12] After recovering the latent skills, we study the three main questions mentioned earlier. First, what influence do math and verbal skills have on decisions in higher education? Second, how do these skills evolve over the course of compulsory education? Third, what is the role of these skills in explaining the gender gaps in university enrollment and in STEM major choice? In this section we explain our factor approach, how we address endogeneity

[12] Our approach follows the spirit of Heckman et al. (2013). However, we use different econometric methods due to the high dimensionality of our data, i.e. around half a million students.

of parental inputs and student effort, and our estimation strategy.

3.1 Factor Model of Skills

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, skills for student i are denoted by θ_{it}. We model a total of nine skills, which include math and verbal skills for each of the four Key Stages and an additional factor in Key Stage 4 that captures the motivation of the student to attend higher education.[13] Let $\theta_i = \begin{bmatrix}\theta'_{i1} & \theta'_{i2} & \theta'_{i3} & \theta'_{i4}\end{bmatrix}'$ be the complete vector of these nine factors, which are modeled jointly as

$$\theta_i = female_i * \psi + \Phi x_i + \xi_i \tag{1}$$

(factor equation)

The factors are determined by characteristics observed by the econometrician (gender and x_i), and an unobserved component ξ_i. We assume that the unobserved component of skills is drawn from a mixture of C multivariate normal distributions, which allows us to approximate many distributions. Specifically, ξ_i is drawn from $\mathcal{N}(\delta_c, \Sigma)$ with probability π_c for $c = [1, 2, \ldots, C]$. Each component of the mixture has a different mean but they share a common covariance. Skills are allowed to be correlated contemporaneously and across time, so Σ is a full covariance matrix. We model the nine factors in Eq. (1) as a joint distribution because this offers more concise matrix notation.[14]

The factors from Eq. (1) are not observed. However, our data contains frequent and extensive measures for each student that we use to recover these latent skills (i.e. θ_i). Let w_{im} for $m =$

[13] We believe this is the most parsimonious factor model given the questions that we are addressing. We also explored a different factor structure that contained a separate general factor in addition to a math and a verbal factor. While the results for the model with a general factor were similar to the ones reported in this paper, the estimation of such a model requires more restrictive identification assumptions.

[14] Appendix A shows how the joint distribution can be re-written in terms of conditional distributions to map our model to a dynamic factor model (cf. Cunha and Heckman 2007). Given our aim is to characterize how skills evolve during compulsory education (among other things) and not to estimate the technology of skill formation, it is unnecessary to express our model in a dynamic factor framework.

$1, 2, \ldots, M$ denote the M measures for individual i. The measures are determined by

$$w_{im} = \mu_m + \theta_i' \lambda_m + \eta_{im}$$

(measurement equation)

where μ_m is the mean of the m^{th} measure and λ_m are the loadings on the factors for measurement m. Our factors span multiple time periods, so only the factors associated with the time period of measurement will have non-zero loadings. Finally, η 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 τ_m^2.

Key Stage 4 is the final period of compulsory schooling. At this stage, each individual is characterized by a vector of skills θ_{i4}, which contains three elements: a math skill, verbal skill, and a measure of motivation for pursuing further education. After this period, students may conclude their formal education or make additional investments. We are interested in understanding how these skill measures influence these decisions. There are K outcomes for each individual, with realization y_{ik}^* for $k = 1, \ldots, K$ that follows

$$y_{ik}^* = female_i * \gamma_k + \theta_{i4}' \alpha_k + x_i' \beta_k + \varepsilon_{ik}$$

(outcome equation)

where γ_k is the influence of gender on the observed outcome, θ_{i4} is the vector of skills at the end of compulsory schooling, x_i includes other control variables (including a constant), and ε_{ik} includes the remaining determinants of the outcome variable that cannot be explained by the other parts of the model and is assumed to be independent. The outcomes we study are university enrollment, university quality, major field of study, and university graduation. All of these outcomes are discrete, so y_{ik}^* represents the underlying latent variable process. We assume that ε is distributed type-I extreme value, and we only observe the outcome $y_{ik} = 1$ if $y_{ik}^* > 0$ and zero otherwise.

3.2 Endogeneity

To identify the impact of math and verbal skills on the outcomes, we need to address certain aspects of endogeneity between the measures used to extract the skills and the outcome equations themselves. Specifically, we need to rule out that unobserved covariates which may impact college enrollment are not differentially affecting math and verbal skills. A first potential problem is parental characteristics. Parental characteristics are crucial for skill development and influential in the college decision. For example, if a given (unobserved) family characteristic has a larger impact on verbal skills than math skills, then we may misattribute to verbal skills the effect of background characteristics. We follow two strategies to deal with this issue. The first is to include observed variables in the explained portion of the factors that can account for heterogeneity in parental inputs. To this extent, we include the IDACI score of student's neighborhood, an indicator if the student qualifies for a free or reduced price lunch, race, the native language of the student's mother, and a full set of elementary school fixed effects. Since we use a single cohort of students, each elementary school has on average 30 students, which gives us more than 15,000 fixed effects for each of the factors (a total of 135,000 fixed effects). The second strategy we use to address the endogeneity of parental inputs and college enrollment exploits the multi-period modeling of skills. Assuming that performance on earlier tests constitute sufficient statistics for parental input, we are able to use marginal changes in skill development in later periods to identify the effect of specific skills on the outcomes.

The second potential source of endogeneity occurs later in the academic career. Many students decide not to go to college prior to the completion of compulsory schooling. These students in turn may put in low-effort and have low-performance on some of the measures. To address this concern, we include a third factor in the final stage of compulsory schooling to capture motivation. While motivation is a difficult characteristic to capture, we identify this factor from three sources. The first two sources are the total number of excused and unexcused absences during KS4. The third source is the number of subject-specific tests taken during KS4. Since students are required to take at least six subject specific tests during KS4 before advancing to A-levels, we use this information to identify their intent for higher education. By including this third factor in KS4, our goal is to identify the causal impact of the other skills on college outcomes, holding motivation constant. Finally, we also analyze the relative importance of skills using earlier measures of skills (e.g. age 7), where academic effort is less likely attached to aspirations of higher education.

3.3 Estimation

Estimation of the factor model requires certain normalizations for interpretation and identification. For a tangible interpretation of the factors as math and verbal skills, we place restrictions on the factor loadings in certain measurements. Specifically, we will choose some measures to load only on the verbal factor and others to load only on the math factor, while others will load on both, e.g., geography. A full list of the factor loading restrictions that we place on the 70 measurements as well as the other normalizations that we impose for identification are outlined in Appendix B.

Given our distributional assumptions on the unobserved data, the parameters can be estimated via maximum likelihood. We use a two-stage estimation approach similar to Heckman et al. (2013). In the first stage, we use all of the observed measurements to jointly estimate the parameters of the measurement system and the factor structure. In the second stage, we use the parameter estimates to construct distributions for the latent skills of each individual and estimate the coefficients of the outcome equations, integrating over these distributions. This process is formalized below.

Given data for N individuals, let $w_i = [w_{i1}, \ldots, w_{iM}]$ be the observed measurements and $y_i = [y_{i1}, \ldots, y_{iK}]$ be the observed outcomes for individual i. To form the likelihood, we write $L(y_i|x_i, \theta)$, the likelihood of observing the outcome y_i for a given value of the unobserved skills, θ. Conditional on θ, the outcome equations are independent, so $L(y_i|x_i, \theta)$ is a product of logit probabilities. In addition, we compute $L(w_i|\theta)$, the likelihood of observing the measurement variables conditional on θ, which is a product of univariate normal probability density functions. Conditional on θ, the remaining unobserved components of the measurements and the outcomes are independent, so the joint probability is the product of these two likelihoods. Therefore, the parameters are estimated by maximizing the integrated likelihood function:

$$LL = \sum_{i=1}^{N} \ln \left[\int_{\theta} L(y_i|x_i, \theta) L(w_i|\theta) f(\theta|x_i) d\theta \right]$$

Where $f(\theta|x_i)$ is the probability density function for a mixture of normals specified in Eq. (1).

Next we apply Bayes' rule to the likelihood to estimate the parameters in the measurement system separately from the parameters in the outcome equation. Let $h(\theta|w_i, x_i)$ be the probability density function of θ conditional on both the measurements and the covariates. From Bayes rule we have the identity

$$L(w_i|\theta)f(\theta|x_i) = h(\theta|w_i, x_i)\left[\int_{\theta'} L(w_i|\theta')f(\theta'|x_i)d\theta'\right]$$

Plugging this into the log-likelihood, the log-likelihood becomes

$$LL = \sum_{i=1}^{N} \ln\left[\int_{\theta} L(y_i|x_i, \theta)h(\theta|w_i, x_i)d\theta\right] + \sum_{i=1}^{N} \ln\left[\int_{\theta'} L(w_i|\theta')f(\theta'|x_i)d\theta'\right]$$

The additive separability of this log-likelihood allows us to estimate the parameters in two stages. First we maximize the last component of the log-likelihood, which contains the parameters in the measurement equations and the factor distribution. Second, using these estimates we construct the conditional distributions $h(\theta|w_i, x_i)$ and maximize the portion of the likelihood containing the outcome equations. As stated by Heckman et al. (2013), the two-step approach is less efficient than joint estimation of the entire model because the information in y is not used to help identify the factor distribution parameters. However, in addition to being more tractable, this approach is beneficial because identification of the latent skills only comes from the measurements, which provides transparency.

In the first step, we search for the parameters that maximize the log-likelihood function

$$LL_{step-one} = \sum_{i=1}^{N} \ln\left[\int_{\theta} L(w_i|\theta)f(\theta|x_i)d\theta\right] \tag{2}$$

One potential concern with this estimation approach in our setting is that we do not observe all measures for all students. Since we do not directly model the selection process, we need to make additional assumptions to proceed. We observe nearly complete coverage for the measures used in KS1 to KS3. For KS4, all students take a mandatory math test and English test. In addition, at

KS4, they choose on average between 4 to 6 additional subject tests. We use the score on these subject tests to recover the factors. This creates a potential selection problem since those with high math ability will likely choose a different set of subjects than those with lower math ability. Our assumption is that, once we condition on the observed scores for all of the mandatory tests, including the math and English test at Key Stage 4, the choice of subject test occurs approximately at random.

Given the dimensionality and complexity of this problem, finding the parameters that maximize Eq. (2) poses a number of challenges. First, we use data on more than 15,000 elementary schools. Since we include elementary school fixed effects for each of the nine factors, this alone yields more than 135,000 parameters. In total, our baseline specification includes around 140,000 parameters. Conventional numerical optimizers based on Newton's method are not possible because the hessian matrix cannot be stored in read/write memory. Second, even if a suitable large-scale algorithm is found, the factor model contains many constraints that are difficult to impose during estimation. For example, the nine dimensional covariance matrix must be positive semi-definite.

