



Gender segregation in higher education: an empirical test of seven explanations

Carlo Barone¹ · Giulia Assirelli²

Published online: 3 September 2019
© Springer Nature B.V. 2019

Abstract

Gender segregation in higher education (GSHE) is recognized as a key factor to explain the persistence of gender inequalities in the labor market despite the reversal of gender gap in educational attainment. Women are systematically overrepresented in fields of study, such as social sciences and the humanities, which offer relatively poor labor market prospects; at the same time, they are underrepresented in fields that perform above the average, as engineering and ICT. Several explanations for GSHE have been proposed in the literature, but their explanatory power has to be assessed yet. Using a rich longitudinal dataset on a recent cohort of Italian upper secondary school leavers, in this paper we jointly test seven potential mechanisms for GSHE. Our results show that rational choice explanations—such as skill-based explanations and gender differences in career preferences—fail to account for GSHE. On the contrary, expressive motivations related to preferences for school subjects and for specific occupations are found to mediate to a significant extent GSHE. However, our most important result concerns the key role of curricular track choice at upper secondary level which, alone, mediates two third of the gender difference in access to the humanities and social sciences and one third of the gender difference in access to engineering and ICT.

Keywords Gender segregation · Higher education · Field of study choice

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10734-019-00396-2>) contains supplementary material, which is available to authorized users.

✉ Giulia Assirelli
giulia.assirelli@gmail.com

Carlo Barone
carlo.barone@sciencespo.fr

¹ Sciences Po, Paris, France

² Independent researcher, Milan, Italy

Introduction

In most western countries, women currently outperform men in education (DiPrete and Buchmann 2013), but continue to experience significant disadvantages in the labor market, even when they complete higher education (Bobbitt-Zeher 2007). These gender inequalities reflect a complex configuration of determinants involving several family- and job-related factors, but it is well documented that gender segregation in higher education (GSHE) is a significant pathway to these labor market inequalities, because women major in fields that typically enhance unemployment and overeducation risks and lead to less lucrative jobs (Barone and Ortiz 2011; Bobbitt-Zeher 2007; Gerber and Schaefer 2004; Nunez and Livanos 2010).

Understanding the mechanisms behind GSHE is therefore a major theoretical challenge which bears important policy implications. Unfortunately, while several potential explanations for GSHE have been proposed, much less progress has been made in the direction of testing their heuristic power. As discussed below, a few previous empirical tests have been already carried out, but they have considered only one or two mechanisms separately and have provided mainly negative results, for instance showing that ability in mathematical and scientific subjects fails to account for GSHE. An additional limitation is that most of these empirical tests are concerned only with the USA, while much less is known about Europe, where educational systems are differently organized in several respects, most notably because curricular differentiation in high school is often more rigid and formalized.

Using a rich longitudinal dataset, in this work we propose a joint test of seven potential explanations for GSHE. We take Italy as test case and we consider a recent cohort of upper secondary school leavers. In Section 2 we discuss these seven potential explanations and in Section 3 we provide some background concerning the Italian case. In Section 4 we present our data, measures, and modeling strategy, which rests on the decomposition technique proposed by Breen et al. (2013). Section 5 presents the empirical results, and Section 6 concludes.

Explaining gender differences in field of study choice

Previous research indicates that the qualitative pattern of GSHE displays remarkable similarities across western countries (Barone 2011; Charles and Bradley 2002, 2009; Vaarmets 2018). Women are systematically underrepresented in STEM fields, such as engineering, computing, technology, and, to a lesser extent, physics; they are strongly overrepresented in the humanities, the social sciences,¹ teacher education, and, to a lesser extent, in medicine and other health-related fields; other fields such as the life sciences, architecture, law, business, and economics tend to be more gender-balanced. Moreover, empirical research indicates that the humanities and the social sciences offer less favorable labor market prospects in terms of unemployment rates, earnings, and the risks of overeducation and skill mismatch, while for the same outcomes graduates from engineering and computing, medicine, and the other health-related fields perform above the average (Assirelli 2015; Barone and Ortiz 2011; Davies and Guppy 1997; Reimer et al. 2011). Hence, the connection between GSHE and gender

¹ We employ a restrictive definition of “social sciences” that refers only to sociology, political science, and communication studies, but not to law nor economics. It may be noted that sociology and communication studies are female-dominated, while political science is not. Similarly, girls dominate university programs in business, while boys are overrepresented in economics. More generally, fine-grained distinctions within broad field categories shed light on subtler forms of gender segregation that we cannot consider in this work.

inequalities in career prospects is largely due the underrepresentation of women in engineering and computing and to their overrepresentation in the humanities and the social sciences. These two specific gender unbalances across fields thus constitute the focus of this paper.

Moreover, previous research reports that GSHE is highly stable across cohorts: the qualitative and quantitative patterns of gendered field of study choices display a remarkable tendency to persistently conform to gender stereotypes (Barone 2011; England and Li 2006; Mann and DiPrete 2013; van de Werfhorst 2017). This resilience is supposed to reflect the persistence of gender essentialist stereotypes (Charles and Bradley 2009). Gender essentialism promotes gender biases in the set of skills, preferences, and beliefs that are transmitted from one generation to the next (Buchmann et al. 2008; Correll 2004).

Several specific mechanisms may relate gender essentialist stereotypes to GSHE. First, gender essentialism can affect skill development through socialization processes operating through significant adults as well as through gender segregated peer networks. Indeed, a common explanation for women's underrepresentation in some math-intensive and rewarding fields relates to gender gaps in math and science (Ceci and Williams 2010). This first explanation may be derived from rational choice theory, which argues that individuals tend to prefer educational options that enhance their chances of success. Indeed, in OECD countries, boys still outperform girls in mathematic skills. According to the PISA study (OECD 2014), the average gender gap is quite modest among 15-year-old students across OECD countries (11 points on the PISA scale), but it doubles among the decile of top performers (20 points), and a similar tendency can be identified as regards scientific skills. In Italy, the gender gap in math skills is comparatively high (18 points) and it reaches 29 points in the top decile. It should be noted that in several countries, including Italy, students have been already tracked by the age of 15, i.e., when they are tested in the PISA study. Therefore, this kind of descriptive evidence is likely to reflect both genuine differences in student skills and the gendered consequences of curricular tracking. At any rate, the empirical studies that have tested whether gender differences in math and science skills mediate GSHE report poor support for this hypothesis (Ellison and Swanson 2010; Hedges and Nowell 1995; Morgan et al. 2013; Xie and Shauman 2005), at least for the USA. There is increasing evidence that the overrepresentation of boys in scientific fields cannot be explained by their advantage in math and science (Vaarmets 2018).

Less is known about a second, complementary skill-based explanation for GSHE, namely, the competitive advantage hypothesis (Jonsson 1999; Vaarmets 2018). If students prefer fields of study in which they enjoy a comparative advantage in terms of academic performance, then girls are on average more likely to select fields that reward language and communication skills more, such as the humanities and the social sciences. The scant empirical evidence regarding this second rational choice hypothesis fails to support this explanation (Jonsson 1999; Riegle-Crumb et al. 2012; van de Werfhorst et al. 2003), but more research is needed to reach conclusive evidence.

A third rational choice mechanism revolves around gender differences in career preferences. In particular, women may opt for less lucrative fields because they are more family-centered and thus attach lower value to prestige, earnings, and career prospects, while they place a higher value on family conciliation and expressive motives, such as self-realization and social utility (Cattaneo et al. 2017; Ceci and Williams 2010; Garcia-Aracil et al. 2007; Zafar 2013). However, gender differences in career orientations are not very pronounced in recent cohorts, and they seem to fall short of explaining GSHE, at least in the USA (Bobbitt-Zeher 2007; Konrad et al. 2000; Mann and DiPrete 2013; Menon et al. 2017). Moreover, these

arguments are difficult to reconcile with the dramatic increase of the presence of women in several legal and business professions, which can be demanding in terms of working time arrangements.

Let us recapitulate the hypotheses that we can derive from these first three rational choice explanations:

H1: gender differences in school performance in scientific subjects mediate the association between gender and field choice;

H2: gender differences in school performance in scientific subjects relative to performance in humanistic subjects mediate the association between gender and field choice;

H3: gender differences in career orientations mediate the association between gender and field choice.

