World Development 195 (2025) 107080

Contents lists available at ScienceDirect

World Development

journal homepage: www.elsevier.com/locate/worlddev

Regular Research Article

Education and ethnic intermarriage: evidence from higher education expansion in Indonesia

Antonio Di Paolo [a,*], Khalifany-Ash Shidiqi [b]

[a] Universitat de Barcelona & AQR-IREA, Spain
[b] Department of Economics, Universitas Muhammadiyah Yogyakarta, Indonesia

ARTICLE INFO

JEL classifications:
I21
I23
J12

Keywords:
Education
Interethnic marriages
Higher education expansion
Indonesia

ABSTRACT

We analyse the effect of educational attainments on interethnic marriages in Indonesia, a multi-ethnic emerging country. The empirical analysis is based on data from the Java Island obtained from the 2014 wave of the Indonesian Family Life Survey, combined with administrative data about the location and year of establishment of Higher Education Institutions (HEI). To estimate causal effects, we exploit variation in exposure to HEI by birth year and district of residence in an IV/TSLS framework. Specifically, we employ as instrument for education the number of HEI located in a radius of 10 km from the centroid of the district of residence at age 18. The analysis is carried separately for males and females. The results indicate that years of schooling, college attendance and completion positively affect the likelihood of exogamy, i.e. having a partner from a different ethnicity. The estimates are somewhat larger for females than for males (although not statistically different), and all the robustness checks provide stable results, supporting their causal interpretation. The effect of schooling is not heterogeneous depending on parental education or mixed parental ethnicity. However, it is lower for individuals with Javanese ethnicity compared to those belonging to other ethnic groups. We also analyse potential mechanisms, highlighting that migration/residential location and changes in social norms could be significant channels underlying the causal chain between higher education expansion, educational attainments, and interethnic marriages. Overall, the results point out that the increase in educational attainments induced by the expansion of higher education could contribute to the reduction of ethnic segregation.

1. Introduction

Education generates several positive effects both at the individual and aggregate levels. The increase in the endowment of human capital is especially important for developing countries, since it shapes economic growth and development (Barro, 2001). Indeed, the governments of several developing countries have undertaken a diverse range of policies to enhance the formation of human capital during the last decades. These policies typically encompass large-scale interventions such as the extension of compulsory schooling and the expansion of educational infrastructures, at the primary, secondary and, especially more recently, at the tertiary education level, following the patterns that have been experienced by developed countries. Indeed, fostering education through the expansion of Higher Education Institutions (HEI) is an effective policy to enhance economic growth (Valero and Van Reenen, 2019). In this paper, we focus on a specific impact of the increase of educational attainments induced by the expansion of HEI: ethnic intermarriages in a multiethnic developing country (Indonesia).

Understanding whether, and to what extent, higher education attainments increase (in a causal sense) the likelihood of interethnic marriages (i.e. exogamy) is relevant in ethnically mixed societies for of several reasons. First, the rate of ethnic intermarriage is a clear indicator of ethnic attachment, which is strongly related to ethnic fractionalization and ethnically-related socioeconomic segregation (Bazzi et al. 2019; Kukić, 2023). Second, lower levels of ethnic fractionalization and segregation can mitigate civil conflicts, which in turn would favour economic development (Esteban et al., 2012; Corvalan & Vargas, 2015; Sanjaya et al., 2023). Indeed, these potential impacts could be relevant channels through which education is likely to a) reduce conflict (Rohner & Saia, 2019) and b) increases interethnic tolerance and diversity in general (Roth & Sumarto, 2015). Therefore, analysing the effect induced by HEI expansion would provide evidence regarding whether this policy is an effective tool to achieve the aforementioned goals.

From the theoretical point of view, there are several possible

* Corresponding author at: Universitat de Barcelona, Faculty of Economics and Business, 690 Diagonal Av. 08034, Barcelona, Spain.
E-mail addresses: antonio.dipaolo@ub.edu (A.D. Paolo), khalifany.ash@umy.ac.id (K.-A. Shidiqi).

https://doi.org/10.1016/j.worlddev.2025.107080
Accepted 1 June 2025
Available online 2 July 2025

justifications for the existence of a positive causal relation between educational attainments and interethnic marriages. First, the (Indonesian) educational system promotes a shared national identity and the adoption of a single language (Bahasa Indonesia, also known as standard Indonesian) and a unitary culture. Indeed, this is in line with the existing papers about the role of education on the formation of identity (e.g. Bandiera et al. 2019; Alesina et al. 2021). In addition, higher education leads to a prolonged exposure to Bahasa Indonesia, improving language proficiency and reducing language barriers, which may, in turn, increase the likelihood of engaging in an interethnic partnership. Second, education might change cultural and social norms, mitigating the degree of attachment to traditional (and possibly ethnically segregated) values, thus favouring interethnic tolerance (Roth and Sumarto, 2015). Third, education increases earnings potential and therefore fosters financial autonomy, thus limiting the degree of dependency on the family, which could be especially important for women living in matrilocal enclaves.[1] Finally, more educated individuals have a higher propensity to migrate, possibly to larger agglomerations characterized by a higher degree of ethnic diversity, which could affect the likelihood of finding a partner from a different ethnic background.

There is a large body of literature regarding the determinants and the socioeconomic effects of ethnic/racial intermarriages for developed countries (mostly the US), mainly focused on first-and-second generation migrants (for a review, see Furtado & Song, 2022). However, despite the relevance of the topic, there is a clear lack of evidence regarding the causal relationship between education and interethnic marriages in multiethnic developing countries. Some recent work has been focused on the determinants of intermarriage (not exclusively on education) in developing countries. For example, Ray et al. (2020) analysed the association between inter-caste marriages and husband's, wife's and parents' education in India. Allendorf and Thornton (2015) examines the determinants of inter-caste marriages in Nepal, including education as an explanatory variable. More recently, Crespin-Boucaud (2020) and Bandyopadhyay and Green (2021) studied the determinants of interethnic marriages in Sub-Saharan countries, both using data from the Demographic and Health Surveys (DHS) but using different methodologies. Nevertheless, these papers are not specifically focused on the causal effect of education on the formation of ethnically-mixed couples.

There are also few papers for the case of Indonesia. The most relevant one is Bazzi et al. (2019), which exploits a large-scale population resettlement program that took place in Indonesia during the'80 (the so-called Transmigration Program) to investigate the causal effect of intergroup contact on national integration. Although educational attainments are not the focus of the paper, the authors consider interethnic marriages as one of the proxies for national integration and show that the interethnic marriage rate is negatively affected by ethnic polarization. There are also other papers about ethnic intermarriage in Indonesia and its determinants (Utomo, 2019; Utomo & McDonald, 2016; 2021), which also consider the association with education. However, these papers do not address the issue of causality, which represents our main contribution to the literature.