We overcome these computational challenges by maximizing Eq. (2) with the expectation-maximization (EM) algorithm (Dempster et al., 1977). The EM algorithm is an iterative procedure for maximizing complicated integrated likelihood functions. In the expectation step, the current iteration parameters are used to construct individual densities of the unobserved data. In the maximization step, new parameters are found by performing complete data maximum likelihood that treats the unobserved data as observed. The appeal of this method is that it is easy to implement. For example, the parameters in Eq. (1) when θ_i is observed can be found using equation-by-equation OLS. This only requires the inversion of a 15,000 by 15,000 matrix of fixed effects, which is the same for each regression and can be calculated and inverted outside of the algorithm. Second, many of the constraints, like positive semi-definiteness of the covariance matrix, are naturally imposed by the algorithm. The steps of the EM algorithm are outlined in Appendix C. One shortcoming of the EM algorithm is that it has, in some cases, a slow rate of convergence. To speed up the convergence of the algorithm we use the SQUAREM accelerator in Varadhan and Roland (2008), which can deliver super-linear rates of convergence.

Once the first stage parameters are estimated, we calculate the conditional distribution of the latent skills for each individual, $h(\theta|w_i, x_i)$ and maximize the likelihood.

$$LL_{step-two} = \sum_{i=1}^{N} \ln\left[\int_{\theta} L(y_i|x_i,\theta)h(\theta|w_i,x_i)d\theta\right]$$

Since the outcome equations are all discrete, this integrated likelihood function represents a mixed logit model. The maximization of this likelihood is only over the parameters in $L(y_i|\cdot)$, denoted $\Psi = [\gamma, \alpha, \beta]$. The parameters that maximize this likelihood are the ones that are a root to the score function.

$$\begin{aligned}
\frac{\partial LL_{step-two}}{\partial \Psi} &= \sum_{i=1}^{N} \frac{1}{\int_{\theta'} L(y_i|x_i,\theta')h(\theta'|w_i,x_i)d\theta'} \int_{\theta} \frac{\partial L(y_i|x_i,\theta)}{\partial \Psi} h(\theta|w_i,x_i)d\theta \\
&= \sum_{i=1}^{N} \int_{\theta} \frac{\partial \ln\left[L(y_i|x_i,\theta)\right]}{\partial \Psi} \frac{L(y_i|x_i,\theta)h(\theta|w_i,x_i)}{\int_{\theta'} L(y_i|x_i,\theta')h(\theta'|w_i,x_i)d\theta'} d\theta \\
&= \sum_{i=1}^{N} \int_{\theta} \frac{\partial \ln\left[L(y_i|x_i,\theta)\right]}{\partial \Psi} h(\theta|y_i,w_i,x_i)d\theta
\end{aligned}$$

The density function $h(\theta|y_i, w_i, x_i)$ is the conditional density of the unobserved factor conditional on all of the data. In this second stage, we impose the restriction on the data that the outcome equations provide no additional information on the factors once we condition on the observed measurements. The primary purpose of this assumption is that it facilitates validation of the model. Our goal is to understand how much variation of the outcome variables can be explained by these skills. By imposing this restriction we risk having less explanatory power. However, as described previously, this approach provides more transparency over the identification of the parameters. A second benefit of this assumption is that the likelihood reduces to an integrated standard logit, which requires maximizing

$$LL_{step-two} = \sum_{i=1}^{N} \int_{\theta} \ln\left[L(y_{ik}|x_i,\theta)\right] h(\theta|w_i,x_i)d\theta$$

Because the integral is outside of the log function, maximizing this likelihood can be done equation-by-equation. There is no known closed form for the integral of this likelihood, so we approximate it with 10 simulated draws for each individual.[15]

4 Main Results

The estimates from the factor model facilitate a deeper investigation of the patterns observed in the simple analysis in Section 2.3. The main benefit of the factor structure is that it allows us to extract measurements of unobserved skills from a large set of data, which can then be used to study the outcome equations. This section summarizes our main results. First, we characterize the distribution of the recovered skills in the population. Second, we provide insights into the nature of the recovered skills by reporting the factor loadings and residual variance for a selection of the KS4 measurements. Finally, we analyze the effect of these skills on educational decisions.

4.1 Factor Distribution

Our empirical strategy recovers nine correlated factors, a math and verbal factor for each of the four Key Stages plus a motivation factor for KS4.[16] Table 4 shows the population correlation matrix as well as the standard deviation of each estimated factor. As expected, the factors are highly correlated over time and across skills. The correlation between contemporaneous math and verbal skills is around 0.8 at each Key Stage. However, the correlation among factors that are more distant in time has a declining trend. For example, the coefficient of correlation between KS1 and KS2 math (verbal) factors is 0.742 (0.818) while the correlation for KS1 and KS4 is around 0.558 (0.641). The high correlation is due to the factors being functions of the same covariates and correlation in the factor error ξ.[17] Towards the end of compulsory education, skills in adjacent periods are highly correlated. KS3 and KS4 math (verbal) factors have a correlation of 0.927 (0.907), suggesting that skills in KS4 are mostly established in previous stages. Finally, the motivation factor measured in KS4 shows a similar correlation with the verbal and math factors.

The last row of Table 4 lists the standard deviation of the factors. Since all of the test score

[15] The integral is outside of the log operator, so there is no simulation bias even using a single draw for each individual.

[16] The motivation factor captures post-secondary education aspirations.

[17] The correlation between math and verbal skills at each Key Stage attributed to the factor error is smaller at around 0.7.

Table 4: Estimated Factor Population Correlation Matrix

	KS1 Math	KS1 Verbal	KS2 Math	KS2 Verbal	KS3 Math	KS3 Verbal	KS4 Math	KS4 Verbal	KS4 Motivation
KS1 Math	1.0	-	-	-	-	-	-	-	-
KS1 Verbal	0.826	1.0	-	-	-	-	-	-	-
KS2 Math	0.742	0.673	1.0	-	-	-	-	-	-
KS2 Verbal	0.712	0.818	0.826	1.0	-	-	-	-	-
KS3 Math	0.736	0.682	0.939	0.826	1.0	-	-	-	-
KS3 Verbal	0.633	0.733	0.725	0.907	0.822	1.0	-	-	-
KS4 Math	0.623	0.603	0.795	0.750	0.927	0.810	1.0	-	-
KS4 Verbal	0.558	0.641	0.659	0.803	0.787	0.907	0.890	1.0	-
KS4 Motive	0.455	0.510	0.555	0.626	0.693	0.722	0.820	0.844	1.0
Std. Dev.	0.872	0.902	0.939	0.888	0.762	0.832	0.922	0.922	1.410

Note: All correlations are statistically significant at the 1% level.

measures are normalized to have standard deviation one and some of the loadings are normalized to one, we can directly link the standard deviations in Table 4 to test scores. For example, both KS4 math and verbal factors have a standard deviation of 0.922. This implies that a one standard deviation increase in the KS4 math factor will increase the KS4 math test by 0.922 standard deviations and a one standard deviation increase in the KS4 verbal factor will increase the KS4 English test by 0.922 standard deviations. Although these factors have no meaningful unit of measure, they map similarly into their respective tests. Therefore, this offers a uniform interpretation as we look at the impact of changes in skills on university outcomes. The remaining parameters governing the factor equation are reported in Appendix D.

Figure 1 shows kernel densities for the KS4 factors by gender. While both groups show similar distributions in math skill, the difference in verbal skill substantially favors females. Table 5 shows that the difference in verbal skills in KS4 between males and females represents 39% of a standard deviation. Similarly, females outperform males in motivation, though the difference is much smaller (14% of a standard deviation). Table 5 also shows the pattern of the gender difference in skills across the Key Stages. With the exception of KS2 where males have a large advantage in math, the gender difference in math across the other Key Stages is small. On the contrary, females' large advantage in verbal is persistent across all Key Stages.

4.2 Factor Loadings

The factor model makes use of the multiple Key Stage test scores to identify the latent skills. Table 6 displays the loadings on the math and verbal factors and the residual variance for a subset of KS4 measurements. By looking at these estimates, it is possible to assess how skills load on each of the different test scores and to analyze the importance of measurement error. For example, Table 6 shows, as expected, that statistics mainly loads on math skill while social science relies heavily on verbal skill. However, courses such as geography and design and technology load similarly on both skills. Finally, the last column of Table 6 shows the variance of the component of each measurement that is independent of the factors, i.e. $var(\eta_{im})$. Given that each measure has variance 1, the residual variance denotes the proportion of the total variance that can be interpreted

Figure 1: Distribution of Key Stage 4 Factors by Gender

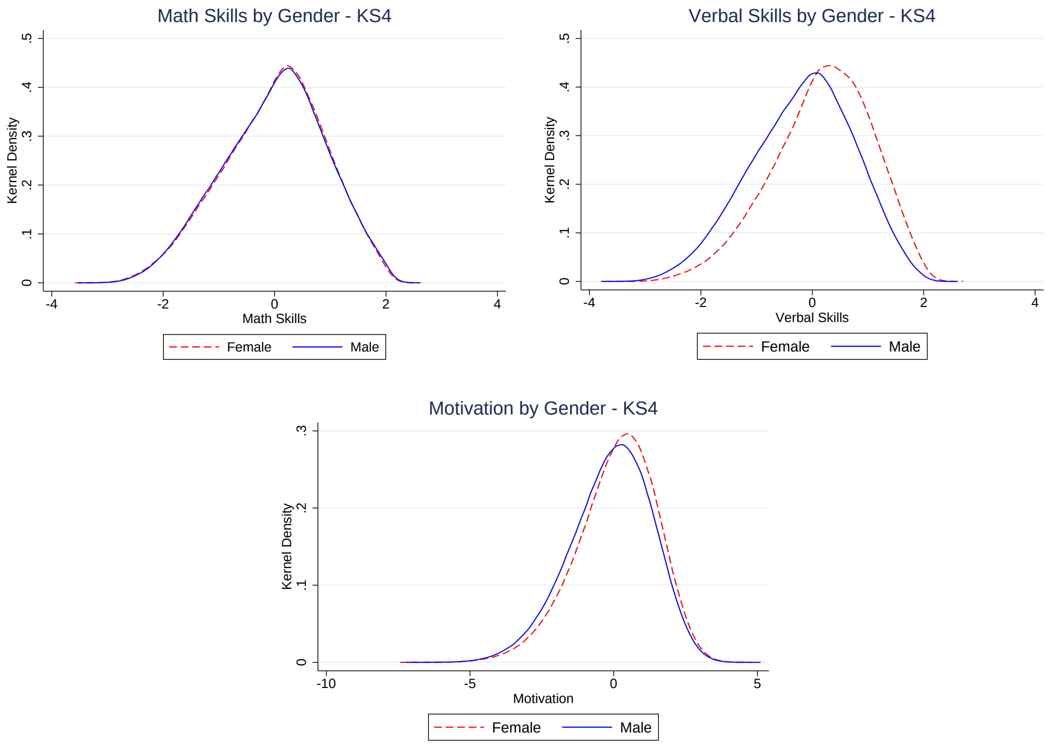


Table 5: Factor Means and Standard Deviations by Gender KS1-KS4

		Male		Female		Difference in Means
		Mean	S.D.	Mean	S.D.	
Math						
	KS1	0.008	0.916	-0.008	0.847	0.016
	KS2	0.051	0.979	-0.051	0.947	0.102
	KS3	0.017	0.785	-0.017	0.770	0.034
	KS4	-0.002	0.937	0.002	0.937	-0.004
Verbal						
	KS1	-0.137	0.930	0.137	0.873	-0.274
	KS2	-0.120	0.916	0.121	0.875	-0.241
	KS3	-0.163	0.844	0.164	0.808	-0.327
	KS4	-0.179	0.932	0.179	0.896	-0.358
Motivation						
	KS4	-0.100	1.431	0.100	1.400	-0.200

Note: Differences in means are statistically significant at the 1% level with the only exception being KS4 math.