However, it is possible that, rather than maximizing their expected utility in a narrow economic sense, girls and boys maximize the consumption value of education by choosing the fields that correspond more closely to the subjects that they have enjoyed in high school, or that lead to their preferred occupations. Hence, their expressive preferences for school subjects and for specific occupations associated with different tertiary fields represent two additional potential explanations for GSHE. These hypotheses have been put forward in an article by Morgan et al. (2013), who stressed the importance of collecting fine-grained measures of occupational plans, which tap the detailed patterns of gendered career orientation, rather than just considering the overall level of aspiration for highly or poorly lucrative jobs. Hence, this argument about expressive preferences differs from the rational choice argument that girls prefer soft fields because they attach less importance to career prospects. The underlying micro-level mechanism in the former case is not a cost-benefit calculation, but rather the emotional, symbolic value attached to specific subjects, educational programs, and occupations, regardless of their economic profitability. Indeed, there is evidence that detailed measures of “horizontal preferences” between school subjects and occupations are strong predictors of field of study choice that partly account for GSHE, at least in the USA (Morgan et al. 2013). For instance, girls develop more often a preference for humanistic subjects (e.g., arts) and for related occupations (e.g., artistic jobs), which favors enrollment in the humanities or the social sciences (Cech 2013; Morgan et al. 2013).

Furthermore, we will assess the contribution of peer influences to GSHE which reflect a variety of mechanisms: expressive motivations reflecting imitative dynamics and the concern to preserve friendship ties, but along the lines of bounded rationality models it may be argued also that peers act as a reference group and can be drivers of information about specific fields. We are aware of some case studies reporting that students are more likely to choose a given field if their classmates do the same (De Giorgi et al. 2010; Ost 2010), but we could not find any empirical test of the explanatory contribution of peer effects to GSHE.

The gender-biased influences of significant adults should be factored in as well. Teachers, parents, and school counselors encourage more or less directly and explicitly girls and boys to follow different educational tracks and tertiary programs (Zafar 2013; Gunderson 2012; Jacobs 1995). In particular, gender differences in curricular choice in high school partly reflect social control mechanisms operating through the gender-biased perception of the “natural” inclinations and preferences of students (Frank et al. 2008; Gabay-Egozi et al. 2015). Therefore, curricular track choices in high school represent an additional explanatory candidate, given

that they are strongly predictive of future field of study choices. The evidence for the USA suggests that gendered patterns in course-taking in high school contribute to GSHE only to a rather limited extent (Frank et al. 2008).² As regards European nations, much less is known. Two comparative analyses report that earlier tracking is associated with higher GSHE at the macro-level (Imdorf et al. 2015; Smyth and Steinmetz 2008), but we could not find any individual-level analysis assessing to what extent track/curricular choices mediate GSHE. This is a significant research gap, because several educational systems in Europe display a fine-grained curricular differentiation in high school that channels girls and boys into different educational pathways at an early age, when the influences of significant adults can be particularly important.

Let us recapitulate the hypotheses that we can derive from culturalist explanations:

H4: gender differences in expressive preferences for specific occupations mediate the association between gender and field choice;

H5: gender differences in expressive preferences for school subjects mediate the association between gender and field choice;

H6: gender differences in the field of study preferences of close classmates mediate the association between gender and field choice;

H7: gender differences in secondary track curricular choices mediate the association between gender and field choice.

The Italian educational system and gender segregation in higher education

In Italy, primary and lower secondary education is comprehensive and takes 8 years to complete, from the age of 6 to 14. Lower secondary school leavers are faced with three main options: academic (*licei*), technical (*istituti tecnici*), and vocational schools (*istituti professionali*). Academic schools are more prestigious and display a marked underrepresentation of male, working class, and immigrant students. Each of these tracks comprises several curricula, which are strongly patterned along gendered lines. Within the academic track, girls are underrepresented in the scientific curriculum and overrepresented in the foreign languages and socio-pedagogic curricula. In technical and vocational tracks, girls are underrepresented in industry-oriented curricula and overrepresented in service-oriented curricula (tourism, catering, accounting, etc.).

Teachers' track and curriculum recommendations are not binding, but they have an important influence on the choices of families. Parents play an important role at this stage, given the young age of their children and the lack of standardized assessments of children's skills³ (Aastrup 2007; Cavalli and Facchini 2001). Italy is thus an interesting case study because, similar to several other countries in continental Europe, curricular choices are more precocious, formalized, and constraining than in the USA, which has been the focus of most previous empirical studies.

² For the USA, there is evidence that also extra-curricular activities may contribute to GSHE (Legewie and DiPrete 2012, 2014).

³ The national bureau for school testing (INVALSI) carries out regular assessments of students' skills, but their individual-level results are not communicated to the teachers, nor to the families.

All of these upper secondary tracks and curricula take 5 years to complete and provide access to any program in higher education. However, universities often set a numerus clausus or entry examinations, which can be more or less demanding across tertiary fields. Higher education in Italy comprises a large university sector and a small sector of short post-secondary vocational programs. University education is articulated into 3-year bachelor courses and 2-year master courses. Field of study choice occurs at the entry into the bachelor level and mobility between fields in the transition to the master level is rare (ANVUR 2016).

Finally, it may be noted that in Italy the level and pattern of GSHE display high similarity with those that have been documented for other western countries (Barone 2011; Charles and Bradley 2009; Triventi 2010). Vertical and horizontal gender segregation in the labor market is quite pronounced in Italy (Almalaurea 2015), but less than in other industrialized nations: possibly owing to the low rate of female labor market participation, employed women are a more selected population in Italy (Charles and Grusky 2004). However, in younger cohorts, the activity rate of female graduates is aligned with those observed in other OECD countries (OECD 2016).

The research design: data, variables, and methods

This study is based on a longitudinal survey carried out in 31 high schools from all types of upper secondary tracks. These schools are located in four provinces (Milan, Vicenza, Bologna, and Salerno) that cover different geographical areas of the country. We first drew a random sample of schools proportionally stratified by province and school track (i.e., the number of students from each province and track in the sample is proportional to their share in the overall population of the four provinces). All senior students (grade 13) of these schools were interviewed in three separate occasions ($N = 4805$). The first wave was conducted at the beginning of the senior school year (October 2013) and involved self-administered questionnaires in the classrooms; the response rate was 99%. The second wave occurred at the end of the same school year (May 2014), and it was based on computer-assisted telephone interviews (CATI); the cumulative response rate was 82.8%. The third wave was conducted in November 2014 and recorded students' university enrollment decisions, again with a CATI methodology; the cumulative response rate was 79.1%. Overall, the high level of participation of the schools and of the students ensures a good level of external validity for our study.

The analytical sample, which is restricted to university students for which we have valid information on the chosen field of study, includes 2403 individuals.⁴

The primary outcome of interest for our analyses is field of study choice among university students (measured in wave 3). More specifically, following our previous arguments concerning the labor market prospects of different fields (Section 2), we will analyze gender differences in enrollment in occupationally weak, highly feminized fields (the humanities and social sciences), as well as in occupationally strong, masculine fields (engineering, computing). In other words, we focus on those fields of study that provide the most significant contributions to gender inequalities in the labor market, because they display a significant unbalance (higher than 10%) in the share of male and female students *and* they are associated

⁴ This survey was part of a broader experimental study that included 31 additional schools where the research team developed a counseling intervention targeting all senior students. For this reason, these schools were excluded from the analyses. There is no evidence of contamination between treated and control schools (Barone et al. 2017).

with significantly above- or below-average labor market outcomes.⁵ We rely on a simple distinction between fields with below-average, average, and above-average occupational returns because it is robust across undergraduate and graduate programs as well as across different occupational indicators, such as unemployment risks, earnings, and overeducation. This basic pattern is well documented in the literature on returns to fields of study concerning the Italian case and other western countries (Almalaurea 2015; Reimer et al. 2011).

As regards school performance in scientific subjects (*H1*), we consider student's self-reported grades in math and in science (the two scientific subjects that are taught in high school tracks) on a scale from 0 to 10. As regards the hypothesis of competitive advantage (*H2*), following Jonsson's (1999) seminal paper, we take the ratio between mean grades in Italian and foreign languages (the two humanistic subjects that are taught in all high school tracks) and mean grades in math and sciences. As regards the high school curriculum (*H7*), we can rely on a detailed eight-category variable that refers to the track (academic, technical, vocational) and to the specific curriculum chosen by the student: scientific, classical studies, psychology-pedagogy, and foreign languages curricula for the academic track; industry-oriented, service-oriented, and foreign languages curricula for the technical track; industry-oriented and service-oriented curricula for the vocational track. The service-oriented curricula in the technical track typically train for skilled white collar professions (e.g., accountant, surveyor), while those in the vocational track are more focused on tourism, catering, and personal services. In order to estimate the model for the KHB decomposition, we must aggregate the industry-oriented technical and vocational track, on the one side, and the service-oriented and foreign languages curricula, on the other side.