In this paper, we analyse the causal effect of educational attainment on the probability of being engaged in an interethnic marriage in Indonesia. As such, this is the first work that provides plausibly causal evidence on this topic, which represents the main contribution of our work to the existing literature. To achieve identification, we leverage on the geographical expansion of Higher Education Institutions that took place in Indonesia, especially in the island of Java (where we focus), since the second half of the 20th century. Therefore, we also contribute to the evidence regarding the effects of investment in educational infrastructures (Duflo, 2001 and related papers), as well as to the growing literature about the local effect of college expansion (Jagnani and Khanna, 2020; Carneiro et al., 2023, among others[2]), with an additional piece of evidence for an emerging country. Moreover, we also provide suggestive evidence regarding potential mechanisms that could be at play in the causal chain between HEI expansion, educational attainments and interethnic marriages, which is an additional value added by this paper. More generally, we contribute to the body of evidence highlighting the role of education as a tool to reduce ethnic-related segregation in multiethnic developing countries.

The empirical analysis integrates various data sources. First, we employ administrative data regarding the year of establishment and exact location of each Higher Education Institutions that provides undergraduate education in the island of Java, the most populated island of Indonesia. A notable aspect of this data is its disaggregation at the campus level, considering the possibility of multiple locations for each institution. Second, we draw on data from the Indonesian Family Life Survey (IFLS). Our primary focus is on the latest available wave in 2014, supplemented by relevant information from preceding waves for specific analytical purposes. Based on information about individual ethnicity and households' identifiers, we can create an indicator for exogamy, that it, the ethnicity of one member of the couple differs from that of the other. This is the main outcome variable of our analysis. Moreover, ILFS data includes details about the district of residence and provides a comprehensive residential history dating back to the year of birth. Therefore, we are able to impute the geographical exposure to available HEI during different stages of adolescence, based on the individual's district of residence. This serves as the basis for constructing our Instrumental Variable (IV) to address the endogeneity of educational attainments in the exogamy equation. More specifically, we instrument education with the number of HEIs present in a radius of 10 km from the district of residence of the individual at age 18. We leverage on variation in geographical exposure to HEI across cohorts and locations, exploiting the expansion of HEI that took place over time in the island of Java. The model that explains the probability of being engaged in an interethnic marriage is separately estimated for males and females. We test for the robustness of the results to the definition of the instrument, particularly with respect to age and radii of exposure. Most importantly, we perform several sensitivity checks to discard the possibility that the instrument captures spurious correlations driven by either unobserved time-varying local factors – related to the demand for higher education – that could be correlated with the propensity for interethnic marriages. Moreover, we conduct various robustness checks to rule out the possibility of endogenous residential sorting of individuals and their families during the educational process—that is, whether families relocate to areas with greater access to higher education institutions to facilitate their children's university education. Furthermore, we test for possible heterogeneous effects of educational attainments on the probability of being married to someone from a different ethnic background. Finally, we provide suggestive evidence regarding the role of possible mechanisms behind the causal chain between HEI exposure, educational attainments and interethnic marriages. Specifically, we examine the potential relevance of migration/residential locations and social norms related to ethnicity. Indeed, this analysis of mechanisms constitutes another significant contribution of our work to the existing literature.

The results indicate that higher educational attainments, induced by

[1] Matrilocality is a social system in which the couple lives in the neighbourhood of the wife or wife's family after marriage. This is different from patrilocality, where the wife moves to her husband's neighbourhood or husband's family. Matrilocality is often associated with matriarchal societies, where women have a central role in the social structure and family decisions.

[2] We are not the first in using college expansion as an instrumental variable to address the endogeneity of educational attainment. Starting from the paper by Currie and Moretti (2003), this approach has been used in several recent works (Kyui, 2016; Kamhöfer et al. 2019; Belskaya et al. 2020; Westphal et al. 2022). In the empirical methodology section, we carefully describe similarities and differences between our identification strategy and the framework adopted in previous papers.

the expansion of HEI, have a positive impact on the likelihood of being in an interethnic marriage. Following Currie and Moretti (2003) and Jagnani and Khanna (2020), among others, we consider different measures for educational attainments: years of schooling, university enrolment and university completion. A positive effect on interethnic marriage is observed for the three outcomes and is somewhat higher for females than for males, although the estimated coefficients are not statistically different by gender. The results are very robust to all the sensitivity checks, pointing to the validity of the underlying assumption behind our IV approach. The analysis of heterogeneous effects highlights that the impact of education on the probability of having a partner from a different ethnic background is the same regardless of parental education and having parents with mixed ethnicities. However, increased educational attainments induced by HEI expansion exert a lower effect on exogamy for individuals with Javanese ethnicity than their counterparts from other ethnic backgrounds. This suggests that education could be a tool to mitigate segregation of ethnic minorities. Finally, the evidence regarding potential mechanism highlights the relevance of both dimensions. On the one hand, individuals who are more educated are more likely to migrate to and reside in large cities, with a higher degree of ethnic diversity, thereby increasing the likelihood of exogamy. On the other hand, the increase in educational attainments induced by the expansion of HEI fosters trust towards individuals from different ethnic backgrounds (our proxy for social norms), which could lead to a higher propensity to form an ethnically-mixed couple.

Overall, the results presented in this paper highlight the relevance of education, and the expansion of higher education, as tools for promoting the social integration of individuals from diverse ethnic backgrounds. Therefore, the beneficial effects on human capital formation induced by the establishment of new HEI could not only materialize in positive impacts in terms of earnings and other labour market outcomes, but can also enhance other social outcomes and, more generally, can mitigate ethnic-related segregation in multiethnic countries and foster social cohesion.

The rest of this paper proceeds as follows: Section 2 summarizes the institutional background regarding ethnicities and marriage in Indonesia, as well as about its education system. Section 3 contains a description of the data used in the empirical analysis and presents some descriptive evidence. Section 4 illustrates the empirical strategy, and Section 5 reports the results. Finally, Section 6 concludes.

2. Institutional background

2.1. Ethnicity and interethnic marriages in Indonesia

Indonesia, with a population of over 240 million, is one of the world's most populous countries. It is also an extremely rich and diverse country from the cultural point of view. Its major religion is Islam, although several other religions coexist. Moreover, the inhabitants of Indonesia belong to a wide and diverse range of ethnic groups, each with its own set of cultural norms and traditions. In Indonesia, ethnicity is largely assigned based on language (Rademakers and van Hoorn, 2021), with minimal variations in terms of physical appearance in the majority of instances. Moreover, the ethnic diversity in Indonesia offers a fascinating chance to explore the interaction between ethnicity, culture, and family dynamics, specifically regarding choices for marriage and family formation.

Every ethnic group deeply values marriage as it represents the union of two individuals and their families. These ceremonies celebrate and maintain the diverse cultural heritage and ethnic identities by following specific ethnic traditions (Buttenheim and Nobles, 2009). Meanwhile, the practice of interethnic marriage encounters notable challenges. For instance, Parker et al. (2014) explored how ethnic and religious groups in Indonesia interact, from socializing to marriage. They observed strong resistance to interreligious relationships, impacting even casual dating, largely due to strict religious teachings. While Indonesian society shows increasing acceptance of interethnic relationships, endogamy remains the most common practice.