Table 6: Factor Loadings: Selected Measurements only loading in Key Stage 4

Measurement	KS4 Math	KS4 Verbal	Residual Variance
Math	$1^{\dagger}$	0	0.127 (0.001)
English	0	$1^{\dagger}$	0.132 (0.001)
Design and Technology: Resistant Materials Technology	0.396 (0.007)	0.429 (0.007)	0.515 (0.004)
Geography	0.508 (0.006)	0.593 (0.006)	0.219 (0.001)
Social Science	0.088 (0.012)	0.864 (0.010)	0.369 (0.003)
Statistics	1.047 (0.010)	0.067 (0.008)	0.246 (0.004)
Physics	1.568 (0.007)	0	0.194 (0.003)
Chemistry	1.548 (0.007)	0	0.198 (0.004)
English Literature	0	1.045 (0.001)	0.220 (0.001)
Home Economics: Child Development	0.240 (0.021)	0.799 (0.021)	0.371 (0.005)
Double Science	1.056 (0.002)	0	0.191 (0.001)

† denotes normalized to 1. The zero values denote that a given skill is not loading on that specific measurement.

as "noise". Our estimates indicate that measurement error is pervasive on some of our measures. For example, noise is substantially larger in design and technology (51.5%) and much smaller in geography (21.9%). A full list of the estimated factor loadings and residual variances for the 70 measurements are in Appendix E.

4.3 The Role of Skills in Education Decisions

This section studies how skills affect the probability of attending and graduating from university. Table 7 shows the average marginal effects of a logistic model of university enrollment using only

Table 7: Logistic Regression: University Enrollment and KS4 Skills

	(1)	(2)	(3)	(4)
Average Marginal Effects				
KS4 Math	0.273 (0.001)	–	–	0.096 (0.001)
KS4 Verbal	–	0.298 (0.001)	–	0.187 (0.001)
KS4 Motive	–	–	0.301 (0.001)	0.029 (0.001)

Note: Results are from a logistic regression with controls for gender, IDACI Index, free school lunch, special education needs, race, mother tongue, and school fixed effects. Skills have been standardized to have mean 0 and standard deviation 1. Bootstrapped standard errors at the school level.

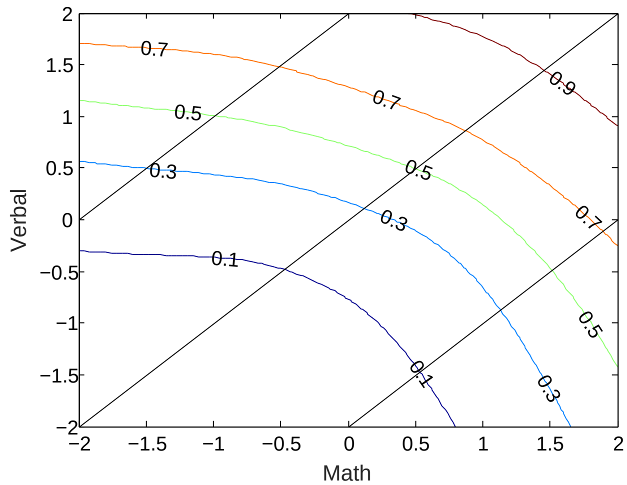
the skills recovered in KS4. We focus on KS4 skills because this is the final period of compulsory schooling when students begin making educational decisions. The regression analysis includes additional controls for background characteristics and school fixed effects. The first three columns of Table 7 show the effect of each skill separately, while the last column considers the skills jointly. Columns (1) to (3) show that each skill independently has a strong effect on university enrollment. For example, the computed average marginal effect of a one standard deviation increase in just one of these skills will improve college enrollment by more than 27 percentage points. However, Column (4) shows that when the skills are jointly considered, verbal stands out, while the impact of the others falls drastically.[18] For example, the effect of motivation becomes substantially less important once conditioning on verbal and math. Similarly, the coefficient on math skill falls by nearly a factor of three. In summary, Table 7 shows that the effect of verbal skill on college enrollment is almost twice as large as the effect of math skill.

To further describe our findings, Figure 2 plots level curves for the predicted probability of enrolling in university for different bundles of math and verbal skills measured in standard deviation units.[19] The slopes of these isoquants represent the technical rate of substitution between math and verbal skills for a given probability of enrollment. These curves are extremely flat for students with

[18]Appendix F shows identical results when we constrain the sample to white (British) students and also when we consider simpler functional form assumptions on the factor distribution.

[19]This three dimensional surface was constructed using locally weighted linear regression with actual enrollment decisions and the recovered distributions of individual skills.

Figure 2: Enrollment Probability Isocurves



low math skill. For example, the 0.3 probability line between -2.0 and -0.5 standard deviations away from the mean in math skill has a slope of 1/8, indicating that the rate of substitution strongly favors verbal skill (8 to 1). As math skill increases, the rate of substitution between math and verbal becomes more balanced. However, even for those with a combination of high verbal and high math skills, the rate of substitution still favors verbal at a rate of 3 to 2. Finally, the isoquants are generally concave, suggesting that, for the college decision, being strong in one skill is more important than being average. For example, a student that is +1.0 std. in math and -0.5 std in verbal has the same probability of attending college (30%) as a student that has -1.0 std in math and +0.5 std in verbal. However, if we considered a third student who had skills that where exactly in-between these students (i.e. 0 std. in math and 0 std. in verbal), she would have a lower probability of enrolling in college (25%).

Despite the fact that we control for a rich set of background characteristics, one concern with our finding is that verbal skill may be proxying for a background characteristic that is not accounted for in the model. To further investigate this issue, we exploit the multi-period nature of the factors. Since we have skill measures at each time period, we can decompose the skills observed in KS4 as a function of the earlier skills. Let $\hat{\theta}_{i4|3,2,1,0} = \mathrm{E}(\theta_{i4}|\theta_{i3}, \theta_{i2}, \theta_{i1}, x_i)$ denote the expected value of skills in KS4, conditional on everything that has happened to the student as of KS3, where conditioning on time zero includes observed background characteristics, i.e. IDACI, gender, race, etc.[20] We can re-write θ_{i4} as

$$\theta_{i4} = \underbrace{\left(\theta_{i4} - \hat{\theta}_{i4|3,2,1,0}\right)}_{\text{residual change in skill occurring in KS4}} + \hat{\theta}_{i4|3,2,1,0}$$

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

$$\theta_{i4} = \underbrace{\left(\theta_{i4} - \hat{\theta}_{i4|3,2,1,0}\right)}_{\text{KS4 Shock}} + \underbrace{\left(\hat{\theta}_{i4|3,2,1,0} - \hat{\theta}_{i4|2,1,0}\right)}_{\text{KS3 Shock}} + \underbrace{\left(\hat{\theta}_{i4|2,1,0} - \hat{\theta}_{i4|1,0}\right)}_{\text{KS2 Shock}} + \underbrace{\left(\hat{\theta}_{i4|1,0} - \hat{\theta}_{i4|0}\right)}_{\text{KS1 Shock}} + \hat{\theta}_{i4|0} \quad (3)$$

where $\hat{\theta}_{i4|2,1,0} = \mathrm{E}(\theta_{i4}|\theta_{i2}, \theta_{i1}, x_i)$, $\hat{\theta}_{i4|1,0} = \mathrm{E}(\theta_{i4}|\theta_{i1}, x_i)$, and $\hat{\theta}_{i4|0} = \mathrm{E}(\theta_{i4}|x_i)$

While the results in Table 7 show the total effect of skills when each of the five components in Eq. (3) 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, for example, all available information through KS1, which includes covariates for background characteristics and performance in KS1 math and verbal national assessments. If earlier years' performance in national tests serve as a sufficient statistic for unobserved background characteristics, in particular parental inputs (Todd and Wolpin, 2003), then conditioning on these tests in this way will correct for the endogeneity.

Table 8 shows the results of a logistic regression that includes the decomposition of KS4 math and verbal skills as described in Eq. (3). This regression allows us to consider the following counterfactual: does performing better than expected in math, once we condition on earlier test performance, have a larger impact on college enrollment than performing better than expected in verbal? The results in this table show that, at each Key Stage, performing better than expected

[20] $\hat{\theta}_{i4|3,2,1,0} = female_i * \psi_{[7:9]} + \Phi_{[7:9,:]} x_i + \Sigma_{[7:9,1:6]} \Sigma^{-1}_{[1:6,1:6]} \left(\begin{bmatrix} \theta_{i1} \\ \theta_{i2} \\ \theta_{i3} \end{bmatrix} - female_i * \psi_{[1:6]} + \Phi_{[1:6,:]} x_i \right)$. Where θ_{i1}, θ_{i2}, and θ_{i3} are observed and the notation $\Phi_{[1:6,:]}$ indicates the first through sixth rows of the matrix Φ.