As regards career orientations (*H3*), we use the following question: "To what extent do you attach importance to your future career prospects when choosing a field of study? Indicate a value between 1 (not at all) and 10 (completely)." As regards students' expressive preferences for specific occupations (*H4*), interviewees were asked to report the occupation that they would like to do within 10 years and their open answers were then coded into a nine-category variable. Similarly, for subject preferences in high school (*H5*), students were asked to indicate up to three favorite subjects and their answers were coded into four categories: humanistic, scientific, technical disciplines, and a separate category for business and law. For the analyses, we take into account the first preference expressed by students.

Finally, as regards peer influences (*H6*), each student was asked to indicate up to three closest friends in the classroom; we could map the final choices of these classmates from wave 3 and we constructed two variables to mark if the student has none, only one, or at least two friends choosing the relevant field (the humanities and social sciences, when modeling access to these fields; engineering and computing, when this is the outcome of interest). As an alternative specification, we can replace the final choices of these classmates with their initial preferences, recorded in wave 1. We can also consider the percentage of classmates or schoolmates choosing the relevant field. We will present the results based on the first solution and we will comment those based on the alternative specifications. It should be noted that all these three specifications refer only to school-based peer influences and that we cannot therefore extend our conclusions to the peer influences occurring outside the school context.

We estimate a sequence of binomial logit models among students who enrolled at university, where we incorporate each of the above measures sequentially, following temporal order

⁵ The original data were recorded using an open format and then recoded into the two-digit Isced classification for field of study, which served as basis for the aggregations reported in the text.

whenever possible. We first look at the corresponding parameters and at the variations of gender effects across models, then we use the KHB method for a formal decomposition exercise for categorical outcomes based on a mediation analysis (Breen et al. 2013). By holding constant the scale and the fit of the error to the assumed distribution, this method presents a generalization to nonlinear probability models of the standard decomposition approach used in linear regression. Birth year, country of birth, province of residence, and parental education are introduced in the models as control variables, as well as measures of general school performance both before track choice in high school and at the final exam in high school. Standard errors are clustered at the school level.

This study improves over previous research by incorporating several potential mechanisms of GSHE altogether, but we would stress that we still miss some potentially important layers of GSHE. For instance, we are not able to directly operationalize teachers' or parents' influences.⁶ To the extent that we omit some potentially important determinants of field of study choices, which could correlate with the determinants incorporated in our models, we cannot discard omitted-variable bias; reverse causality is also a potential concern, for instance girls may want to enroll more often in the humanities for some reason other than those discussed above and they may then tend to develop occupational plans that are coherent with these choices as a mechanism of cognitive-dissonance reduction. Our longitudinal design reduces these risks of reverse causality because information on the determinants of field of study choices is collected 1 year before (wave 1) the information on these field choices (wave 3), but this does not eliminate the problem. Hence, we cannot interpret the estimates of the coefficients of the variables in our models as truly "causal effects."

Results

Descriptive results

In this sub-section, we present some descriptive analyses concerning the evolution of gender differences in field preferences across waves, as well as the determinants of these preferences. These analyses provide the background for the multivariate results that we will present in Section 5.2. Figure 1 refers to the variations over time in the field preferences of male and female students who enrolled at university. As explained above, the data for wave 1 refer to intended field of study choices at the beginning of the high school senior year, while wave 3 refers to the actual decisions of the students. Starting with strong fields, it can be seen that for both girls and boys engineering and ICT display a moderate increase over time in student preferences, while medicine and health-related fields decline considerably—most likely because of the strong selectivity of entry examinations in these highly regulated fields.

Conversely, both the humanities and the social sciences increase significantly for female and male students when comparing initial preferences and final decisions. Overall, we observe a significant decline over time of strong fields and a corresponding increase of weak fields for both genders. Interestingly, this trend is more pronounced for girls: 27.9% of them planned to select an occupationally strong field, but only 13.9% finally chose one. This decline is twice as

⁶ Another hypothesis that we cannot test with our data is that GSHE reflects anticipated discrimination in university or at the workplace (Polachek 1978), but the evidence in this respect is generally not supportive (Mastekaasa and Smeby 2008).

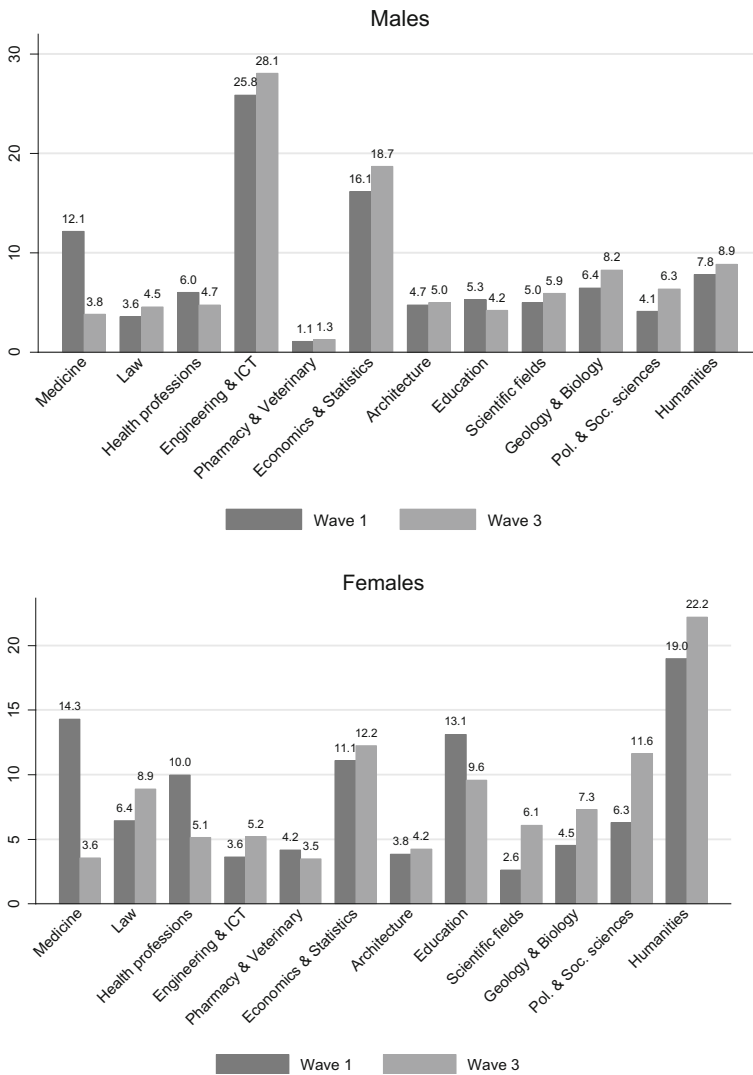


Fig. 1 Variation over time of field preferences by gender, detailed classification of fields of study. Descriptive analysis

large as the reduction observed among their male counterparts (from 44 to 36.6%); conversely, the share of girls who choose a weak field (33.8%) is substantially higher (+ 8.5%) than the share of girls who initially planned to do so (25.3%). We detect a much smaller increase among boys (+ 3.3%). Hence, gender differences in access to rewarding fields of study widen substantially over the senior high school year (Alon and DiPrete 2017).

Table 1 displays the distributions by gender of the independent variables. These distributions are informative of the potential explanatory power of each factor: if its distribution does not vary by gender, it cannot contribute to gender differences in field of study choice. As can be seen, gender differences across high school curricula are strong.