The island of Java, the focus of our study and the most densely populated island in Indonesia, is largely inhabited by the Javanese people, who make up more than 55 % of its population. They predominantly reside in Central Java, D.I. Yogyakarta, and East Java Province. The Sundanese, constituting around 25 % of the population, primarily reside in West Java. The Betawi and Madurese, with approximately 5 % of the populace each, are primarily concentrated in Jakarta and Madura Island, situated immediately north of East Java, respectively. The remaining portion of the population, approximately 10 %, comprises various minority ethnic groups (Statistics Indonesia, 2010). Considering this demographic context, Utomo and McDonald (2016) found notable disparities in marriage trends between Jakarta, the primary urban and economic hub, which displayed the lowest propensity for endogamous marriages at 67 %, and regions heavily influenced by Javanese culture, where this rate surpasses 95 %. According to Utomo (2019), Jakarta has lower rates of endogamy since it serves as a hub for migrants and represents a place where different cultures mix together. The city's heterogeneous population, particularly in its higher education institutions that attract students from across the entire country, favour the formation of interethnic partnerships and marriages. More broadly, the relative size of ethnic groups within local communities, along with migration patterns, plays a significant role in shaping (or diminishing) the frequency of ethnically mixed marriages. This is because endogamy rates within a given ethnic group are positively associated with the fraction of co-ethnic residents and negatively associated with the local stock of migrants (Utomo and McDonald, 2021). Beyond sociodemographic factors, social norms and cultural traditions also serve as key determinants of the prevalence of endogamous marriage in multiethnic countries such as Indonesia. For example, Utomo (2019) also highlights that individuals do not engage in random marriage pairings, but instead take ethnicity into consideration as a significant aspect in their decision-making process. In general, the primary challenge in interethnic unions often lies in adapting to the spouse's customs, traditions, and culture, as well as the strictness of customs. These strict traditional norms often lead to a preference for marrying within the same ethnicity (Ida Bagus, 2008; Parker et al., 2014).

2.2. The educational system and higher education in Indonesia

Indonesia's educational system follows the 6-3-3-4 model, that is, 6 years of elementary school, 3 years of junior high, 3 years of senior high, and up to 4 years of higher education (Mukminin et al., 2019). The system of higher education is composed of vocational degrees, whose duration ranges between one and four years, and undergraduate degrees, which typically consists of four-year programs. After completing their undergraduate studies, graduates can pursue either a two-year master's degree or a doctoral program, which typically lasts three to five years.

Indonesia's Higher Education Institutions (HEI) include universities, institutes, colleges, polytechnical institutes, and academies, which can be either public or private. Public institutions are funded through public subsidies and tuition fees. The funding of private institutions relies primarily on tuition fees and other financing sources. Public HEIs are under the authority of a government-appointed administration and adhere to stringent regulations. In contrast, private HEIs have greater independence in their governance and management, although they may encounter varying degrees of government influence that impact their funding, governance, and regulatory supervision (Welch, 2007; Ngo and Meek, 2019). In general, for both types of HEI, the enrolment cost paid by students varies according to the institution (and its quality) and the field of study. However, when enrolling in private institutions, students also have to pay an entry fee, which is not fixed and is specific to each institution and study program. On average, the overall cost paid by students is generally higher in private institutions, although there could

be specific undergraduate degrees that are more expensive in prestigious public institutions than in less renowned private centres.

From the historical perspective, the expansion of HEIs in Indonesia began immediately after the country achieved independence in 1945. Just between 1945 and 1950, national student enrolment in higher education increased from 1600 to 5200 (Buchori and Malik, 2004). The Higher Education Act of 1961 was one of the first substantial advances the newly independent nation made. DGHE (2003) outlined that this legislation established the foundation for future HE advancements and brought about significant improvements. Following this new law, HEI adopted an ordered framework with a precise division of faculties. The legislation defined the requirements for establishing universities, colleges, academies, and other HEIs.

HEIs are established through different processes, depending on whether they are public or private. Public institutions are opened through a public procedure (and inaugurated directly by the President of the Republic of Indonesia), while private institutions are typically initiated by private corporations or foundations, which are obliged to inform the Ministry of Education of their intent (Welch, 2007). This notification requires the submission of a notarial deed confirming the existence of a legal entity governing the HEI, its articles of association, assets, expected sources of funding for its operation, curricular plans, and a complete description of each faculty member's credentials and teaching position. The government supervises and guides private HEIs to ensure quality and compliance with standards, through an agency called the Private Higher Education Coordinator (KOPERTIS). This agency, led by the Minister of Education, is present in all Indonesian provinces (Buchori and Malik, 2004). In terms of admission to undergraduate degree programs, initially the only requirement was a senior high school diploma. To unify standards, the government and the major public HEIs in the island of Java implemented a general admissions test in 1976 (SKALU). The admission system changed in 1989 (UMPTN), mostly because of the introduction of specialized exams based on the chosen major. On the contrary, private HEIs have maintained independent admission processes at the college level, without a unified testing system.

The number and variety of Indonesian HEIs have grown significantly since the HE Act was enacted in 1961. According to Pannen et al. (2018), there were 450 HEIs in 1970, with a student population of 237 thousand. However, by 1990, the number had risen dramatically to 900 schools, serving nearly 1.5 million students. Fig. 1 depicts the number of public and private HEIs offering undergraduate degrees in the island of Java, by year of establishment. From 1945 to the mid-1960s, the development of both public and private HEIs was relatively moderate and steady. Around the mid-1960s, there was a pronounced increase in the establishment of public HEIs, which continued to grow steadily during the following decades. Private HEIs, however, experienced a constant rise during the'70s, but a sharp increase during the'80s. The increase in the number of private HEIs was more moderate, although still very pronounced, during the following decades. At the end of the 20th century, there were more than double the number of private HEIs than public HEIs in the island of Java. Buchori and Malik (2004) argued that the rapid growth of private HEIs in the 1980s was driven by the increasing demand for HE that emerged in the 1970s. During this period, the state's budget was insufficient to satisfy this demand (Ngo and Meek, 2019). Notably, private foundations or organizations responded by creating schools such as universities, institutes, colleges, polytechnics, and academies, which provide a variety of programs and degrees.

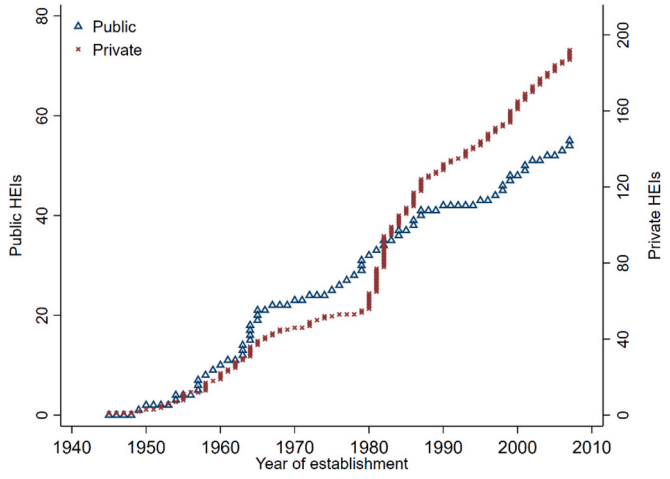

Fig. 1. Year of establishment (public and private HEI).