Table 8: Logistic Regression: College Enrollment and Shocks to KS4 Skills

Average Marginal Effects	
KS1 Math Shock	0.075 (0.004)
KS2 Math Shock	0.099 (0.001)
KS3 Math Shock	0.113 (0.001)
KS4 Math Shock	0.122 (0.010)
KS1 Verbal Shock	0.194 (0.004)
KS2 Verbal Shock	0.197 (0.001)
KS3 Verbal Shock	0.200 (0.001)
KS4 Verbal Shock	0.180 (0.006)
Motivation (not shown)	

Note: Results are from a logistic regression that includes all the shocks, and controls for gender, IDACI Index, free school lunch, and special education needs. Shocks are defined in Eq. (3). Note that the shocks have been normalized into KS4 standard deviation units. Bootstrapped standard errors at the school level.

in verbal has a much larger impact on college enrollment than performing better than expected in math. For example, a one standard deviation 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 college enrollment of 19.7 percentage points, while a similar shock in math would lead to an increase of 9.9 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.

A second source of endogeneity that we need to address is motivation. While we model this directly, it is possible that it is not completely captured and might be driving our main result. The issue 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

Table 9: Logistic Regression: University Enrollment and Early Factors

	KS1	KS2	KS3
	(1)	(2)	(3)
Average Marginal Effects			
Female	0.031 (0.001)	0.032 (0.001)	0.010 (0.001)
Math	0.067 (0.001)	0.062 (0.001)	0.107 (0.001)
Verbal	0.116 (0.001)	0.184 (0.001)	0.173 (0.001)

Notes: Results are from a logistic regression with controls for gender, IDACI Index, free school lunch, special education needs, race, mother tongue, and school fixed effects. Skills are normalized to mean zero, standard deviation 1. Column (1) to (3) show results where math and verbal correspond to KS1, KS2 and KS3 respectively. Bootstrapped standard errors at school level.

motivation at school are less likely to be determined by decisions that will be made 11 years later. Table 9 repeats a regression similar to Table 7 except using skills from earlier Key Stages. Table 9 shows 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. For example, Table 9 indicates that the effect of KS1 verbal skills on university enrollment is 60% larger than KS1 math.

Next we study the effects of skills on other university outcomes: selectiveness of the university attended, field of study, and graduation. The first column in Table 10 looks at the effect of math and verbal skills on university enrollment when we remove those attending the most selective institutions, which is about 20% of enrollees. The purpose of this analysis is to distinguish if our main result is 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 7. Therefore, the larger effect of verbal skill 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 more insight, the second column of Table 10 shows the effect of skills on the intensive margin of school selectivity. This analysis only looks at college enrollees and shows that math and verbal skills have a similar effect on the probability of attending a selective institution. This implies that the selective universities are enrolling students equally from the top of both skill distributions. The last column

Table 10: Logistic Regression: Institution Type, Major and Key Stage 4 Skills

	Enrolled in Non-Selective Institution (conditional on not attending selective institution)	Enrolled in Selective Institution (conditional on attending university)	Enrolled in STEM (conditional on attending university)
Average Marginal Effects			
KS4 Math	0.090 (0.001)	0.147 (0.001)	0.218 (0.002)
KS4 Verbal	0.185 (0.001)	0.151 (0.002)	-0.159 (0.003)
KS4 Motive	0.032 (0.001)	-0.008 (0.001)	0.003 (0.001)
Obs.	463141	180830	180830

Note: Results are from a logistic regression with controls for gender, IDACI Index, free school lunch, special education needs, race, mother tongue, and school fixed effects. Bootstrapped standard errors at the school level.

shows results for enrollment in STEM fields conditional on university enrollment.[21] As expected, math skill has a large positive effect on enrolling in STEM fields, while verbal skill has a negative effect.[22]

Of similar interest is the effect of math and verbal skills on college completion. In Table 11 we look at graduation probabilities, splitting the sample by field of study, STEM and non-STEM, because it is likely that skills play different roles in different majors. While Table 10 indicates that STEM majors are heavily concentrated in the upper end of the math skill distribution and the lower end of the verbal skill distribution, the first column in Table 11 shows that for this population, marginal changes in verbal skill has a larger effect on graduation than marginal changes in math skill. Specifically, conditional on enrolling in STEM, increasing verbal skill by one standard deviation increases the graduation rate by 10 percentage points, while increasing math skill by one standard deviation only increases the graduation rate by 3.2 percentage points. This result provides evidence that not only does having higher verbal skill promote college enrollment, but it also increases the probability of college completion even in math intensive STEM fields.

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

[22]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 11: Logistic Regression: Graduation and Key Stage 4 Skills

	Conditional on Enrolled in STEM	Conditional on Enrolled in Non-STEM
Average Marginal Effects		
KS4 Math	0.038 (0.002)	0.004 (0.002)
KS4 Verbal	0.101 (0.003)	0.085 (0.001)
KS4 Motive	0.032 (0.005)	0.028 (0.002)
Obs.	54206	126624

Note: Results are from a logistic regression with controls for gender, IDACI Index, free school lunch, special education needs, race, mother tongue, and school fixed effects. Bootstrapped standard errors at school level.

Collectively, these results demonstrate that verbal skill plays a key role in many educational outcomes. In the next section, we look beyond our data and factor model to provide further evidence to substantiate this claim.

5 Reinforcing Our Main Findings

Our main result shows that verbal skill has a larger effect on university enrollment than math skill. In this section we rule out a number of alternative explanations for this result that we cannot directly address with our data or model. This further validates our findings. First, using supplemental data that contains detailed information on student externalizing behavior,[23] parent background characteristics, and student IQ, we show that our measure of verbal skill is *not* disproportionately serving as a proxy to any of these other important student characteristics. Second, our main analysis does not include any features of the supply-side of the university market, which could exaggerate the importance of verbal skill relative to math skill. 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 a byproduct of university offerings. Third, our data only covers students in the English education

[23]Childhood behaviors characterized by impulsivity, disruptiveness, aggression, antisocial features, and overactivity are called externalizing behavior.

system. It is possible that the relative importance of verbal is strictly a phenomenon among this population. To address this critique, we analyze university enrollment decisions using data from the United States and show that similar patterns to our main findings persist in this data as well. Finally, we conclude this section by showing that our main result holds even under simpler econometric methods, which demonstrates that our main result is not driven by the assumptions of the factor framework.

5.1 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. 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 the multi-period factor model addresses this concern by controlling for special education needs, 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.[24] 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.

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

This database also contains aggregate scores on math and English exams for each Key Stage.[25] 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.[26] Table 12 shows regression outcomes where each coefficient corresponds to a separate regression (in each of them we control for gender).[27] Panel A of Table 12 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.[28]

Similarly, as we have discussed, family background characteristics might disproportionately impact verbal skill more than 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 include the following controls for family background characteristics: free school meal eligibility, race, mother tongue, special education needs, IDACI index (a poverty index), and school attended. Second, we further examine our findings by conditioning on early Key Stage skills, which perhaps serve as sufficient statistics for any unobserved family background characteristic.[29] 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

[25] 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.

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

[27] 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.

[28] 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).

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

Table 12: 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.

characteristics. Panel B of Table 12 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 12 shows the interrelation between the Wechsler Intelligence Scale for Children (WISC) IQ test and KS2 exams.[30] 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.

5.2 Can University Supply Explain the Relative Importance of Verbal Skill?

It is possible 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.[31] 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.[32] Using the same method, but broadening our definition to include programs

[30]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.

[31]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.

[32]The remaining 55.5% enrolled in programs with no science requirement, with 31% having a non-science requirement, and 24.5% having no requirement.

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.

5.3 Is the Larger Effect of Verbal Skill on University Enrollment a Specific Phenomenon of the UK?

To assess the possibility that our main result 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 on 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 contains 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 thats 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 last interview was less than 21 where 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 skill appears to be substantially more important than verbal skill 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.[33] 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 declines and the importance of math 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 to the char-

[33] 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.

acteristics of the US education system) that is 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.[34]

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 7. 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.[35] This analysis leaves out students who took Algebra I in the eighth grade (high-achieving students) and students taking Pre-Algebra in the ninth grade (low-achieving students).[36] Column (7) of

[34]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.

[35]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."

[36]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.

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 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 skill in explaining college enrollment and graduation does not appear to be unique to the UK education system.

5.4 Beyond the Factor Model Assumptions: Instrumental Variables Approach

Given the questions that we are studying (i.e. the role of skills in explaining educational outcomes, and the evolution of skills during compulsory education), and the large number of test score outcomes that are available in our database, a factor model approach constitutes the most natural empirical strategy to follow. However, we also explore an alternative strategy to show that our results are not driven by any of the normalizations or mild assumptions that we impose in the factor model. In this regard, we implement a simple instrumental variables approach that not only corrects for measurement error in test scores but also deals with possible endogeneity driven by student motivation. The specific concern is that the decision to attend university has already been made in the last year of high school, and therefore it affects exam performance.

Columns (1) to (4) of Table 14 show IV regression results where the dependent variable is university enrollment, and the independent variables math and English performance in Key Stage 4 are instrumented (depending on the specification) with math and English performance in Key Stage 1, 2, or 3. Column (2) of Table 14 shows the baseline (OLS) results where we only control for family background and school characteristics (i.e. free school meal eligibility, special education needs, IDACI index, race, gender, mother tongue, and school fixed effects). Columns (2) to (4) present IV regression results where math and English performance in KS4 are instrumented by their analogues in KS1 to KS3, respectively. Overall, the results in these columns are largely consistent with the findings of the factor model. Regardless of how we instrument performance in KS4, English has a substantially larger effect on university enrollment than math skills.

Table 14: IV - University Enrollment

	Baseline KS4	IV - KS1	IV - KS2	IV - KS3
KS4 Overall Math Test Score	0.139	0.108	0.079	0.083
KS4 Overall Verbal Test Score	0.162	0.198	0.253	0.264
R-Squared	0.378	0.375	0.367	0.367

Note: Math and verbal test scores have been standardized to have mean zero and standard deviation 1. Standard errors not shown. All coefficients are statistically significant at the 95% level. Earlier Key Stage test scores are no available for all students so those with missing values where dropped from the analysis. Specifications include controls for free school meal eligibility, special education needs, mother tongue, race, and IDACI score.

6 Lessons from the Main Findings

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).[37] 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. 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 15 shows the impact of the scores in paragraph comprehension (PC)

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

Table 15: 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 Bachelors degree or more at their most recent interview. Those whose age at last interview was less than 25 where 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. 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 skill suffer a substantially larger drop (i.e. from 0.0253 to 0.0031) than math skill (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 15, may significantly understate the importance of verbal skills for labor market outcomes.

Columns 4 and 5 in Table 15 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 a ninth grade English. These results are consistent with our findings for the ASVAB subject tests, where the return to the verbal skill 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).