Table 1 The distribution of the determinants of field of study choices across male and female students

	Males	Females	<i>p</i> value <i>t</i> test	<i>p</i> value Regression ^a
Track and curriculum (%)				
General: scientific curriculum	51.6	31.0	0.000	0.000
General: classical studies curriculum	10.3	14.7	0.001	0.024
General: foreign languages and psycho-pedagogical curricula	6.5	34.2	0.000	0.000
Technical: industry-oriented curricula	12.9	2.7	0.000	0.029
Technical: business curricula	13.5	8.0	0.000	0.042
Technical: foreign languages curricula	1.5	5.7	0.000	0.089
Vocational: industry-oriented curricula	2.9	1.8	0.080	0.391
Vocational: service-oriented curricula	0.8	1.8	0.058	0.087
Grades in scientific subjects (1–10)				
Maths	6.8	7.0	0.000	0.001
Sciences	7.2	7.3	0.001	0.013
Comparative advantage (0.6–3)	0.99	1.03	0.000	0.000
Preferred subject (%)				
Scientific subjects	43.9	28.3	0.000	0.000
Technical subjects	7.3	1.2	0.000	0.007
Humanistic subjects	39.0	63.9	0.000	0.000
Business and law	7.4	5.4	0.048	0.188
NR	2.3	1.3	0.058	0.057
Occupational preferences (%)				
Managers and entrepreneurs	12.8	7.1	0.000	0.000
Legal and business professionals	7.3	6.3	0.342	0.519
Medical and scientific professionals	18.0	24.8	0.000	0.001
Architects	2.3	2.6	0.669	0.714
Engineers & ICT specialists	10.8	1.4	0.000	0.000
Humanistic professions	7.6	8.0	0.764	0.783
Teachers	1.4	5.1	0.000	0.000
Care-oriented professions	4.3	11.4	0.000	0.000
Unqualified workers	4.4	4.3	0.821	0.797
NR	31.0	29.1	0.314	0.417
Importance of career prospects (1–10)	7.5	7.7	0.035	0.036
Friends enrolled in the humanities and social sciences (%)				
None	74.1	52.5	0.000	0.000
One	20.0	27.8	0.000	0.001
At least two	5.9	19.7	0.000	0.000
Friends enrolled in engineering and ICT (%)				
None	70.6	89.3	0.000	0.000
One	22.0	9.5	0.000	0.000
At least two	7.4	1.2	0.000	0.000

^a The statistical significance of the difference is estimated by means of regression models where the characteristic to be tested is the dependent variable, gender is the independent variable (standard errors are clustered at the school level)

In particular, as regards the general track, girls are overrepresented in the curricula of classical studies (14.7 vs. 10.3% for boys) and, more markedly, of foreign languages and psychology-pedagogy (34.2 vs. 6.5% for boys). Conversely, boys are heavily overrepresented in scientific curricula (51.6 vs. 31% for girls) of the general track, as well and in the industry-oriented curricula of technical (12.9%) and vocational tracks (2.9%), while girls dominate in the foreign language and service-oriented curricula of these tracks. Hence, curricular differences may be a relevant mechanism of GSHE.

To the contrary, when inspecting the average grades in math and science of female and male students, it is apparent that girls outperform boys across both subjects in this recent cohort of

students. Moreover, when contrasting performance in the two humanistic subjects against performance in the two scientific subjects, we see that girls enjoy a slight but statistically significant comparative advantage over boys in humanistic subjects. However, for both genders, school performance varies only moderately across domains. In particular, for girls, the average grade in Italian and in English is, respectively, 7.4 and 7.3; the corresponding values for boys are 6.9 in both subjects.⁷ Overall, girls outperform boys in all subjects, and only slightly more in humanistic than in scientific subjects. Hence, skill-based explanations are not promising candidates to account for GSHE. It should be noted also that in Italy, due to the comparatively high drop-out rates in the first 2 years of upper secondary education, high school seniors are a selected population with respect to school performance.

Instead, subject preferences reveal strong gender differences along expected lines: boys more often prefer scientific subjects (43.9 vs. 28.3% for girls) or technical subjects (7.3 vs. 1.2%), while girls display a much stronger interest in humanistic subjects (63.9 vs. 39% for boys). Also gender differences in the preferred types of occupations are very pronounced, particularly as regards engineering and ICT professions (10.8% for boys vs 1.4% for girls), managerial jobs (12.8 vs. 7.1%), teaching (1.4 vs. 5.1%), and care-oriented occupations (4.3 vs. 11.4%). Hence, the expressive preferences for school subjects and future occupations are strongly patterned along gender lines and may thus contribute to the explanation of GSHE. At the same time, as regards the subjective importance of finding jobs with high career prospects, we observe a small significant difference in favor of girls, which casts further doubts on the explanatory potential of the rational choice mechanisms.

Finally, girls mention more often as close friends one classmate (27.8%) or two (19.7%) who enroll in the humanities or in the social sciences, while boys mention more often one (22%) or two (7.4%) classmates enrolling in engineering or ICT. Hence, the gender segregation of friendship networks may contribute to GSHE.

Multivariate analyses

In this section, we consider whether the above-described factors affect field of study choices by means of two sets of binomial logistic regression models referring to enrollment in the humanities and social sciences (Table 2), or in engineering and ICT (Table 3). We consider only students who enrolled at university and report the average marginal effects, which can be interpreted as differences between categories in the probability of enrolling in a given field. We will comment the parameters referring to the potential mediators of GSHE, as well as the variations across models in the coefficients for gender which provide some first indications on the explanatory power of these mediators that will be later tested more formally using the KHB method.

Model 1 displayed on Table 2 refers to the total effect of gender on the chances of choosing a humanistic or social science field (boys are the reference category), controlling for socio-demographic covariates and for the final school mark at lower secondary examinations.⁸ As can be seen, we detect an overall strong gender gap, which amounts to 24.6%.

⁷ In Italian high schools, boys still slightly outperform girls in math at standardized tests in grade 10 (Invalsi 2017), but at any given level of performance in these tests girls obtain higher grades (Argentin and Triventi 2016). To repeat, students, parents, and teachers do not have access to the individual results of these standardized tests; therefore, the only ability signal that families receive refers to grades.

⁸ This is the standard measure for overall school performance before tracking in Italy.

Table 2 Determinants of enrollment in the humanities and social sciences at university: binomial logistic regression models, average marginal effects

	M1	M2	M3	M4	M5	M6	M7	M8
Gender (Ref: male)	0.246*** (0.021)	0.072*** (0.022)	0.082*** (0.023)	0.081*** (0.024)	0.068*** (0.024)	0.064*** (0.026)	0.067*** (0.027)	0.065*** (0.027)
Province of residence (Ref: Bologna)	0.057 (0.084)	0.029 (0.021)	0.025 (0.019)	0.025 (0.018)	0.028 (0.018)	0.023 (0.019)	0.024 (0.018)	0.021 (0.017)
Salerno	-0.060 (0.075)	-0.043 (0.028)	-0.036 (0.036)	-0.072** (0.035)	-0.056 (0.035)	-0.041 (0.031)	-0.035 (0.031)	-0.030 (0.033)
Vicenza	-0.014 (0.111)	-0.034 (0.024)	-0.025 (0.024)	-0.024 (0.024)	-0.016 (0.020)	-0.002 (0.022)	-0.000 (0.021)	0.003 (0.021)
1995	-0.014 (0.032)	-0.036 (0.026)	-0.004 (0.025)	-0.004 (0.025)	0.006 (0.024)	-0.001 (0.021)	-0.000 (0.021)	-0.005 (0.021)
1996 or later	-0.088 (0.079)	-0.119* (0.066)	-0.074 (0.060)	-0.074 (0.060)	-0.057 (0.052)	-0.038 (0.048)	-0.040 (0.048)	-0.046 (0.049)
Missing	0.107 (0.091)	-0.056 (0.041)	-0.012 (0.041)	-0.012 (0.041)	-0.014 (0.036)	-0.039 (0.031)	-0.040 (0.030)	-0.048 (0.031)
Abroad	-0.044 (0.077)	-0.064 (0.064)	-0.080 (0.063)	-0.080 (0.063)	-0.085 (0.068)	-0.079 (0.062)	-0.077 (0.063)	-0.078 (0.064)
Missing	-0.019 (0.044)	-0.005 (0.032)	-0.022 (0.034)	-0.022 (0.034)	-0.023 (0.034)	-0.042 (0.026)	-0.043 (0.027)	-0.041 (0.028)
Upper secondary	0.032 (0.033)	0.024 (0.028)	0.023 (0.029)	0.023 (0.029)	0.019 (0.028)	0.014 (0.024)	0.015 (0.024)	0.020 (0.023)
Tertiary	-0.030 (0.035)	-0.030 (0.030)	-0.029 (0.030)	-0.030 (0.030)	-0.037 (0.028)	-0.029 (0.024)	-0.028 (0.024)	-0.024 (0.023)
Missing	0.074 (0.143)	0.130 (0.099)	0.100 (0.086)	0.101 (0.084)	0.098 (0.089)	-0.001 (0.086)	-0.023 (0.087)	-0.003 (0.089)
Good	0.037 (0.052)	0.024 (0.047)	0.023 (0.044)	0.023 (0.044)	0.020 (0.042)	0.017 (0.040)	0.017 (0.040)	0.017 (0.040)
Very good	-0.002 (0.059)	0.012 (0.045)	0.047 (0.042)	0.047 (0.043)	0.040 (0.041)	0.034 (0.038)	0.034 (0.038)	0.036 (0.038)
Excellent	-0.041 (0.066)	-0.028 (0.046)	0.062 (0.045)	0.061 (0.047)	0.053 (0.046)	0.052 (0.043)	0.054 (0.042)	0.056 (0.041)
Classic lyceum	0.251*** (0.057)	0.269*** (0.068)	0.269*** (0.068)	0.269*** (0.067)	0.219*** (0.059)	0.207*** (0.054)	0.206*** (0.054)	0.181*** (0.050)
Track (Ref: Scientific lyceum)	0.492*** (0.066)	0.492*** (0.066)	0.494*** (0.067)	0.493*** (0.067)	0.441*** (0.059)	0.356*** (0.054)	0.357*** (0.054)	0.311*** (0.050)