Fig. 2 display the temporal evolution of the geographical location of public and private HEIs, again focussing on institutions that offer undergraduate programs. In 1960, the few existing HEIs were concentrated in the major urban centres, notably Jakarta, Bandung, Semarang, Yogyakarta, and Surabaya. From 1980 to 1995, the sector of higher education expanded considerably, with public institutions increasingly present in medium and large agglomerations. Nevertheless, throughout this period, many private institutions emerged, both in urban centres and in small towns. At the end of the relevant period (2007[3]), the presence of HEIs was more widely spread, geographically, providing a general coverage of all Java's provinces, especially due to the extensive expansion of private institutions.

3. Data and descriptive statistics

Our empirical analysis focuses on the island of Java, the most populated island of Indonesia and the island in which its capital and most populous city – Jakarta – is located. Indeed, the majority of higher educational institutions are located in the island of Java (PDSP Kemdikbud, 2013). We combine different data sources. First, we employ data regarding all HEIs obtained from the National Accreditation Body for Higher Education (BAN-PT). This dataset includes information about the exact location of each campus for both public and private HEIs, the year of establishment, as well as details on the type of higher education offered by each institution and their accreditation status. For the empirical analysis, we retain only institutions offering undergraduate degrees that achieved a minimum accreditation score.[4] The sites of the campuses of the HEIs have been geolocated using their detailed addresses (see Fig. 2).

Second, we use individual and family-level data from the Indonesian Family Life Survey (IFLS) database,[5] which is representative of more than 80 % of the Indonesian population within the survey area (Strauss et al., 2016). We mostly use data from the last wave of 2014, although we also exploit information from previous waves for specific purposes. The survey provides information about several individual and parental characteristics, including detailed information about educational attainments. Most importantly, the last two waves (2014 and 2007) of the IFLS database contain information about the respondent's ethnicity, as well as the ethnicity of his/her parents. The questionnaire includes 29 different ethnicities, representing the large majority of ethnic groups in terms of the country's population. Thanks to household identifiers, we

[3] We consider 2007 as the end of the relevant period because, as explained in what follows, we mainly consider exposure to HEI at age 18, and the youngest individual in our estimation sample turned 18 in that year.

[4] Based on the BAN-PT (National Accreditation Board for Higher Education) Regulation No. 2 of 2017, which details the mechanisms for accreditation, Higher Education Institutions in Indonesia are evaluated and classified into three categories of accreditation: A (excellent compliance with the standards), B (good compliance with the standards), and C (represents the minimum fulfilment of national standards).

[5] IFLS data are freely available from this link: https://www.rand.org/well-being/social-and-behavioral-policy/data/FLS/IFLS.html.

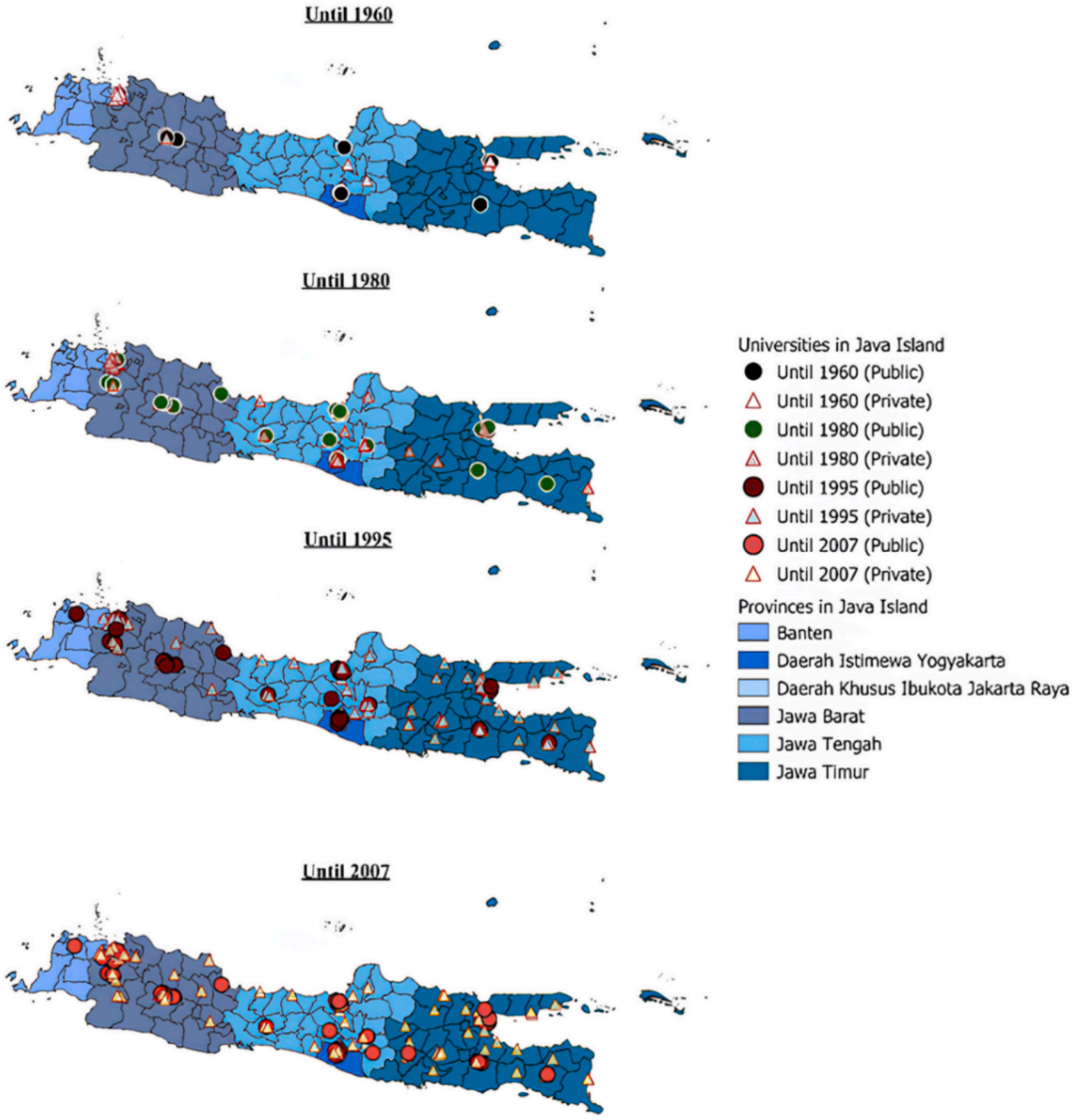

Fig. 2. The geographical location of hei on java island over time.

are able to construct our outcome variable, exogamy, which is an indicator for having a partner from a different ethnic background. We consider several measures of educational attainment. Specifically, we use, as explanatory variables of interest, a) years of schooling, b) college attendance and c) college completion These variables have been constructed combining information about the highest grade attended and the highest completed grade.[6]

Moreover, the IFLS database also includes information about the place of birth and the current place of residence, defined according to two main administrative geographical units – provinces and districts – as well as the entire migration history. Given the lack of information about the precise place of residence of households within the districts, we combine the two data sources based on the centroids of the districts. Specifically, as better explained in the next section, we construct different measures of geographical exposure to HEI during adolescence. These are defined according to the number of HEIs located within a certain distance ("radius") from the district's centroid, covering the period from the year of birth until the year in which the individual turned 18 years old.[7]

The estimation sample has been obtained by retaining married individuals with age comprised between 25 and 65. In this way, we avoid including individuals who could be still studying, and limit selection issues related to the age at marriage.[8] Moreover, we also exclude older individuals due to potential issues of selective mortality. We retain only individuals who were born in Java and lived on the island throughout the entire relevant period, due to the design of our identification strategy and the robustness checks we implement to validate its underlying

[6] That is, if an individual's highest level of education is junior high school and his/her highest grade ever completed is 2, then we impute 8 years of schooling. Furthermore, the indicator for college attendance is equal to one if an individual attended at least one year of college, while the indicator for college completion takes the value 1 if the individual attended and completed college.