7 Timing of Skill Development

We now study how math and verbal skills evolve over the course of a student's compulsory education career. Table 4 presents the estimated correlation matrix of these nine factors, which shows they are highly correlated. It is likely that a large portion of the KS4 math and verbal factors are determined from outcomes in KS1 to KS3. To formalize the analysis, we return to the tools used in Eq. (3), which shows that current skills can be decomposed as the summation of the expected value given realizations of earlier skills and deviations from the expected value. From our model we can empirically compute how likely deviations from the expectations occur in the data. For example, we can compute $\text{Var}(\hat{\theta}_{i4|1,0} - \hat{\theta}_{i4|0})$, which describes the extent to which the events of KS1 affect the expected skills in KS4, conditional on background characteristics.[38] This is an empirical measure in the population of the variance in KS4 skills that can be attributed to outcomes in KS1.

Table 16 computes the contribution of each Key Stage shock in explaining the total variance of KS4 skills, measured in terms of R-squared. In addition, this table also lists the computed probability of receiving a skill shock that would move a student up 0.5 standard deviation units (or more) in the population skill distribution. The first row of the table shows the contribution of background characteristics (IDACI index, free school meal eligibility, special education needs, elementary school attended, mother tongue, race and gender) in explaining the total variation (in terms of R-squared) in each of the KS4 factors. As expected, these elements have large explanatory power, on the order of 46% in math and verbal skills and 50% in motivation. Moreover, since these family background variables show similar predictive power in math and verbal skills, it suggests that these skills are similarly shaped at home. Moving forward, the outcome in KS1 adds an additional 14% and 13% to the explained variance of KS4 math and verbal skills, respectively. Therefore, approximately 60% of the variance in both skills at age 16 is determined by outcomes occurring before age 8. This result highlights a key point. While a large portion of skill variation in KS4 is already determined at an early age, there remains a large amount (40%) to be explained at later stages, suggesting some scope for overcoming initial disadvantages. However, by the end

[38] Here we are quantifying to what extent the new information that arrives in each Key Stage contributes to explaining the total variability of skills by the end of compulsory schooling.

Table 16: Development of Key Stage 4 Skills

	Contribution to R-squared			Probability of Receiving Shock to Move Up 0.5 S.D. in the Population Distribution		
	Math	Verbal	Motive	Math	Verbal	Motive
Background Characteristics	0.465 (0.002)	0.466 (0.002)	0.501 (0.003)	0.232 (0.000)	0.232 (0.000)	0.240 (0.001)
Key Stage 1 Shocks	0.140 (0.001)	0.130 (0.001)	0.067 (0.002)	0.091 (0.001)	0.082 (0.001)	0.027 (0.001)
Key Stage 2 Shocks	0.173 (0.002)	0.160 (0.001)	0.081 (0.002)	0.114 (0.001)	0.106 (0.001)	0.039 (0.002)
Key Stage 3 Shocks	0.209 (0.001)	0.149 (0.002)	0.154 (0.004)	0.137 (0.001)	0.098 (0.001)	0.101 (0.003)
Key Stage 4 Shocks	0.040 (0.001)	0.117 (0.001)	0.211 (0.003)	0.006 (0.000)	0.072 (0.001)	0.138 (0.002)
Total	1.00	1.00	1.00	–	–	–

Note: Key Stage 4 shocks are defined as $\theta_{i4} - \mathrm{E}(\theta_{i4}|\theta_{i3}, \theta_{i2}, \theta_{i1}, x_i)$. Key Stage 3 shocks are defined as $\mathrm{E}(\theta_{i4}|\theta_{i3}, \theta_{i2}, \theta_{i1}, x_i) - \mathrm{E}(\theta_{i4}|\theta_{i2}, \theta_{i1}, x_i)$. The other Key Stage shocks are defined similarly.

of KS3 (age 14), around 96% (88%) of the variation in KS4 math (verbal) skills can be explained by KS1-KS3 outcomes and family background characteristics.

The last three columns of Table 16 indicate the likelihood of a student receiving a skill shock that would move them up 0.5 standard deviation units in the aggregate skill distribution at each Key Stage. For example, the probability that a student moves up 0.5 standard deviations in her KS4 math skill given a positive shock in KS3 is 0.13, which would require receiving a positive shock that is more than 1 standard deviation larger than the mean shock in that period. Since Table 7 reports the average marginal effect for a one standard deviation increase in KS4 skills, the impact of this shock would be a 4.5 percentage point increase in the probability of college enrollment.

In summary, this section highlights two main points. First, around half of the variation in skills at the end of compulsory education is explained by "new information" that is arriving between KS1 and KS4, suggesting that a sizable portion of skills is determined at later stages. Second, the probability of receiving a sufficiently positive shock at any given key stage is relatively small.[39]

While the inclusion of the motivation factor into our analysis was largely to correct for endogeneity, the results on motivation in Table 16 are interesting on their own. The entirety of the motivation factor is either explained by background characteristics or events occurring towards the end of compulsory education. One of the measurements used to extract motivation was the number of subject-specific exams taken in Key Stage 4. The results in this table suggest that the events that occur during Key Stage 1, after conditioning on background characteristics, have virtually no predictive power on the number of subject tests taken in Key Stage 4. Therefore, if the KS4 skills are thought to be endogenous with university enrollment, the results from motivation in this table validate the use of these early skills as exogenous regressors, which was assumed in the analysis in Table 9.

8 Understanding the Gender Gap in College Enrollment and STEM Major

Section 4.1 shows that, while there are small differences in math skill between genders, females have a large advantage in verbal skill. This empirical regularity, combined with our findings on the importance of verbal skill for college enrollment, suggests a possible explanation for the gender gap in college enrollment and other college outcomes. Table 17 studies whether the gender gap in college enrollment and STEM fields can be explained by differences in skills. Column (1) shows that females are 4.4% more likely to attend university than males. This gap is smaller than the 6.6% reported earlier because this specification controls for special educational needs.[40] Column (2) shows that, after controlling for skills in KS4, females are slightly less likely to attend university than males, suggesting that gender differences in verbal skill play a key role in explaining the gender gap in college enrollment. Columns (3) and (4) analyze whether skills have a differential effect on college enrollment once conditioning on gender. These columns show results that are based on the

[39] In order to provide a robustness check, Appendix G performs a similar decomposition for KS3. Results show a similar pattern.

[40] Males are substantially more likely to be assigned into this category.

Table 17: Logistic Regression: Gender, University Outcomes and Key Stage 4 Skills

	University Enrollment				Enrollment in STEM (conditional on enrollment)			
	All		Male	Female	All		Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average Marginal Effects								
Female	0.044 (0.001)	-0.008 (0.001)	–	–	-0.171 (0.003)	-0.105 (0.002)	–	–
KS4 Math	–	0.096 (0.001)	0.099 (0.002)	0.095 (0.002)	–	0.218 (0.002)	0.273 (0.003)	0.167 (0.004)
KS4 Verbal	–	0.187 (0.001)	0.179 (0.000)	0.200 (0.002)	–	-0.159 (0.003)	-0.225 (0.002)	-0.097 (0.004)
KS4 Motive	–	0.029 (0.001)	0.024 (0.001)	0.031 (0.001)	–	0.003 (0.001)	-0.002 (0.003)	0.005 (0.001)

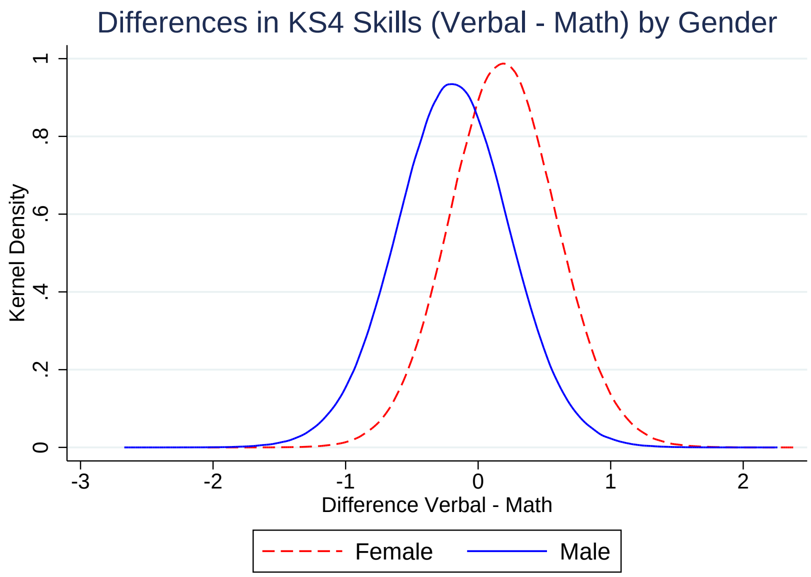
Notes: Results are from a logistic regression with controls for IDACI Index, free school lunch, race, mother tongue, and special education needs. Bootstrapped standard errors at school level.

male and female sample separately. Differences in the average marginal effects between Columns (3) and (4) are small, suggesting that math and verbal skills have a consistent effect on males and females in the decision to enroll in college.

We also investigate the gender gap in STEM fields, which favors males. Columns (5) to (8) analyze enrollment in STEM fields conditional on enrolling in college. Column (5) shows that, after controlling for family background characteristics (IDACI Index, free school lunch, race, mother tongue, and special education needs), females are 17.1 percentage points less likely to enroll in STEM fields. Column (6) adds controls for skills to the regression. The large positive effect on math skill and large negative effect on verbal skills in these regressions indicate that comparative advantage in math skill is critical for the STEM enrollment decision. Figure 3 plots the distribution of skill differences for males and females, which shows males are more likely to perform better (in percentile terms) in math than in verbal, while for females the opposite is true. In fact, 65% of males have a comparative advantage in math compared to 35% for females.[41] Consequently, accounting for skill differences by gender reduces the raw gender gap in STEM enrollment 40%, down to 10.5 percentage points.

[41] Comparative advantage in math is defined as the math skill percentile being larger than the verbal skill percentile.

Figure 3: Differences in Skills



Finally, Columns (7) and (8) of Table 17 perform separate regressions on STEM enrollment by gender to analyze if skills have a differential effect between genders when deciding to enroll in STEM. Comparing Column (7), the regression for males, and Column (8), the regression for females, indicates that males are much more responsive to skills than females. Interestingly, the stronger responsiveness of males to skills occurs in both directions. For example, a one standard deviation increase in math skill will lead to a 27 percentage point increase in STEM enrollment for males, while a similar increase will only increase female enrollment by 16 percentage points. However, looking at the effect of verbal skill, a one standard deviation increase in verbal skills will lead to a 22 percentage point *drop* in STEM enrollment for males, while a similar increase in verbal skill only generates a 10 percentage point drop in STEM enrollment for females. To conclude, these results suggest that males are more sensitive to the differential level of skills than females when deciding to enroll in STEM fields, indicating that skill comparative advantage seems to be substantially more important for males in their decision to enroll in scientific fields.