Table 2 (continued)

	M1	M2	M3	M4	M5	M6	M7	M8
Foreign lang. or psych. lyceum		(0.034) -0.077**	(0.033) -0.055	(0.033) -0.055	(0.032) -0.044	(0.034) -0.052	(0.033) -0.050	(0.034) -0.043
Technical: industry-oriented		(0.038)	(0.039)	(0.039)	(0.041)	(0.037)	(0.037)	(0.038)
Technical: business		0.034	0.066**	0.066**	0.080***	0.092***	0.097***	0.101***
Technical: foreign languages		(0.034)	(0.031)	(0.031)	(0.029)	(0.027)	(0.027)	(0.027)
Vocational: industry-oriented		0.390***	0.421***	0.421***	0.375***	0.297***	0.304***	0.275***
Vocational: services		(0.069) -0.096***	(0.063) -0.063	(0.063) -0.063	(0.054) -0.062	(0.062) -0.057	(0.062) -0.053	(0.058) -0.040
Grade in maths in high school (1–10)		(0.031)	(0.040)	(0.040)	(0.044)	(0.046)	(0.046)	(0.051)
Grade in sciences in high school (1–10)		0.303*** (0.113)	0.374*** (0.094)	0.374*** (0.094)	0.366*** (0.090)	0.224*** (0.083)	0.235*** (0.082)	0.253*** (0.083)
Upper secondary graduation mark			-0.056*** (0.007)	-0.055*** (0.009)	-0.053*** (0.009)	-0.052*** (0.008)	-0.052*** (0.008)	-0.053*** (0.008)
Comparative advantage (range 0.6–3)			-0.039*** (0.010)	-0.038*** (0.010)	-0.035*** (0.010)	-0.025*** (0.010)	-0.026*** (0.010)	-0.028*** (0.010)
Preferred school subject (Ref: Scientific)			0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)
				0.017 (0.074)	-0.081 (0.071)	-0.112* (0.065)	-0.112* (0.066)	-0.125* (0.065)
					0.066 (0.053)	0.058 (0.046)	0.053 (0.045)	0.050 (0.046)
					0.162*** (0.016)	0.113*** (0.013)	0.111*** (0.013)	0.111*** (0.013)
					0.006 (0.030)	-0.013 (0.025)	-0.013 (0.025)	-0.011 (0.024)
					0.062 (0.039)	0.035 (0.044)	0.034 (0.046)	0.036 (0.047)
						-0.134*** (0.044)	-0.136*** (0.046)	-0.137*** (0.047)

Table 2 (continued)

	M1	M2	M3	M4	M5	M6	M7	M8
Occupational expressive preferences (Ref: Manager)								
Legal and business professions						(0.037) -0.066*	(0.038) -0.068*	(0.037) -0.071*
Medical and scientific professionals						(0.038) -0.159***	(0.038) -0.160***	(0.040) -0.158***
Architects						(0.060) -0.165***	(0.059) -0.165***	(0.061) -0.166***
Engineers & ICT specialists						(0.055) 0.356***	(0.056) 0.349***	(0.056) 0.342***
Humanistic professions						(0.037) 0.305***	(0.035) 0.289***	(0.037) 0.284***
Teachers						(0.071) 0.271***	(0.070) 0.266***	(0.072) 0.260***
Care-oriented professions						(0.054) 0.158***	(0.054) 0.151***	(0.054) 0.143***
Unskilled occupations						(0.047) 0.071*	(0.048) 0.066*	(0.049) 0.065*
NR						(0.040) -0.013***	(0.039) -0.013***	(0.039) -0.012***
Importance of career prospects (1–10)						(0.003) 0.056**	(0.003) 0.056**	(0.003) 0.056**
Friends enrolled in humanities and social sciences (Ref: None)						(0.023) 0.085***	(0.023) 0.085***	(0.023) 0.085***
Observations	2403	2403	2403	2403	2403	2403	2403	2403

Robust standard errors in parentheses

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table 3 Determinants of enrollment in engineering and ICT at university: binomial logistic regression, average marginal effects

	M1	M2	M3	M4	M5	M6	M7	M8
Gender (Ref: male)	-0.223*** (0.019)	-0.167*** (0.012)	-0.168*** (0.012)	-0.165*** (0.012)	-0.151*** (0.011)	-0.116*** (0.009)	-0.117*** (0.009)	-0.110*** (0.010)
Province of residence (Ref: Bologna)	0.024 (0.043)	0.046 (0.054)	0.052 (0.052)	0.050 (0.053)	0.049 (0.049)	0.037 (0.045)	0.036 (0.044)	0.035 (0.041)
Salerno	0.078* (0.043)	0.094* (0.052)	0.104** (0.050)	0.102** (0.051)	0.099** (0.048)	0.068 (0.043)	0.064 (0.043)	0.059 (0.041)
Vicenza	0.015 (0.057)	0.031 (0.062)	0.035 (0.059)	0.034 (0.060)	0.024 (0.054)	0.020 (0.049)	0.018 (0.049)	0.018 (0.046)
1995	0.066*** (0.021)	0.059*** (0.022)	0.050** (0.023)	0.050** (0.023)	0.044* (0.024)	0.044** (0.021)	0.043** (0.021)	0.040* (0.020)
1996 or later	0.053 (0.043)	0.058 (0.047)	0.040 (0.049)	0.040 (0.050)	0.042 (0.050)	0.022 (0.042)	0.021 (0.043)	0.016 (0.040)
Missing	0.080 (0.062)	0.175** (0.075)	0.179** (0.077)	0.181** (0.076)	0.177** (0.071)	0.170*** (0.063)	0.172*** (0.063)	0.160*** (0.062)
Abroad	0.042 (0.042)	0.050 (0.041)	0.043 (0.036)	0.043 (0.036)	0.036 (0.038)	0.028 (0.029)	0.025 (0.027)	0.023 (0.027)
Missing	0.020 (0.031)	0.016 (0.033)	0.023 (0.034)	0.024 (0.034)	0.023 (0.033)	0.034 (0.029)	0.029 (0.030)	0.024 (0.030)
Parental education (Ref: Lower secondary)	0.016 (0.023)	0.016 (0.026)	0.019 (0.025)	0.020 (0.025)	0.021 (0.026)	0.013 (0.024)	0.012 (0.024)	0.011 (0.023)
Upper secondary	0.049** (0.025)	0.045* (0.025)	0.046* (0.024)	0.047* (0.024)	0.048** (0.024)	0.037 (0.024)	0.036 (0.024)	0.034 (0.023)
Tertiary	-0.008 (0.065)	-0.028 (0.057)	-0.024 (0.058)	-0.027 (0.057)	-0.014 (0.069)	-0.021 (0.064)	-0.017 (0.066)	-0.013 (0.064)
Missing	0.017 (0.031)	0.001 (0.030)	-0.001 (0.033)	0.001 (0.032)	0.006 (0.025)	0.013 (0.026)	0.012 (0.026)	0.011 (0.026)
Good	0.045 (0.032)	0.021 (0.028)	0.007 (0.033)	0.010 (0.032)	0.014 (0.027)	0.023 (0.028)	0.021 (0.028)	0.022 (0.028)
Very good	0.071* (0.038)	0.049 (0.031)	0.019 (0.035)	0.025 (0.036)	0.023 (0.032)	0.028 (0.032)	0.025 (0.032)	0.027 (0.032)
Excellent	-0.136*** (0.027)	-0.136*** (0.027)	-0.147*** (0.024)	-0.145*** (0.024)	-0.127*** (0.023)	-0.099*** (0.026)	-0.100*** (0.026)	-0.087*** (0.024)
Classic lyceum	-0.183*** (0.038)	-0.183*** (0.038)	-0.191*** (0.038)	-0.190*** (0.038)	-0.178*** (0.038)	-0.167*** (0.038)	-0.167*** (0.038)	-0.154*** (0.038)

Table 3 (continued)