[7] Actually, in order to perform a robustness check for our empirical framework, we also consider exposure to HEI at age 25.

[8] According to the World Bank (2023), the average age at marriage in Indonesia is 27.1 and 22.4 for male and female respectively. Notice that, using information about the year of marriage, we also perform a robustness check in which we only retain individuals who got married after completing education.

assumptions. Finally, we also exclude observations with missing values in the variables of interest. After applying these conditions, we obtain a sample of 6352 males and 6181 females.[9]

Table 1 reports descriptive information about ethnicity and exogamy/endogamy for the estimation sample by gender. The largest ethnic group is Javanese (64 %-65 %), followed by Sundanese (20 %).

Madurese and Betawi ethnicities are significantly less common (we grouped other minority ethnic groups due to the low number of observations, although all ethnic groups are used for the construction of the exogamy indicator). Overall, around 13 % of individuals in the sample are engaged in an interethnic marriage, with this proportion being significantly lower for individuals from the Javanese ethnicity, which is the largest ethnic group in Java. In Table 2, we also display the proportion of interethnic marriages according to educational level. The probability of having a partner from a different ethnic background increases with educational attainment. More concretely, among individuals with less than the compulsory education (junior high school), the exogamy rate is 7.8 % males and 9.2 % for females. However, this proportion increases up to around 20 % for individuals with university education.

Similar descriptive evidence about the association between educational attainment (share of individuals with at least senior secondary education) and exogamy (fraction of interethnic marriages) can be appreciated in Fig. 3, which is based on collapsed data at the district level obtained from the 2010 Census. Of course, these differences in the likelihood of exogamy associated with educational attainments cannot be interpreted in causal terms, since there could be differences in observed and unobserved characteristics that affect both education and the propensity for interethnic marriages.

Table 3 displays basic summary statistics for all the variables that we use in the empirical analysis for males and females. Besides exogamy and the three measures of educational attainment, we also report descriptive information about the number of available HEIs within a certain radius from the district of residence at age 18 (exposure at other ages in not reported for reasons of space). As expected, exposure increases with the radius. Moreover, exposure is higher for private than for public HEIs, which is in line with the figures reported in Section 2. To provide suggestive information about the changes across the cohort in exposure to HEI, driven by the expansion process, in Fig. 4 we display a scatter plot and a lowess fit of the average number of HEI surrounding the district of residence at age 18 by year. For both genders we observe a pronounced positive trend, indicating that exposure to HEI increases across the cohorts.

As control variables, we use own ethnicity and religion and family background. Specifically, we consider the number of siblings, a dummy for having low-educated parents, and an indicator for mixed parental ethnicity (i.e. father's ethnicity different than mother's ethnicity).

We also employ additional variables that are used for the analysis of potential mechanisms. Using information about residential history, we construct an indicator for having changed district of residence between the year in which the individual turned 18 and 2014. Moreover, combining this information with the district of residence in 2014, we constructed a dummy that is equal to 1 if the individual resided in a large city: Jakarta, Bandung, Semarang, Surabaya, Surakarta, and Yogyakarta, the largest urban areas in the island of Java.

The indicator for being a member of a minority group in the place of residence is directly obtained from the IFLS data, combining the information about the largest ethnic group in the community of residence[10] in 2014 and own ethnicity. We also imputed ethnic fractionalization in the district of residence in 2010. In order to do this, we computed the fractionalization index at the district level using information on individual ethnicity from the 2010 Census (10 % sample), following Bazzi et al. (2019). Finally, we constructed a proxy for social norms based on the question regarding trust in individuals from the same ethnic group relative to individuals from different ethnic backgrounds. Specifically, the question asks whether the individual: 1) strongly agrees, 2) agrees, 3) disagrees, or 4) strongly disagrees with the statement that they place more trust in individuals from the same ethnic group than they do in others. Therefore, we use an indicator that takes the value of one if the individual agrees or strongly agrees with the above statement. Unfortunately, this variable is missing for 25 % of the estimation sample.

4. Empirical strategy

Our objective is to estimate the (causal) impact of education on the likelihood of being in a relationship with a partner from a different ethnic background (exogamy). The equation of interest takes the form

$$EXO_i = \alpha + \delta EDUC_i + \beta' X_i + \theta_{tp(i)} + \varepsilon_i. \quad (1)$$

here, EXO_i is an indicator that takes the value 1 for individuals who have a partner with a different ethnicity (exogamy), and 0 for those who have a partner belonging to the same ethnic group (endogamy). The variable $EDUC_i$ encompasses different proxies for educational attainment, namely i) years of schooling, ii) college attendance, and iii) college completion, which represent our main explanatory variables of interest. The model also includes a set of control variables (X_i), which comprise dummy variables for one's own ethnicity and religion, the number of siblings, an indicator for having low-educated parents, and another dummy for having ethnically-mixed parents. We also control for year of birth (t) × province of residence (p) fixed effects, which capture province-cohort specific trends in local time-varying factors that might affect the outcome.[11] Throughout the whole empirical analysis, we estimate the model separately for males and females.

We start with the OLS estimation of equation (1). However, the causal interpretation of the OLS estimate of the δ parameter is challenging, mostly because of the likely relevance of unobserved factors that correlate both with educational attainment and with the propensity to form an ethnically-mixed couple. Moreover, beyond the issue of unobservable factors, measurement error in educational attainment could also introduce additional bias. To deal with these issues and obtain a plausibly causal estimate of the effect of education on exogamy, we employ an Instrumental Variable (IV) approach that leverages the presence of Higher Education Institutions (HEI) in the place of residence during adolescence, exploiting the massive geographical expansion of HEIs that took place in Java over time. More specifically, our instrument ($HEI^r_{d(i)\tau(i)}$) consists in the number of existing HEIs (at the relevant age, τ) in a certain radius (r) from the centroid of the district of residence (d).[12] In our preferred specification, we define the instrument based on the

[9] The estimation sample contains a slightly higher number of males than females, since there are cases in which the wife is younger than 25 and, therefore, does not satisfy the 25–65 age range criteria.

[10] This information proceeds from "Community-Facility Survey" of IFLS and is reported by the official village/township leader.