9 Conclusion

This paper estimates a multi-period factor model of skills that uses administrative data that follows a cohort of students in England from elementary school to university to assess the impact of math and verbal skills on schooling outcomes. First, we show that the effect of verbal skill on university enrollment is almost twice as large as the effect of math skill. Second, our empirical strategy contributes to understanding how math and verbal skills evolve during the years of compulsory education. We show that 60% of the variance in skills measured at age 16 can be attributed to outcomes occurring before age 7 and 40% of the variance can be attributed to outcomes after age 7, suggesting some scope for overcoming initial skills disadvantages. Third, we document that females have a large advantage in verbal skill relative to males, which explains the gender gap in college enrollment. We find that, after controlling for math and verbal skills, females are slightly *less* likely to attend college than males. On the other hand, males' comparative advantage in math skill over verbal skill contributes to increase male representation in STEM majors.

Our finding on the predominant effect of verbal skill on college enrollment suggests that the role of this skill 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 skill 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 skill affects wages, that is increasing years of schooling.

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 and D. Autor. Skills, tasks and technologies: Implications for employment and earnings. *Handbook of labor economics*, 4:1043–1171, 2011.

M. Almlund, A. L. Duckworth, J. Heckman, and T. Kautz. Chapter 1 - personality psychology and economics. In S. M. Eric A. Hanushek and L. Woessmann, editors, *Handbook of The Economics of Education*, volume 4 of *Handbook of the Economics of Education*, pages 1 – 181. Elsevier, 2011. doi: http://dx.doi.org/10.1016/B978-0-444-53444-6.00001-8.

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.

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.

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.

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.

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. 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.

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.

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

P. A. Ruud. Extensions of estimation methods using the EM algorithm. *Journal of Econometrics*, 49(3):305–341, 1991.

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.

R. Varadhan and C. Roland. Simple and globally convergent methods for accelerating the convergence of any EM algorithm. *Scandinavian Journal of Statistics*, 35(2):335–353, 2008.

Appendix

A Mapping the joint factor model to a dynamic factor model

In our factor specification, we allow the factor error to be correlated across time periods. An alternative specification would allow the factors to be a function of the earlier factors and model the factor error as independent (once we condition on the earlier factors). This alternative approach is commonly used in dynamic factor models. This section demonstrates that for any value of the parameters, the factor structure we use in Eq. (1) maps directly to a dynamic factor model.

In a dynamic factor model, the current period factors are written as a linear function of previous factor realizations and an additive, independent error term. For example, the factors in KS3 would be written as

$$\theta_{i3} = female_i * \widetilde{\psi}_3 + \widetilde{\Phi}_3 x_i + \Omega_3 \begin{bmatrix} \theta'_{i1} & \theta'_{i2} \end{bmatrix}' + \widetilde{\xi}_{i3} \tag{4}$$

The tilde's represent that the parameters in this specification perform a similar role to the ones in Eq. (1) but are not identical. The additional variable Ω_3 is a weighting matrix that describes how KS1 and KS2 factors influence KS3 factors. Finally, the error term, $\widetilde{\xi}_{i3} \sim \mathcal{N}(0, \widetilde{\Sigma}_3)$.

We now show how the parameters $\widetilde{\psi}_3$, $\widetilde{\Phi}_3$, Ω_3, and $\widetilde{\Sigma}_3$ in Eq. (4) map into the parameters in Eq. (1): ψ, Φ, and Σ. First, from Eq. (4),

$$\theta_{i3} \sim \mathcal{N}\left(female_i * \widetilde{\psi}_3 + \widetilde{\Phi}_3 x_i + \Omega_3 \begin{bmatrix} \theta'_{i1} & \theta'_{i2} \end{bmatrix}', \widetilde{\Sigma}_3 \right)$$

To map this model to the parameters in Eq. (1), let $\psi_{[5:6]}$ denote the fifth to sixth element of the vector ψ, which represents the effect of gender on the KS3 math and verbal skills. Similarly, let $\Phi_{[5:6,\ ]}$ denote the fifth and sixth rows of the matrix Φ. Using the knowledge of θ_{i1} and θ_{i3} we can compute the moments of the marginal distribution

$$\begin{aligned} \mathrm{E}(\theta_{i3}|\theta_{i1}, \theta_{i2}) = & female_i * \psi_{[5:6]} + \Phi_{[5:6,\ ]} x_i \\ & + \Sigma_{[5:6,1:4]} \Sigma^{-1}_{[1:4,1:4]} \left[\begin{array}{c} \theta_{i1} \\ \theta_{i2} \end{array} - female_i * \psi_{[1:4]} + \Phi_{[1:4,\ ]} x_i \right] \end{aligned} \tag{5}$$

$$\mathrm{Var}(\theta_{i3}|\theta_{i1}, \theta_{i2}) = \underbrace{\Sigma_{[5:6,5:6]} - \Sigma_{[5:6,1:4]} \Sigma^{-1}_{[1:4,1:4]} \Sigma_{[1:4,5:6]}}_{\widetilde{\Sigma}_3}$$

This defines $\widetilde{\Sigma}_3$. We can define the rest of the dynamic factor parameters in Eq. (4) by re-arranging terms in Eq. (5): $\Omega_3 = \Sigma_{[5:6,1:4]} \Sigma^{-1}_{[1:4,1:4]}$, $\widetilde{\psi}_3 = \psi_{[5:6]} - \Omega_3 \psi_{[1:4]}$, and $\widetilde{\Phi}_3 = \Phi_{[5:6,\ ]} - \Omega_3 \Phi_{[1:4,\ ]}$. To complete the mapping to a dynamic factor model, we can apply similar ideas to re-define all of the skills in KS1, KS2 and KS4.

B Normalizations

To identify the model, we need to make three sets of assumptions. First, the mean of the measurement equation is not separately identified from the mean of the factors. We choose to estimate μ_m in each of the measurement equations and set the mean of the factors to zero. To do this, the factor equation (Eq. (1)) does not include a constant and one of the mixture-means is set to a vector of zeros, i.e. $\delta_1 = 0$. Second, the scale of the nine factors is not separately identified from all of the factor loadings. We choose to estimate the variance of the factors and normalize at least one factor loading for each of the factors to 1. Third, the full covariance is not identified if all of the measurements load in multiple factors. To estimate the covariance, we restrict some of the measurements to only load on a single factor. Table 18 provides a full list of our measurements and the normalizations we make on the factor loadings. The cells labeled "na" in the table indicate that the loading is estimated.

C Estimation Algorithm for Factor Model

This section outlines the iterative algorithm used to estimate the factor model in the first stage. Estimating these parameters requires maximizing Eq. (2). The vector of unobserved factors θ_i is of dimension $J \times 1$ and is a function of observed characteristics, x_i, and the unobserved factor error ξ_i. The functional form for the factors is as follows:

$$\theta_i = \begin{bmatrix} \theta_i(1) \\ \vdots \\ \theta_i(J) \end{bmatrix} = [\phi_1 \ \cdots \phi_J]^\top x_i + \begin{bmatrix} \xi_i(1) \\ \vdots \\ \xi_i(J) \end{bmatrix} = \Phi^\top x_i + \xi_i$$

Where ϕ_j is a vector describing how observed characteristics x_i influence factor j.

In this section, we will focus on the case where the unobserved error structure is distributed multivariate normal $\xi_i \sim \mathcal{N}(0, \Sigma)$. In our main specification we allow this to follow a mixture of normals. For each individual $i = 1, 2, \ldots N$ we observe M measures that contain information about the unobserved factors. We assume here for simplicity that all measures are observed for all individuals. The measures are determined by.

$$w_i = \begin{bmatrix} w_i(1) \\ \vdots \\ w_i(M) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_M \end{bmatrix} + [\lambda_1 \cdots \lambda_M]^\top \theta_i + \begin{bmatrix} \eta_i(1) \\ \vdots \\ \eta_i(M) \end{bmatrix} = \mathrm{M} + \Lambda^\top \theta_i + \eta_i$$

Where μ_m represents the mean of measurement m and $\mathrm{M} = [\mu_1, \mu_2, \cdots, \mu_M]^\top$. All of the measure specific means are identified because the mean of ξ is normalized to zero. λ_m are the measure specific loadings, where the appropriate loadings have been normalized so that the full covariance of the factor error, Σ is identified. Finally, $\eta_i(m)$ is the error in measurement m that is not explained by the factor. Since the factors are correlated, we assume the errors in the measurement are not correlated and $\eta_i(m) \sim \mathcal{N}(0, \tau_m^2)$. Let T denote an $M \times M$ matrix with τ_m^2 as the m^{th} diagonal element, with zeros for all off diagonal elements, then $\eta_i \sim \mathcal{N}(0, \mathrm{T})$.

Using the current iterations estimates of the parameters, $\Psi^{(r)} = [\Phi^{(r)}, \Sigma^{(r)}, \mathrm{M}^{(r)}, \Lambda^{(r)}, \mathrm{T}^{(r)}]$, the EM algorithm finds new parameters $\Psi^{(r+1)}$ and iterates until $|(LL(\Psi^{(r+1)}) - LL(\Psi^{(r)}))/LL(\Psi^{(r)}| <$