	M1	M2	M3	M4	M5	M6	M7	M8
Foreign lang. or psych. lyceum	(0.019) 0.053 (0.091) -0.144*** (0.023) -0.099* (0.057) -0.069 (0.052) 0.000 (0.000)	(0.018) 0.043 (0.086) -0.157*** (0.020) -0.110** (0.053) -0.086* (0.046) 0.000 (0.000)	(0.017) 0.040 (0.086) -0.156*** (0.020) -0.110** (0.052) -0.086* (0.047) 0.000 (0.000)	(0.019) -0.008 (0.065) -0.162*** (0.023) -0.099** (0.047) -0.099*** (0.035) 0.000 (0.000)	(0.022) -0.026 (0.055) -0.161*** (0.023) -0.100** (0.048) -0.065** (0.031) 0.000 (0.000)	(0.023) -0.027 (0.055) -0.163*** (0.024) -0.104** (0.050) -0.068** (0.032) 0.000 (0.000)	(0.021) -0.018 (0.050) -0.148*** (0.022) -0.091* (0.048) -0.049 (0.032) 0.000 (0.000)	
Grade in maths in high school (1–10)					0.025*** (0.007)	0.019*** (0.006)	0.019*** (0.006)	0.017*** (0.006)
Grade in sciences in high school (1–10)					-0.007 (0.011)	-0.006 (0.011)	-0.006 (0.012)	-0.005 (0.012)
Upper secondary graduation mark					0.000 (0.009) -0.001 (0.001)	0.000 (0.000) -0.000 (0.001)	0.000 (0.000) -0.001 (0.001)	0.000 (0.000) -0.001 (0.001)
Comparative advantage (range 0.6–3)					-0.041 (0.072)	-0.059 (0.070)	-0.061 (0.071)	-0.067 (0.068)
Preferred school subject (Ref: Scientific)					0.104* (0.061)	0.069 (0.058)	0.073 (0.059)	0.067 (0.056)
Technical					-0.078*** (0.011)	-0.069*** (0.011)	-0.068*** (0.011)	-0.065*** (0.011)
Humanistic					0.010 (0.045)	0.027 (0.052)	0.029 (0.053)	0.028 (0.052)
Business and Law					-0.078** (0.036)	-0.071** (0.032)	-0.070** (0.031)	-0.062* (0.034)
Missing								
Occupational expressive preferences (Ref: Manager)					-0.057 (0.038)	-0.057 (0.038)	-0.056 (0.038)	-0.049 (0.035)
Legal and business professions					-0.090*** (0.036)	-0.088*** (0.036)	-0.088*** (0.036)	-0.079*** (0.036)

Table 3 (continued)

	M1	M2	M3	M4	M5	M6	M7	M8
Medical and scientific professionals						(0.024)	(0.023)	(0.020)
Architects						-0.068*	-0.069**	-0.062*
Engineers & ICT specialists						(0.035)	(0.035)	(0.034)
Humanistic professions						0.193***	0.193***	0.193***
Teachers						(0.042)	(0.042)	(0.044)
Care-oriented professions						-0.075**	-0.071**	-0.064*
Unqualified workers						(0.031)	(0.032)	(0.033)
NR						-0.134***	-0.127***	-0.118**
						(0.045)	(0.048)	(0.047)
						-0.074*	-0.071	-0.063
						(0.044)	(0.044)	(0.043)
						-0.048	-0.046	-0.035
						(0.032)	(0.032)	(0.027)
						-0.004	-0.001	0.008
						(0.025)	(0.025)	(0.023)
							0.008	0.007
							(0.005)	(0.005)
								0.028
Importance of career prospects (1–10)								(0.019)
Friends enrolled in engineering & ICT (Ref: One None)	2403	2403	2403	2403	2403	2403	2403	0.074**
								(0.031)
Observations	2403	2403	2403	2403	2403	2403	2403	2403

Robust standard errors in parentheses

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

In model 2, we introduce a set of dummy variables that refer to the upper secondary curriculum chosen by each student. This is a powerful predictor of field of study choice: relative to students attending scientific curricula in the general track (reference category), those in classical studies curricula have 25.1% higher chances of enrolling in the humanities and social sciences, while those in foreign language and psycho-pedagogical curricula have 49.2% higher chances; these fields are chosen more often also by students of foreign language curricula in technical tracks (+39%) and of service-oriented curricula in vocational tracks (+30.3%). Crucially, following the introduction of this single variable, the effect of gender is reduced from 24.6% (model 1) to 7.2% (model 2), in line with hypothesis 7. This reflects the strong gender unbalances between secondary curricula described in the previous section and the strong effects that these curricula exert on field of study choices.

Model 3 indicates that one additional point in the average grade in math and in science reduces by, respectively, 5.6 and 3.9% the chances of enrolling in humanistic and social science fields. Hence, performance in scientific subjects is a powerful predictor of field of study choice. However, contrary to hypothesis 1, it does not seem to mediate gender differences across fields, given that the parameter for gender has changed little in model 3. This could be expected given that, in this recent cohort of students, boys do not report higher grades than girls in scientific subjects. The measure of comparative advantage (model 4) does not have a statistically significant effect on field of study choice (contrary to hypothesis 2), and the same applies to the overall mark at the final upper secondary examinations. Overall, general school performance in lower and upper secondary education (models 1 and 4) impacts little on the outcome; the only relevant skill-based predictor is performance in scientific subjects.

Models 5 and 6 refer to student preferences for school subjects and for specific occupations to test hypotheses 4 and 5. These preferences are strong predictors of the outcome: a preference for humanistic subjects in high school boosts by 16.2% the preference for the corresponding fields relative to students mentioning a preference for scientific subjects (reference category, model 5). Moreover, aspiring to service-oriented, humanistic, and teaching professions correlates strongly with later enrollment in the humanities and the social sciences: the corresponding parameters are 27.1, 35.6, and 30.5%. When adding these variables referring to student expressive preferences in models 5 and 6, the residual effect of gender further decreases from 8.1 to 6.3%. Interestingly, these two factors also mediate the effect of secondary curricula to some extent. In particular, we observe substantial reductions—between one fourth and one third of the values in model 4—in the parameters referring to curricula in classical studies, other humanistic tracks, foreign languages in the technical track, and service-oriented curricula in vocational tracks.

Model 7 assesses the relevance of instrumental preferences for field of study choices (hypothesis 3): in line with rational choice theory, students who are more career-oriented display a lower propensity to enroll in the humanities and the social sciences: one point on this scale, which ranges from 0 to 10, reduces by 1.3% the chances to enroll in these fields. However, the effect of gender is virtually unchanged. This was expected, because girls are not less career-oriented than boys. Finally, model 8 indicates that, controlling for all the previous covariates, the share of close classmates choosing the humanities and social sciences is a significant predictor of respondent's own field of study choice. In particular, having at least two friends who take these fields enhances the chances of choosing them by 8.5%. The parameter for gender is unaffected by the inclusion of this variable, contrary to hypothesis 6.

The pattern of results for enrollment in engineering and computing (Table 3) displays strong similarities with those that we have just commented. The overall effect of gender, which is now

negative, is almost as strong (-22.3%), and curricular choices in secondary education display strong influences on field of study choice. Moreover, performance in math is a relevant predictor of enrollment in these fields, while grades in science as well as the other skill-based predictors do not correlate with the outcome. Expressive preferences for scientific and technical subjects and for engineering and ICT professions are powerful predictors, while career aspirations do not seem to play any role. The effects of classmates' field of study choices are confirmed, too.

Moreover, the initial gender differential in access to these fields is reduced substantially following the introduction of these mediators. However, in this case, the role of secondary curricula choices seems less prominent, while expressive preferences for specific occupations seem to play a more important role.

The results of the KHB decomposition, reported in Fig. 2, confirm the pattern of findings based on the logistic regression. In particular, curricular track alone accounts for almost two thirds (64%) of gender differences in access to the humanities and social sciences, and for almost one third (29%) of the gender gap in engineering and computing. The two skill-based explanations as well as career orientations do not hold any explanatory power (actually for the first outcome they even have negative explanatory power, meaning that they promote an underrepresentation of girls in occupationally weak fields). Expressive preferences for specific subjects and occupations give a small but significant contribution to explain gender differences in the humanities and social sciences (8% if we compare models 4 and 6), while their explanatory power is twice as high (15%) when it comes to engineering and computing. Overall, these determinants of field of study choices altogether account for 67% of the gender gap in soft fields and for 47% of this gap in the hard fields.