[11] We primarily focus on the province of residence at age 18 due to reasons related to our identification strategy. Notice that accounting for province-cohort-specific trends enables us to implicitly control for variations in the size of ethnic groups at the local level and their evolution over time, which, among other factors, represents an important determinant of interethnic marriages (as mentioned in Section 2.1).

[12] We adopt clustered standard errors at the district level, which represents the primary level of variation for the instrument. We experimented with two-way clusters at the district-year of birth level, yielding similar results (available upon request). Note also that, due to the low number of cases with more than 10 HEIs surrounding the district of residence at age 18, we capped the variable at 10, as this yielded higher F-statistics in the first-stage regressions. However, the results from the structural equation, in terms of the estimated coefficient of interest, remain virtually identical to those obtained using the original exposure variable (available upon request).

Table 1
Endogamy and Exogamy by Ethnicity.

	Males			Females		
Variable	% sample	Endogamy	Exogamy	% sample	Endogamy	Exogamy
Javanese	0.640	0.931	0.069	0.648	0.928	0.072
Sundanese	0.209	0.833	0.167	0.213	0.817	0.183
Madurese	0.049	0.877	0.123	0.049	0.904	0.096
Betawi	0.067	0.611	0.389	0.062	0.652	0.348
Other Ethnicities	0.035	0.413	0.587	0.028	0.489	0.511
Total	1	0.868	0.132	1	0.874	0.126
Observations	6352	5548	843	6181	5403	778

Table 2
Endogamy and Exogamy by Level of Education.

	Males			Females		
Variable	% sample	Endogamy	Exogamy	% sample	Endogamy	Exogamy
Less than Compulsory Education	0.399	0.922	0.078	0.459	0.908	0.092
Post Compulsory Education	0.601	0.832	0.168	0.541	0.845	0.155
No University Attendance	0.863	0.880	0.120	0.870	0.884	0.116
University Attendance	0.137	0.791	0.209	0.130	0.805	0.195
No University Completion	0.884	0.878	0.122	0.880	0.884	0.116
University Completion	0.116	0.792	0.208	0.120	0.802	0.198
Total		0.868	0.132		0.874	0.126
Observations	6352	5548	843	6181	5403	778

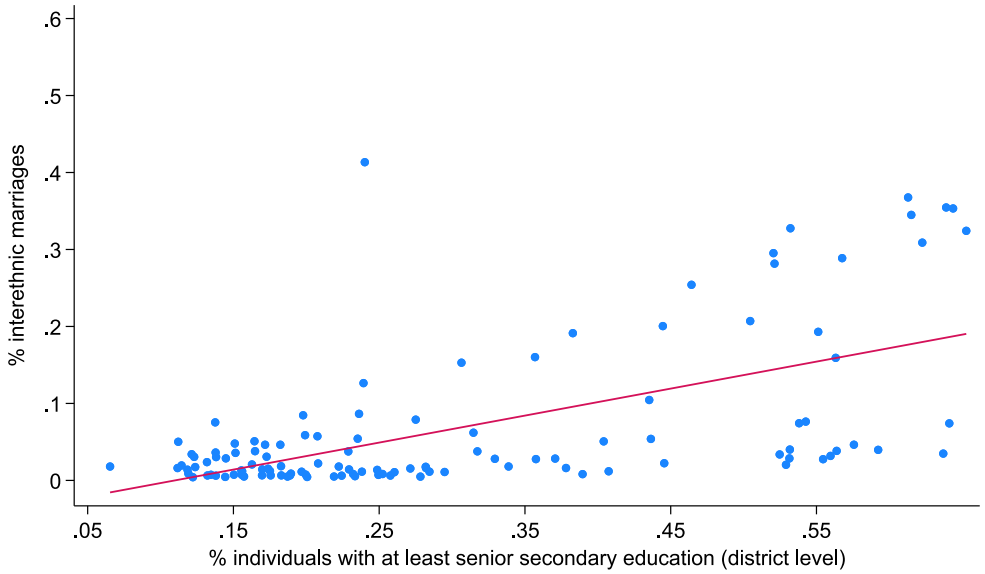

Fig. 3. Share of individuals with at least senior secondary education and interethnic marriages at the district level (2010 Census) *Note: the proportions are computed by collapsing the data at the district level, using observations of married individuals (couples) with age comprised between 25 and 65.*

district of residence at age 18, which is the typical age in Indonesia for entering an university. Similarly, we consider a radius of 10 km to compute the number of available HEIs surrounding the district of residence. For both dimensions of the instrument, we select the option that maximizes the instrument's strength. However, we also conduct robustness tests using alternative reference ages and different radii. The corresponding first-stage equation is

$$EDUC_i = \mu + \rho HEI^r_{d(i)\tau(i)} + \gamma' X_i + \omega_{tp(i)} + u_i \quad (2)$$

Therefore, we use within birth cohort and province variation in the geographical exposure to HIE as an exogenous source of variation in educational attainment. This approach is valid under the assumptions that the presence of HEIs at the local level is a strong predictor of educational attainment while not being directly related to ethnic exogamy. The IV counterpart of equation (1) is thus

$$EXO_i = \alpha + \delta_{IV}\widehat{EDUC}_i + \beta' X_i + \theta_{tp(i)} + \varepsilon_i. \quad (3)$$

Under the validity of the underlying assumptions, the coefficient associated with educational attainment (δ_{IV}) can be interpreted as the causal effect of education on interethnic marriages among individuals induced by higher educational attainments due to the geographical expansion of higher education (in a LATE framework).

This IV approach resembles the one employed in Currie and Moretti (2003), and its variants that have been adopted by others in more recent papers (Kyui, 2016; Kamhöfer et al. 2019; Belskaya et al. 2020; Westphal et al. 2022, among others). Nevertheless, there are certain notable differences in our setting that warrant further discussion. On the one hand, an advantage of our dataset is that it provides retrospective information about the district of residence since birth, year by year. Hence, we are able construct our instrument based on the district of

Table 3
Descriptive Statistics.

Variable	Males		Females	
	Mean	*S.D.*	Mean	*S.D.*
Exogamy	0.132	*0.338*	0.126	*0.332*
Years of Schooling	8.687	*4.178*	8.176	*4.291*
University Attendance	0.137	*0.344*	0.130	*0.337*
University Completion	0.116	*0.320*	0.120	*0.324*
HEI within 5 Km radius at age 18	1.264	*2.495*	1.290	*2.597*
HEI within 10 Km radius at age 18	3.146	*5.851*	3.158	*6.009*
HEI within 15 Km radius at age 18	5.292	*9.195*	5.331	*9.372*
HEI within 20 Km radius at age 18	6.762	*11.108*	6.855	*11.379*
HEI within 25 Km radius at age 18	8.197	*12.568*	8.361	*12.929*
Public HEI within 10 Km radius at age 18	0.685	*1.504*	0.690	*1.536*
Private HEI within 10 Km radius at age 18	2.479	*4.806*	2.493	*4.941*
Javanese	0.640	*0.480*	0.648	*0.477*
Sundanese	0.209	*0.406*	0.213	*0.409*
Madurese	0.049	*0.217*	0.049	*0.216*
Betawi	0.067	*0.250*	0.062	*0.241*
Other Ethnicities	0.035	*0.184*	0.028	*0.165*
Moslems	0.971	*0.168*	0.969	*0.174*
Christians	0.027	*0.161*	0.029	*0.169*
Hindus	0.001	*0.028*	0.001	*0.031*
Other Religions	0.002	*0.040*	0.001	*0.031*
Number of Siblings	3.069	*2.221*	3.363	*2.500*
Low Parental Education	0.122	*0.328*	0.136	*0.342*
Ethnically-Mixed Parents	0.070	*0.255*	0.063	*0.243*
Change District of Residence (18–2014)	0.138	*0.345*	0.124	*0.330*
Move to Large Cities (18–2018)	0.035	*0.183*	0.039	*0.193*
Fractionalization	0.427	*0.495*	0.423	*0.494*
Being a Minority in 2014	0.216	*0.411*	0.207	*0.405*
Trust Own Ethnicity	0.636	*0.481*	0.687	*0.464*
Observations	6391		6181	