Table 18: List of measurement variables and normalizations of factor loadings

No.	Key Stage	Description	KS1 Math	KS1 Verbal	KS2 Math	KS2 Verbal	KS3 Math	KS3 Verbal	KS4 Math	KS4 Verbal	KS4 Motive
1	1	Math Test	1	0	0	0	0	0	0	0	0
2	1	Math Using and Applying TA	na	0	0	0	0	0	0	0	0
3	1	Math Number and Algebra TA	na	0	0	0	0	0	0	0	0
4	1	Math Shapes and Measure TA	na	0	0	0	0	0	0	0	0
5	1	Writing Test	0	1	0	0	0	0	0	0	0
6	1	Writing TA	0	na	0	0	0	0	0	0	0
7	1	Reading TA	0	na	0	0	0	0	0	0	0
8	1	Listening TA	0	na	0	0	0	0	0	0	0
9	2	Math Test Paper A	0	0	1	0	0	0	0	0	0
10	2	Math Test Paper B	0	0	na	0	0	0	0	0	0
11	2	Math Arithmetic Test	0	0	na	0	0	0	0	0	0
12	2	Math TA	0	0	na	0	0	0	0	0	0
13	2	Reading Test	0	0	0	1	0	0	0	0	0
14	2	Writing Test	0	0	0	na	0	0	0	0	0
15	2	Spelling Test	0	0	0	na	0	0	0	0	0
16	2	English TA	0	0	0	na	0	0	0	0	0
17	3	Math Test Paper 1	0	0	0	0	na	0	0	0	0
18	3	Math Test Paper 2	0	0	0	0	na	0	0	0	0
19	3	Math Arithmetic Test	0	0	0	0	1	0	0	0	0
20	3	Math TA	0	0	0	0	na	0	0	0	0
21	3	Writing Test (Longer)	0	0	0	0	0	1	0	0	0
22	3	Reading Test	0	0	0	0	0	na	0	0	0
23	3	Writing Test (Shorter)	0	0	0	0	0	na	0	0	0
24	3	Reading Test (Shakespeare)	0	0	0	0	0	na	0	0	0
25	3	English TA	0	0	0	0	0	na	0	0	0
26	4	Math	0	0	0	0	0	0	1	0	0
27	4	English	0	0	0	0	0	0	0	1	0
28	4	Design and Technology: Graphic Products	0	0	0	0	0	0	na	na	0
29	4	Design and Technology: Resistant Materials Technology	0	0	0	0	0	0	na	na	0
30	4	Design and Technology: Textiles Technology	0	0	0	0	0	0	na	na	0
31	4	Art and Design	0	0	0	0	0	0	na	na	0
32	4	History	0	0	0	0	0	0	0	na	0
33	4	Geography	0	0	0	0	0	0	na	na	0
34	4	French	0	0	0	0	0	0	na	na	0
35	4	German	0	0	0	0	0	0	na	na	0
36	4	Business Studies	0	0	0	0	0	0	na	na	0
37	4	Religious Studies	0	0	0	0	0	0	0	na	0
38	4	Short Religious Studies	0	0	0	0	0	0	0	na	0
39	4	Physical Education	0	0	0	0	0	0	na	na	0
40	4	Physics	0	0	0	0	0	0	na	0	0
41	4	Chemistry	0	0	0	0	0	0	na	0	0
42	4	Biology	0	0	0	0	0	0	na	na	0
43	4	Drama	0	0	0	0	0	0	na	na	0
44	4	Information Technology	0	0	0	0	0	0	na	na	0
45	4	Short Information Technology	0	0	0	0	0	0	na	na	0
46	4	Spanish	0	0	0	0	0	0	na	na	0
47	4	Music	0	0	0	0	0	0	na	na	0
48	4	Social Science	0	0	0	0	0	0	na	na	0
49	4	Design and Technology: Electronic Products	0	0	0	0	0	0	na	na	0
50	4	Design and Technology: System and Control	0	0	0	0	0	0	na	na	0
51	4	English Literature	0	0	0	0	0	0	0	na	0
52	4	Design and Technology: Food Technology	0	0	0	0	0	0	na	na	0
53	4	Science	0	0	0	0	0	0	na	na	0
54	4	Statistics	0	0	0	0	0	0	na	na	0
55	4	Medial, Film and Television Studies	0	0	0	0	0	0	na	na	0
56	4	Fine Art	0	0	0	0	0	0	na	na	0
57	4	Office Technology	0	0	0	0	0	0	na	na	0
58	4	Home Economics: Child Development	0	0	0	0	0	0	na	na	0
59	4	Italian	0	0	0	0	0	0	na	na	0
60	4	Urdu	0	0	0	0	0	0	na	na	0
61	4	Additional Applied Science	0	0	0	0	0	0	na	na	0
62	4	Leisure and Tourism	0	0	0	0	0	0	na	na	0
63	4	Applied ICT	0	0	0	0	0	0	na	na	0
64	4	Applied Science	0	0	0	0	0	0	na	na	0
65	4	Health and Social Care	0	0	0	0	0	0	na	na	0
66	4	Applied Business	0	0	0	0	0	0	na	na	0
67	4	Double Science	0	0	0	0	0	0	na	0	0
68	4	Total GCSE Exams Taken	0	0	0	0	0	0	0	0	1
69	4	Authorize Absences	0	0	0	0	0	0	0	0	na
70	4	Unauthorized Absences	0	0	0	0	0	0	0	0	na

Note: The normalizations are either 0 or 1. Na denotes the loading is estimated. TA denotes teacher assessment.

$1E-8$.

The iteration proceeds as follows.

Using $\Psi^{(r)}$, compute, $\mathrm{K}^{(r)} = \Sigma^{(r)}\Lambda^{(r)}\left[\Lambda^{(r)\top}\Sigma^{(r)}\Lambda^{(r)} + \mathrm{T}^{(r)}\right]^{-1}$

For each individual i, compute the following expected values

$$\mathbf{e}_i^{(r)} = \mathrm{E}(\theta_i|\Psi^{(r)}, x_i, w_i) = \Phi^{(r)\top}x_i + \mathrm{K}^{(r)}\left[w_i - \mathrm{M}^{(r)} - \Lambda^{(r)\top}\Phi^{(r)\top}x_i\right] \quad \text{(first moment)}$$

$$\mathbf{v}_i^{(r)} = \mathrm{E}(\theta_i\theta_i^\top|\Psi^{(r)}, x_i, w_i) = \left[\mathrm{I} - \mathrm{K}^{(r)}\Lambda^{(r)\top}\right]\Sigma^{(r)} + \mathbf{e}_i^{(r)}\mathbf{e}_i^{(r)\top} \quad \text{(uncentered second moment)}$$

Where I is the identity matrix that is $M \times M$.

The new parameters are found by computing

$$\phi_j^{(r+1)} = \left[\sum_{i=1}^N x_i x_i^\top\right]^{-1}\left[\sum_{i=1}^N x_i \mathbf{e}_i^{(r)}(j)\right]$$

$$\Sigma^{(r+1)} = \frac{1}{N}\sum_{i=1}^N \left[\mathbf{v}_i^{(r)} - \Phi^{(r+1)\top}x_i x_i^\top \Phi^{(r+1)}\right]$$

$$\mu_m^{(r+1)} = \frac{1}{N}\sum_{i=1}^N \left[w_i(m) - \lambda_m^{(r+1)\top}\mathbf{e}_i^{(r)}\right]$$

Given that λ^m is the $J \times 1$ vector of factor loadings for measure m, let ι_m be a $J \times 1$ indicator vector of the loadings which are not constrained (and must be estimated in vector λ_m, and $-\iota_m$ be the indicator vector for the factors that are normalized. This means we only need to update $\lambda_{m\{\iota_m\}}$. Letting $\widetilde{w}_i = w_i - \mathrm{M}^{(r+1)}$, the equation to update the factor loadings is,

$$\lambda_{m\{\iota_m\}}^{(r+1)} = \left[\sum_{i=1}^N \mathbf{v}_{i\{\iota_m,\iota_m\}}^{(r)}\right]^{-1}\left[\sum_{i=1}^N \left(\mathbf{e}_{i\{\iota_m\}}^{(r)}\widetilde{w}_i(m) - \mathbf{v}_{i\{\iota_m,-\iota_m\}}^{(r)}\lambda_{m\{-\iota_m\}}^{(r)}\right)\right]$$

$$\tau_m^{(r+1)} = \frac{1}{N}\sum_{i=1}^N \left(\widetilde{w}_i(m)^2 - 2\widetilde{w}_i(m)\lambda_m^{(r+1)\top}\mathbf{e}_i^{(r)} + \lambda_m^{(r+1)\top}\mathbf{v}_i^{(r)}\lambda_m^{(r+1)}\right)$$

D Factor Coefficients

Table 19: Factor Coefficients

	KS1 Math	KS1 Verbal	KS2 Math	KS2 Verbal	KS3 Math	KS3 Verbal	KS4 Math	KS4 Verbal	KS4 Motive
Demographics									
female	-0.118	0.159	-0.222	0.116	-0.124	0.232	-0.083	0.274	0.096
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.005)
SEN	-0.939	-1.104	-1.069	-1.190	-0.838	-0.916	-0.837	-0.849	-1.075
	(0.004)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.009)
FSM	-0.185	-0.233	-0.205	-0.224	-0.210	-0.251	-0.336	-0.347	-0.700
	(0.003)	(0.002)	(0.004)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.008)
IDACI	-0.399	-0.468	-0.470	-0.517	-0.514	-0.635	-0.784	-0.818	-1.371
	(0.012)	(0.014)	(0.010)	(0.008)	(0.009)	(0.010)	(0.009)	(0.013)	(0.017)
Elementary School Fixed Effects									
Std. of 15,353 FE	0.323	0.314	0.281	0.274	0.230	0.276	0.303	0.303	0.661
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.005)
Distribution of Unobservables									
Type 1 (prob type 1 = 0.24)	0	0	0	0	0	0	0	0	0
Type 2 (prob type 2 = 0.76)	0.531	0.449	1.177	0.633	0.758	0.475	0.659	0.501	0.527
	(0.004)	(0.006)	(0.004)	(0.004)	(0.003)	(0.005)	(0.003)	(0.006)	(0.017)
Std(ξ)	0.663	0.649	0.572	0.595	0.517	0.597	0.692	0.687	1.009
	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.006)

Note: These estimates correspond to the main coefficients of Eq. (1) (i.e. skill coefficients)