Finally, let us comment the results of some robustness checks (available upon request to the authors). First, the pattern of findings that we have reported does not significantly change if we run the models on each field separately (the humanities, the social sciences, engineering, computing), but the estimates become more unstable (due to lower numbers for the positive outcome). Moreover, as reported in the [online appendix](#), using a three-category multivariate logit model (humanities and social sciences, engineering and ICT, gender-neutral fields) leads to the same results as using two sets of binomial logits, and the same holds for a four-category specification where gender-neutral fields are further disaggregated (economics and law vs. scientific fields). Second, we have experimented more and less detailed specifications for the variable referring to secondary curricula: more detailed specifications display slightly higher explanatory power as

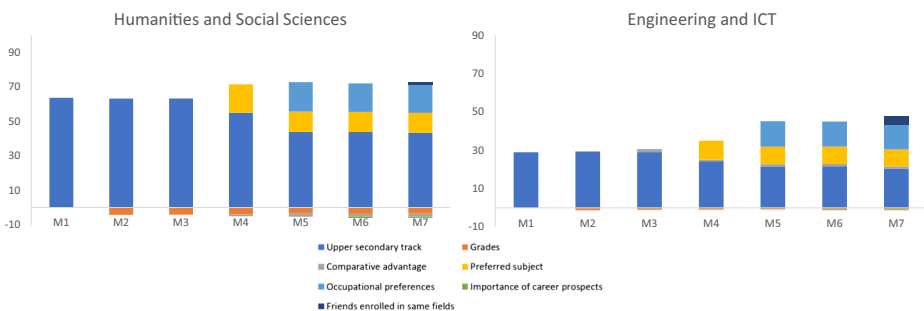


Fig. 2 KHB decomposition of gender differences across fields, % explained by: secondary curricular track, school performance in scientific subjects, school performance in scientific subjects relative to performance in humanistic subjects, career orientations, expressive preferences for school subjects, expressive preferences for specific occupations, field of study preferences of close classmates

expected, but the differences are small. Third, we can measure peer effects using a broader range of fields or considering the percentage of class- or schoolmates selecting more and less rewarding fields: the conclusion that peer effects matter for field of study choice but account for GSHE only marginally still holds. Finally, in the models reported in Tables 2 and 3, we have followed a chronological criterion concerning the order of introduction of the covariates (general marks in lower secondary education, track choice at the beginning of upper secondary education, subject-specific marks in the previous school year, student preferences for subjects and occupations in the last year, peers' final field of study choices). If we invert the sequence of introduction of the variables that refer to subject preferences, occupational preferences, career orientations, and peer influences, the pattern of results does not change and our substantive conclusions are confirmed. The same conclusion holds if we enter occupational preferences in high school before track choice (assuming that these preferences are a proxy of unmeasured preferences in junior school, i.e., before choosing the secondary track), as reported in the [online appendix](#) web address to be reported.

Concluding remarks

This article has used fresh, longitudinal data on a recent cohort of Italian high school seniors to assess seven potential explanations for gender differences across fields. Let us summarize our three main sets of results. First, skill-based explanations referring to gender differences in scientific subjects or the comparative advantage in humanistic subjects do not hold any explanatory power: school performance in math affects field of study choices, but in recent cohorts gender differentials are negligible in this domain, at least among high school seniors (*see* footnote 7). Similarly, the higher the importance attached to future career prospects, the lower are the chances of choosing the humanities and social sciences, but this factor fails to account for GSHE because girls are not less career-oriented than boys. Overall, this first set of results suggests that rational choice explanations for GSHE do not find empirical support. At least in this recent cohort of high school graduates, girls do not seem to experience any lack of relevant skills or of ambition that would prevent them from enrolling in more rewarding fields.

Second, expressive motivations related to preferences for school subjects and for specific occupations mediate to a significant extent GSHE because these two factors are strongly patterned along gender lines and affect field of study choices. Additionally, our results suggest that the educational choices of the closest classmates correlate with respondents' own choices, although these factors do not mediate GSHE. Overall, our analyses suggest that culturalist approaches stressing the role of expressive preferences over and above cost-benefit calculations display significant heuristic value. Their importance is stronger for gender differences in engineering and ICT than for the humanities and social sciences.

Our third and most important result concerns the key role of curricular track choices. This single factor mediates 64% of the gender differences in access to the humanities and social sciences and 29% of the gender differences in access to engineering and ICT. This is because curricular choices in high school are heavily segregated along gender lines and curricular track displays a strong influence on field of study choices. Selecting a specific track in high school is a choice which has several important consequences, such as being exposed to the following: a specific curriculum and thus developing more interest in specific disciplines; a specific form of anticipatory socialization shaping the range of potential occupations that they envisage; different classmates with their own educational and occupational preferences which can affect

student's own preferences. Indeed, our results indicate that the strong influence of curricular track is partly mediated by the expressive preferences for school subjects and for specific occupations and to a minor extent by peer influences. However, even after taking into account these determinants of field choices, the effects of track choices remain strong. We suspect that the effect of secondary curricula reflects also the different chances to succeed in different fields afforded by these curricula. More generally, a limitation of our data is that they do not contain information on the subjective beliefs concerning the chances to succeed in different fields nor about their perceived profitability, which could contribute to further account for gender differences across tertiary fields (Barone et al. 2017).

With the data at hand, we cannot examine what shapes curricular choices at early stages. As suggested in Section 2, there is evidence from the literature that the influences of significant adults (teachers, parents, counselors) play a significant role, particularly at early educational transitions, when students are younger. However, the social norms with regard to the appropriateness for men and women of various educational fields may not be very strong: for instance, students choosing gender-atypical fields do not seem to receive less support from their parents (Mastekaasa and Smeby 2008). A limitation of this study is that we do not have direct information on significant other's influences.

Moreover, it is possible that earlier expressive preferences for specific subjects and occupations influenced secondary track choices. If these earlier preferences affected track choice, and if they exert a direct impact on field choice net of track choice, this would lead us to overestimate the role of track choices. More generally, it is important to stress that our results refer to one specific educational transition, namely, access to university education. Our longitudinal data do not cover the earlier stages of the educational career and thus we cannot assess the dynamic evolution of the mechanisms that we consider. In this study, we can only observe the net influences of each of the above-mentioned mechanisms as they operate at the end of upper secondary education. This may explain why we are not able to entirely account for gender differences across fields, which is another limitation of our study.

Even if we cannot analyze what lies behind the influence of secondary curricula, the finding that this factor accounts for a substantial part of GSHE seems important to reorient the scientific and policy debate on this matter, because it suggests that GSHE is to a substantial extent *predetermined* well before students complete high school and choose a field of study. This conclusion differs from the results concerning the USA, where upper secondary education is not tracked and curricular choices are more fluid. Our results may not be mechanically transferred to other countries, but we would expect to observe similar patterns in other European countries, where school-based tracking is common. At the macro-level, there is already evidence that GSHE is stronger where early tracking is prevalent (Imdorf et al. 2015; Smyth and Steinmetz 2008). The scientific debate on this matter that we have discussed in Section 2 seems too focused on the proximate mechanisms operating when students must choose a field of study. Future research should devote more attention to the role of curricular tracking for GSHE in European contexts and to the mechanisms that lie behind it.

Future developments of our work involve also a deeper understanding of the interplay between the different determinants of field of study choices. Since our research questions focus on the mediating mechanisms of GSHE, we carried out a standard mediation analysis, but it is possible that the influence of different determinants of field choices varies across genders, thus opening the room for a moderation analysis. It seems also important to keep in mind that, while our modeling approach is unidirectional in suggesting a relation between track choice and field of study, reality is necessarily more complex, in that it involves multi-directional

progressions, which can be at the same time reversible and open to anticipation effects, whereby occupational plans have retroactive influences on anticipated field of study choice and/or track choices (Xie and Shauman 2005, pp. 57–60).

As regards the policy implications, many interventions to reduce GSHE take the form of educational and counseling activities trying to enhance girls' interest toward STEM fields of study and the related occupations or to raise the awareness of gender stereotypes on the side of students, teachers, and parents in high school (Timmers et al. 2010). The risk is that these interventions arrive too late, if they target only university choices and leave gendered curricular choices unchanged.