Note: Low parental education = 1 if parents did not complete primary education. Fractionalization has been defined according to ethnicity, based on district-level information from the 10 % of the 2010 Census. Being a minority in 2014 = 1 if the individual's ethnicity is different than the most prevalent ethnicity in the district of residence in 2014. Trust own ethnicity = 1 if the individual declares he/she completely agrees or agrees with the sentence "I trust individuals from my own ethnic group more than others". This last variable is available only for 4515 males and 4872 females (i.e. is missing for 25 % of the estimation sample). The corresponding descriptive statistics have been obtained only with valid observations.

residence at age 18, a pivotal year when individuals in Indonesia typically enrol in an university (although we also explore previous ages for robustness, as elaborated below). Indeed, data about the place of residence during adolescence is not always available and several works rely on information about residence at birth. On the other hand, unfortunately, to the best of our knowledge, information about the size of the cohort of individuals in the age range to attend college is not available for the case of Indonesia, neither at the district nor at the province level. This constitutes a data limitation for our identification strategy. In fact, as noticed by Currie and Moretti (2003), the geographical variation in the number of HEIs across cohorts could be capturing both the demand and supply for university education. While the supply-side can be reasonably taken as exogenous, demand-side factors can (directly) correlate with other local-level variables that could be associated with the decision to form an ethnically-mixed couple. Although controlling for cohort × province of residence specific fixed effects should account for local-level confounders varying across birth cohorts, questions may still arise regarding the exogeneity of the instrument. That is, there could be unobserved local factors correlated with both the demand for higher education and the propensity for exogamy, influencing individuals born in a given cohort in different ways within their province of residence. An additional but related potential concern that might undermine the validity of the instrument is the endogenous residential sorting of families and/or individuals during the educational process. This is because decisions about residential location could be influenced by the desire to live in areas with better access to higher education, with the aim of facilitating university attendance for their children, which could in turn affect the formation of interethnic marriages. Nevertheless, we conduct several robustness checks that are aimed at providing evidence in favour of the validity of the instrument and the causal interpretation of the corresponding estimate of the parameter of interest (δ_{IV}).

4.1. *Alternative specifications and robustness checks*

To validate our IV approach and the general empirical framework, we perform a battery of sensitivity tests. First, we test for the robustness of the results with respect to the two main dimensions along which we construct the instrument: the radius (r) and age at exposure (τ). Regarding the former element, we compute the number of HEIs

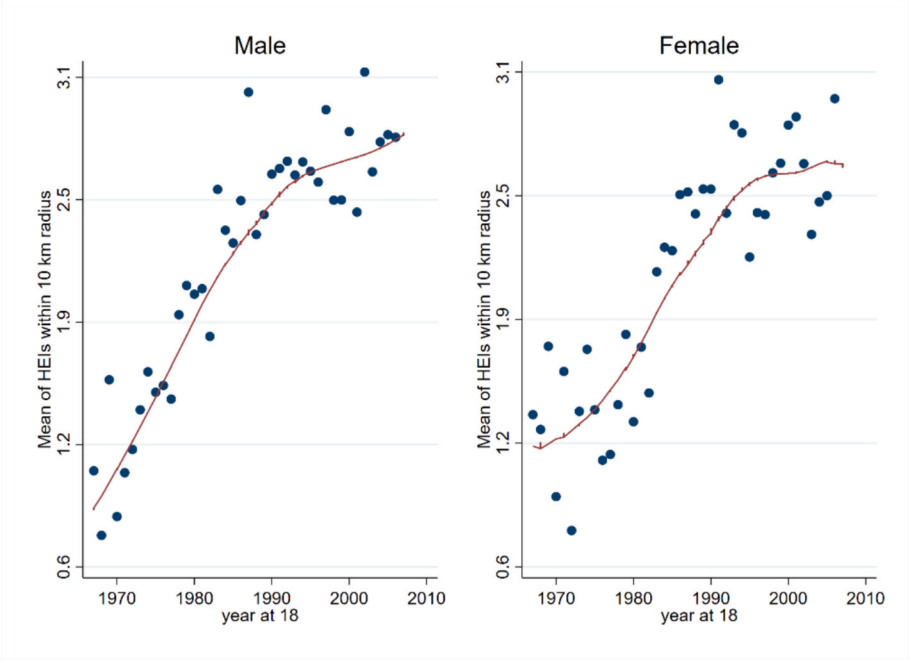

Fig. 4. Average HEIs within a 10 km radius by year at age 18.

surrounding the individual's district of residence using buffers of a certain radius from the district's centroid. We adopt this strategy to define the availability of HEI because the IFLS data contain information on two main geographical identifiers: the province, which is possibly too broad to define the relevant area of influence, and the district, which is likely to be too narrow.[13] Of course, the choice of the radius is, by definition, subject to some degree of arbitrariness. We therefore computed the instrument based on different radii of exposure: 5 km, 10 km, 15 km, 20 km and 25 km. Moreover, we also adopt a similar approach to that in Kamhöfer et al (2019) and Westphal et al. (2022), which consists in considering data on the location of all university campuses in the island of Java and computing the number of available HEIs weighted by their distance from the centroid of the district of residence using Gaussian Kernel weights (using Silverman's rule for bandwidth selection). To determine the best specification, we select the option that maximizes the strength of the instrument, i.e. maximizes the first stage *F*-statistic. Second, we also check for the sensitivity to the choice of the relevant age at exposure. Although the natural choice consists in selecting the typical age at which people enrol in college (18 in the case of Indonesia), as done in other papers, to some extent this is also an arbitrary choice. Moreover, using age 18 could also be related to the issue of endogenous residential sorting, because individuals and families might decide to relocate to areas in which there is more accessibility to higher education, but there is also a more favourable environment for the formation of ethnically-mixed couples. Therefore, we defined the instrument based on the district of residence at ages 18, 15, 12, 6, and at birth, and check for the stability of the coefficient of interest. Subsequently, we selected the option that yields a higher *F*-statistic in the first stage.[14] Indeed, assessing the robustness of the results by varying the age at exposure is also important for addressing concerns about endogenous residential sorting during the educational process, which could lead to a violation of the exclusion restriction assumption. To further demonstrate that endogenous residential sorting does not affect our identification strategy, we also re-estimate the model, restricting the sample to individuals who did not change their district of residence between their year of birth and the year they turned 18. Finally, we use an alternative instrument based on the minimum distance (in kilometres) from the district's centroid to the nearest HEI at age 18 and assess the robustness of the results.