E Factor Loadings

Table 20: Factor Loadings

No.	Key Stage	Description	KS1 Math	KS1 Verbal	KS2 Math	KS2 Verbal	KS3 Math	KS3 Verbal	KS4 Math	KS4 Verbal	KS4 Motive	Residual Variance
1	1	Math Test	$1^\dagger$	0	0	0	0	0	0	0	0	0.224 (0.002)
2	1	Math Using and Applying TA	0.985 (0.002)	0	0	0	0	0	0	0	0	0.246 (0.002)
3	1	Math Number and Algebra TA	1.013 (0.002)	0	0	0	0	0	0	0	0	0.202 (0.001)
4	1	Math Shapes and Measure TA	0.992 (0.004)	0	0	0	0	0	0	0	0	0.235 (0.002)
5	1	Writing Test	0	$1^\dagger$	0	0	0	0	0	0	0	0.172 (0.001)
6	1	Writing TA	0	0.962 (0.002)	0	0	0	0	0	0	0	0.231 (0.001)
7	1	Reading TA	0	0.937 (0.002)	0	0	0	0	0	0	0	0.271 (0.001)
8	1	Listening TA	0	0.827 (0.003)	0	0	0	0	0	0	0	0.431 (0.001)
9	2	Math Test Paper A	0	0	$1^\dagger$	0	0	0	0	0	0	0.126 (0.000)
10	2	Math Test Paper B	0	0	0.981 (0.001)	0	0	0	0	0	0	0.158 (0.000)
11	2	Math Arithmetic Test	0	0	0.960 (0.001)	0	0	0	0	0	0	0.194 (0.001)
12	2	Math TA	0	0	0.908 (0.001)	0	0	0	0	0	0	0.237 (0.002)
13	2	Reading Test	0	0	0	$1^\dagger$	0	0	0	0	0	0.250 (0.001)
14	2	Writing Test	0	0	0	0.878 (0.002)	0	0	0	0	0	0.421 (0.001)
15	2	Spelling Test	0	0	0	0.876 (0.001)	0	0	0	0	0	0.424 (0.002)
16	2	English TA	0	0	0	0.972 (0.002)	0	0	0	0	0	0.233 (0.001)
17	3	Math Test Paper 1	0	0	0	0	0.984 (0.002)	0	0	0	0	0.423 (0.001)
18	3	Math Test Paper 2	0	0	0	0	0.996 (0.003)	0	0	0	0	0.409 (0.001)
19	3	Math Arithmetic Test	0	0	0	0	$1^\dagger$	0	0	0	0	0.406 (0.001)
20	3	Math TA	0	0	0	0	1.189 (0.001)	0	0	0	0	0.147 (0.001)
21	3	Writing Test (Longer)	0	0	0	0	0	$1^\dagger$	0	0	0	0.343 (0.002)
22	3	Reading Test	0	0	0	0	0	1.061 (0.002)	0	0	0	0.259 (0.001)
23	3	Writing Test (Shorter)	0	0	0	0	0	1.016 (0.001)	0	0	0	0.322 (0.002)
24	3	Reading Test (Shakespeare)	0	0	0	0	0	0.932 (0.003)	0	0	0	0.429 (0.002)
25	3	English TA	0	0	0	0	0	0.999 (0.002)	0	0	0	0.295 (0.001)
26	4	Math	0	0	0	0	0	0	$1^\dagger$	0	0	0.127 (0.001)
27	4	English	0	0	0	0	0	0	0	$1^\dagger$	0	0.132 (0.001)
28	4	Design and Technology: Graphic Products	0	0	0	0	0	0	0.301 (0.006)	0.579 (0.007)	0	0.477 (0.006)

(*Continued on next page*)

Note: TA denotes teacher assessment.

Table 20: Factor Loadings

No.	Key Stage	Description	KS1 Math	KS1 Verbal	KS2 Math	KS2 Verbal	KS3 Math	KS3 Verbal	KS4 Math	KS4 Verbal	KS4 Motive	Residual Vari- ance
29	4	Design and Technology: Resistant Materials Technology	0	0	0	0	0	0	0.396 (0.007)	0.429 (0.007)	0	0.515 (0.004)
30	4	Design and Technology: Textiles Technology	0	0	0	0	0	0	0.379 (0.008)	0.557 (0.009)	0	0.415 (0.002)
31	4	Art and Design	0	0	0	0	0	0	0.064 (0.006)	0.711 (0.005)	0	0.530 (0.003)
32	4	History	0	0	0	0	0	0	0	1.078 (0.002)	0	0.245 (0.001)
33	4	Geography	0	0	0	0	0	0	0.508 (0.006)	0.593 (0.006)	0	0.219 (0.001)
34	4	French	0	0	0	0	0	0	0.320 (0.006)	0.766 (0.006)	0	0.322 (0.002)
35	4	German	0	0	0	0	0	0	0.392 (0.007)	0.720 (0.009)	0	0.358 (0.003)
36	4	Business Studies	0	0	0	0	0	0	0.405 (0.009)	0.700 (0.006)	0	0.322 (0.001)
37	4	Religious Studies	0	0	0	0	0	0	0	1.052 (0.004)	0	0.281 (0.002)
38	4	Short Religious Studies	0	0	0	0	0	0	0	0.970 (0.002)	0	0.337 (0.002)
39	4	Physical Education	0	0	0	0	0	0	0.593 (0.008)	0.342 (0.008)	0	0.469 (0.002)
40	4	Physics	0	0	0	0	0	0	1.568 (0.007)	0	0	0.194 (0.003)
41	4	Chemistry	0	0	0	0	0	0	1.548 (0.007)	0	0	0.198 (0.004)
42	4	Biology	0	0	0	0	0	0	1.102 (0.011)	0.241 (0.008)	0	0.187 (0.003)
43	4	Drama	0	0	0	0	0	0	-0.005 (0.007)	0.860 (0.007)	0	0.478 (0.003)
44	4	Information Technology	0	0	0	0	0	0	0.423 (0.013)	0.503 (0.010)	0	0.411 (0.004)
45	4	Short Information Technology	0	0	0	0	0	0	0.423 (0.008)	0.421 (0.006)	0	0.461 (0.003)
46	4	Spanish	0	0	0	0	0	0	0.295 (0.012)	0.781 (0.015)	0	0.363 (0.004)
47	4	Music	0	0	0	0	0	0	0.298 (0.008)	0.584 (0.008)	0	0.465 (0.004)
48	4	Social Science	0	0	0	0	0	0	0.088 (0.012)	0.864 (0.010)	0	0.369 (0.003)
49	4	Design and Technology: Electronic Products	0	0	0	0	0	0	0.528 (0.022)	0.330 (0.020)	0	0.492 (0.009)
50	4	Design and Technology: System and Control	0	0	0	0	0	0	0.540 (0.021)	0.334 (0.026)	0	0.493 (0.012)
51	4	English Literature	0	0	0	0	0	0	0	1.045 (0.001)	0	0.220 (0.001)
52	4	Design and Technology: Food Technology	0	0	0	0	0	0	0.202 (0.009)	0.742 (0.006)	0	0.372 (0.003)
53	4	Science	0	0	0	0	0	0	0.670 (0.006)	0.342 (0.006)	0	0.282 (0.003)
54	4	Statistics	0	0	0	0	0	0	1.047 (0.010)	0.067 (0.008)	0	0.246 (0.004)
55	4	Medial, Film and Television Studies	0	0	0	0	0	0	-0.027 (0.011)	1.026 (0.009)	0	0.333 (0.003)
56	4	Fine Art	0	0	0	0	0	0	0.011 (0.015)	0.792 (0.014)	0	0.487 (0.006)
57	4	Office Technology	0	0	0	0	0	0	0.484 (0.016)	0.489 (0.015)	0	0.354 (0.003)
58	4	Home Economics: Child Development	0	0	0	0	0	0	0.240 (0.021)	0.799 (0.021)	0	0.371 (0.005)

(*Continued on next page*)

Note: TA denotes teacher assessment.

Table 20: Factor Loadings

No.	Key Stage	Description	KS1 Math	KS1 Verbal	KS2 Math	KS2 Verbal	KS3 Math	KS3 Verbal	KS4 Math	KS4 Verbal	KS4 Motive	Residual Variance
59	4	Italian	0	0	0	0	0	0	0.224 (0.053)	0.710 (0.043)	0	0.521 (0.023)
60	4	Urdu	0	0	0	0	0	0	-0.033 (0.074)	0.763 (0.078)	0	0.637 (0.019)
61	4	Additional Applied Science	0	0	0	0	0	0	0.678 (0.050)	0.481 (0.043)	0	0.311 (0.008)
62	4	Leisure and Tourism	0	0	0	0	0	0	0.112 (0.011)	0.876 (0.007)	0	0.446 (0.007)
63	4	Applied ICT	0	0	0	0	0	0	0.302 (0.016)	0.572 (0.013)	0	0.506 (0.004)
64	4	Applied Science	0	0	0	0	0	0	0.478 (0.016)	0.559 (0.018)	0	0.457 (0.005)
65	4	Health and Social Care	0	0	0	0	0	0	0.164 (0.010)	0.827 (0.012)	0	0.450 (0.007)
66	4	Applied Business	0	0	0	0	0	0	0.416 (0.012)	0.602 (0.012)	0	0.407 (0.004)
67	4	Double Science	0	0	0	0	0	0	1.056 (0.002)	0	0	0.191 (0.001)
68	4	Total GCSE Exams Taken	0	0	0	0	0	0	0	0	$1^{\dagger}$	1.710 (0.017)
69	4	Authorize Absences	0	0	0	0	0	0	0	0	-0.016 (0.000)	0.005 (0.000)
70	4	Unauthorized Absences	0	0	0	0	0	0	0	0	-0.017 (0.000)	0.003 (0.000)

Note: This table provides the factor loadings (i.e. λ_m) corresponding to equation (2) of the paper. † indicates that the factor loading has been normalized to be equal 1. The zero values denote that a given skill was not loading in that specific measurement variable. Residual variance can be interpreted as the proportion of the total variance that could be attributed to noise. TA denotes teacher assessment. Finally, when estimating the model, if a given student has a missing value in a given mandatory test, then we keep this student (in order to use his/her remaining observations on the other tests) but we include a dummy indicating the missing value.

F Robustness Check: College Enrollment

Table 21: Logistic Regression, University Enrollment, Alternative Specifications

	Full Sample – No Mixture	Full Sample – No Mixture, KS4 Factors only (controlling for earlier test scores)	White, British Sample Only – No Mixture, KS4 Factors only (controlling for earlier test scores)‡
Average Marginal Effects			
KS4 Math	0.097	0.096	0.095
KS4 Verbal	0.189	0.190	0.196
KS4 Motive	0.030	0.030	0.025

Note: No mixture implies factor error comes from a multivariate normal distribution. Model includes controls for gender, IDACI Index, free school lunch, special education needs, and school fixed effects. Skills are normalized to mean zero, standard deviation 1.

G Robustness Check: Variance Decomposition of KS3 Skills

Table 22: Development of Key Stage 3 Skills

	Contribution to R-squared	
	Math	Verbal
Background Characteristics	0.581 (0.002)	0.511 (0.002)
Key Stage 1 Shocks	0.196 (0.001)	0.190 (0.001)
Key Stage 2 Shocks	0.204 (0.001)	0.212 (0.002)
Key Stage 3 Shocks	0.061 (0.001)	0.112 (0.000)
Total	1.00	1.00

Note: Key Stage 3 shocks are defined as $\theta_{i3} - \mathrm{E}(\theta_{i3}|\theta_{i2}, \theta_{i1}, x_i)$. Key Stage 2 shocks are defined as $\mathrm{E}(\theta_{i3}|\theta_{i2}, \theta_{i1}, x_i) - \mathrm{E}(\theta_{i3}|\theta_{i1}, x_i)$. The remaining Key Stage shock is defined similarly.