References

- Aastrup, W. (Ed.). (2007). *Guidance and counselling in higher education in European Union member states*. Aarhus: Narayana Press.
- AlmaLaurea (2015). XVII Indagine sulla Condizione Occupazionale dei Laureati. Bologna: Consorzio Interuniversitario AlmaLaurea.
- Alon, S., & DiPrete, T. A. (2017). Gender differences in the formation of a field of study choice set. *Sociological Science*, 2, 50–81.
- Anvur. (2016). *Rapporto sullo Stato del Sistema Universitario e della Ricerca*. Roma: Anvur.
- Argentin, G., & Triventi, M. (2016). Come mi giudichi? Analisi delle pratiche e degli standard di attribuzione dei voti agli studenti nelle scuole italiane. In P. Falzetti (Ed.), *Concorso idee per la ricerca*. Padova: Cleup.
- Assirelli, G. (2015). Credential and skill mismatches among tertiary graduates: the effect of labour market institutions on the differences between fields of study in 18 countries. *European Societies*, 17(4), 535–568.
- Barone, C. (2011). Some things never change: gender segregation in higher education across eight nations and three decades. *Sociology of Education*, 84(2), 157–176.
- Barone, C., & Ortiz, L. (2011). Overeducation among European university graduates: a comparative analysis of its incidence and the importance of higher education differentiation. *Higher Education*, 61(3), 325–337.
- Barone, C., Schizzerotto, A., Abbiati, G., & Argentin, G. (2017). Information barriers, social inequality, and plans for higher education: evidence from a field experiment. *European Sociological Review*, 33(1), 84–96.
- Bobbitt-Zeher, D. (2007). The gender income gap and the role of education. *Sociology of Education*, 80(1), 1–22.
- Breen, R., Karlson, K. B., & Holm, A. (2013). Total, direct, and indirect effects in logit and probit models. *Sociological Methods & Research*, 42(2), 164–191.
- Buchmann, C., DiPrete, T. A., & McDaniel, A. (2008). Gender inequalities in education. *Annual Review of Sociology*, 34(1), 319–337.
- Cattaneo, M., Horta, H., Malighetti, P., Meoli, M., & Paleari, S. (2017). Effects of the financial crisis on university choice by gender. *Higher Education*, 74(5), 775–798.
- Cavalli, A., & Facchini, C. (2001). *Scelte cruciali*. Bologna: Il Mulino.
- Cech, E. (2013). The self-expressive edge of occupational sex segregation. *American Journal of Sociology*, 119(3), 747–789.
- Ceci, S. J., & Williams, W. M. (2010). Sex differences in math-intensive fields. *Current Directions in Psychological Science*, 1(5), 11–23.
- Charles, M., & Bradley, K. (2002). Equal but separate? A cross-national study of sex segregation in higher education. *American Sociological Review*, 67(4), 573–599.
- Charles, M., & Bradley, K. (2009). Indulging our gendered selves? Sex segregation by field of study in 44 countries. *American Journal of Sociology*, 114(4), 924–976.
- Charles, M., & Grusky, D. B. (2004). *Occupational ghettos*. Stanford: Stanford University Press.
- Correll, S. J. (2004). Constraints into preferences: gender, status and emerging career aspirations. *American Sociological Review*, 69(1), 93–113.
- Davies, S., & Guppy, N. (1997). Fields of study, college selectivity, and student inequalities in higher education. *Social Forces*, 75(4), 1417–1438.
- De Giorgi, G., Pellizzari, M., & Redaelli, S. (2010). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics*, 2, 241–275.
- DiPrete, T. A., & Buchmann, C. (2013). *The rise of women: the growing gender gap in education and what it means for American schools*. New York: Russell Sage Foundation.

- Ellison, G., & Swanson, A. (2010). The gender gap in secondary school mathematics at high achievement levels: evidence from the American mathematics competitions. *Journal of Economic Perspectives*, 24(2), 109–128.
- England, P., & Li, S. (2006). Desegregation stalled: the changing gender composition of college majors, 1971–2002. *Gender & Society*, 20(5), 657–677.
- Frank, K. A., Muller, C., Schiller, K. S., Riegle-Crumb, C., Mueller, A. S., Crosnoe, R., & Pearson, J. (2008). The social dynamics of mathematics course-taking in high school. *American Journal of Sociology*, 113(6), 1645–1696.
- Gabay-Egozi, L., Shavit, Y., & Yaish, M. (2015). Gender differences in fields of study: the role of significant others and rational choice motivations. *European Sociological Review*, 31(3), 284–297.
- Garcia-Aracil, A., Gabaldon, D., Gines Mora, J., & Vila, L. (2007). The relationship between life goals and fields of study among young European graduates. *Higher Education*, 53(6), 843–865.
- Gerber, T. P., & Schaefer, D. R. (2004). Horizontal stratification of higher education in Russia: trends, gender differences, and labour market outcomes. *Sociology of Education*, 77(1), 32–59.
- Gunderson, E. A. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles*, 66, 153–166.
- Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 269(5220), 41–45.
- Imdorf, C., Hegna, K., Eberhard, V., & Doray, P. (2015). Educational systems and gender segregation in education—a three-country comparison of Germany, Norway & Canada. In C. Imdorf, K. Hegna, & L. Reisel (Eds.), *Gender segregation in vocational education* (Vol. 32, pp. 83–122). Bingley: Emerald Insight.
- Invalsi (2017). Rilevazioni Nazionali degli Apprendimenti—2016. http://www.invalsi.it/invalsi/doc_evidenza/2016/07_Rapporto_Prove_INVALSI_2016.pdf
- Jacobs, J. A. (1995). Gender and academic specialties: trend among recipients of college degrees in the 1980s. *Sociology of Education*, 68(2), 81–98.
- Jonsson, J. O. (1999). Explaining sex differences in educational choice: an empirical assessment of a rational choice model. *European Sociological Review*, 15(4), 391–404.
- Konrad, A. M., Ritchie, J. E., Jr., Lieb, P., & Corrigan, E. (2000). Sex differences and similarities in job attribute preferences: a meta-analysis. *Psychological Bulletin*, 126(4), 593–641.
- Legewie, J., & DiPrete, T. A. (2012). School context and the gender gap in educational achievement. *American Sociological Review*, 77(3), 463–485.
- Legewie, J., & DiPrete, T. A. (2014). The high school environment and the gender gap in science and engineering. *Sociology of Education*, 87(4), 259–280.
- Mann, A., & DiPrete, T. A. (2013). Trends in gender segregation in the choice of science and engineering majors. *Social Science Research*, 42(6), 1519–1541.
- Mastekaasa, A., & Smeby, J. (2008). Educational choice and persistence in male- and female-dominated fields source. *Higher Education*, 55(2), 189–202.
- Menon, M. E., Markadjijs, E., Theodoropoulos, N., & Socratous, M. (2017). Influences on the intention to enter higher education: the importance of expected returns. *Journal of Further and Higher Education*, 41(6), 831–843.
- Morgan, S. L., Gelbgiser, D., & Weeden, K. A. (2013). Feeding the pipeline: gender, occupational plans, and college major selection. *Social Science Research*, 42(4), 989–1005.
- Nunez, I., & Livanos, I. (2010). Higher education and unemployment in Europe: an analysis of the academic subject and national effects. *Higher Education*, 59(4), 475–487.
- OECD. (2014). *The ABC of gender equality in education*. Paris: OECD.
- OECD. (2016). *Education at a glance—2016 edition*. Paris: OECD.
- Ost, B. (2010). The role of peers and grades in determining major persistence in the sciences. *Economics of Education Review*, 29, 923–934.
- Polachek, S. W. (1978). Sex differences in college major. *Industrial and Labor Relations Review*, 31, 498–508.
- Reimer, D., Noelke, C., & Kucel, A. (2011). Labor market effects of field of study in comparative perspective: an analysis of 22 European countries. *International Journal of Comparative Sociology*, 49(4), 232–256.
- Riegle-Crumb, C., King, B., Grodsky, E., & Muller, C. (2012). The more things change, the more they stay the same? Prior achievements fails to explain gender inequality in entry into STEM college majors over time. *American Educational Research Journal*, 49(6), 1048–1073.
- Smyth, E., & Steinmetz, S. (2008). Field of study and gender segregation in European labour markets. *International Journal of Comparative Sociology*, 49(4), 257–281.
- Timmers, T., Willemsen, T., & Tijdens, K. (2010). Gender diversity policies in universities: a multi-perspective framework of policy measures. *Higher Education*, 59(6), 719–735.
- Triventi, M. (2010). Something changes, something not. Long-term trends in gender segregation of fields of study in Italy. *Italian Journal of Sociology of Education*, 5(2), 47–80.

- Vaarmets, T. (2018). Gender, academic abilities and postsecondary educational choices. *Journal of Applied Research in Higher Education*, 10(3), 380–398.
- van de Werfhorst, H. G. (2017). Gender segregation across fields of study in post-secondary education: trends and social differentials. *European Sociological Review*, 33(3), 449–464.
- van de Werfhorst, H. G., Sullivan, A., & Cheung, S. Y. (2003). Social class, ability and choice of subject in secondary and tertiary education in Britain. *British Educational Research Journal*, 29(1), 41–62.
- Xie, Y., & Shauman, K. A. (2005). *Women in science: career processes and outcomes*. Cambridge: Harvard University Press.
- Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources*, 48(3), 545–595.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.