After determining the preferred specification of the instrumental variable ($HEI^{r}_{d(i)\tau(i)}$), we implement other checks that are aimed at validating its exogeneity and the exclusion restriction assumption.[15] We are especially concerned about the possibility that the number of available HEIs could be capturing time-varying demand-side local factors that directly affect the outcome, which would imply a violation of the exclusion restriction. For these checks we also focus on the following reduced-form equation:

$$EXO_i = \alpha + \lambda_{RF} HEI^{r}_{d(i)\tau(i)} + \beta' X_i + \theta_{tp(i)} + \varepsilon_i. \quad (4)$$

First, we compare the estimate of the reduced-form coefficient (λ_{RF}) from equation (4) and from the structural equation (3) estimated by IV/2SLS using the number of HEIs in a radius of 10 km from the district of residence at 18 as instrument, to the same estimates obtained from an alternative specification in which we also include the presence of HEIs surrounding the district of residence at birth ($\tau = 0$) as additional control. This additional control variable should capture long-standing unobservables at the local level that could directly correlate with both the demand for higher education and interethnic marriages. If these factors are actually relevant, the reduced-form coefficient of our instrument should be significantly lower, which would suggest that the exclusion restriction assumption is not satisfied. Conversely, finding results similar to our baseline estimates, along with null coefficients for the additional control variable in both the reduced-form and second-stage regressions, would provide supporting evidence for the validity of the exclusion restriction underlying our IV approach. In a similar vein, we re-estimate the model while conditioning to the presence of at least one HEI at birth around the district of residence. This implies considering only individuals who were born in districts that should be generally similar in terms of local characteristics. Second, we aim to account for potential recent changes in local demand-related factors by including an additional control for the presence of "new" HEIs established between the individual's birth year and the year they turned 18, located in proximity to the district. If what really matters in both the reduced-form and first-stage equations is the number of newly established HEIs and not the overall stock, this is probably indicative of the higher relevance of (potentially endogenous) demand-side factors rather than supply-side elements. Third, borrowing from Currie and Moretti (2003), we include as an additional control variable the number of available HEIs at age 25. In the hypothetical case in which our instrument is capturing a spurious correlation with local-level unobservables that are directly related to the propensity of having a partner from a different ethnicity, we would observe a significant coefficient for the number of HEIs at age 25, a clear reduction in the coefficient of the instrument relative to the baseline reduced-form estimation, and a different (possibly insignificant) estimate of the δ_{IV} coefficient, suggesting a possible violation of the exclusion restriction assumption. Fourth, also following Currie and Moretti (2003), we compute the exposure to public and private HEIs separately and re-estimate the model with each of these two instruments. As the establishment of private universities is more likely to be related to (potentially endogenous) geographical characteristics such as the price of land, but also to the expected demand. Therefore, finding larger effects of the presence of private HEIs than public HEIs would be indicative of the lack of exogeneity of the instrument.[16] Finally, we conduct a falsification exercise based on a permutation test, in which we randomly assign the district of residence at age 18, and impute the number of HEIs within 10 km of the fake district of residence. This process is repeated 10,000 times, and we estimate the reduced-form equation for each replication, generating a distribution of placebo reduced-form coefficients. If the distribution of these placebo estimates is symmetric and centred around 0, it would be evidence that the real instrument could be capturing some kind of spurious correlation, suggesting a possible violation of the exclusion restriction assumption. Moreover, we also estimate an overidentified model in which we use dummies for the presence of HEIs at the local level and present the

[13] Authors of existing papers have focused on the number of colleges within administrative geographical units that are in between provinces and districts, such as US counties (Currie and Moretti, 2003) and municipalities (Kyui et al, 2016).

[14] In conducting this exercise, we also change the year of birth × province of residence accordingly, considering the province of residence at the corresponding age.

[15] Following the suggestion of an anonymous referee, we also examined the validity of the monotonicity assumption. To this end, we transformed the instrument into a binary variable equal to 1 if there is at least one HEI within 10 km of the centroid of the district of residence at age 18. We then estimated the cumulative density function (CDF) for years of education separately for individuals without a nearby university and those with at least one university nearby, following Fiorini and Stevens (2021). The CDF for the latter group is clearly shifted to the right and does not intersect with the CDF of the former group, supporting the validity of the monotonicity assumption (results available upon request).

[16] Currie and Moretti (2003) also refer to a potential issue related to the differences in tuition fees between public and private institutions. This concern is less relevant for the case of Indonesia. As explained in the institutional background section, differences in prices between public and private Higher Education Institutions (HEI) are not very pronounced, though they are indeed field- and university-specific. The primary difference in cost lies in the entry fee for private institutions.

results of the Hansen J-test for overidentification.

Besides this battery of sensitivity checks regarding the definition and the validity of our instrumental variable, we also perform two additional checks to provide further evidence about the internal validity of our estimations. On the one hand, the causal chain that we hypothesized is that the expansion of HEIs fostered educational attainment, and this in turns increased the propensity to find a partner from a different ethnicity. However, although rare, there could be cases where marriage occurs before completing education. To address this, we re-estimate the model after excluding individuals who married before leaving the education system. On the other hand, we observe the ethnicity of both members of the couple in 2014, which is after marriage. Many existing papers on ethnicity assume this to be a predetermined and immutable feature. However, Rademakers and van Hoorn (2021) provide evidence of the likelihood of changing ethnicity in Indonesia, noting that this pattern is more prevalent among members of interethnic marriages. In IFLS ethnicity is reported from the last two waves (2014 and 2007). Therefore, we also repeat our estimations considering only individuals who i) are interviewed in both waves and ii) report the same ethnicity in 2014 as in 2007.

4.2. Analysis of heterogeneous effects and potential mechanisms

The additional evidence that we report in this paper concerns the analysis of heterogeneous effects of education on exogamy, as well as potential mechanisms that lie behind the link between HEI, educational attainment, and interethnic marriage.

As for heterogeneous effects, we consider whether the impact of education differs along three main features: own ethnicity, parental education and having parents from a mixed ethnic background. In doing that, we use interactions rather than splitting the sample, with the aim of avoiding small sample issues. Therefore, for each of these three variables, we estimate separately the model that includes interactions with the instrument as an additional exclusion restriction, as well as interaction with educational attainment as an additional endogenous regressor.

In terms of potential mechanisms, while there are several factors that could be relevant in this setting, we are limited by data availability.[17] Consequently, we focus on two main elements: migration/residential location and social norms. Regarding the former, the hypothesis is that the expansion of higher education leads individuals to attain higher educational levels, influencing their propensity to migrate and, possibly, settle in larger and more ethnically diverse cities. This, in turn, could increase their likelihood of marrying someone from a different ethnic background. Therefore, we consider alternative outcomes related to these factors: i) an indicator for having changed place of residence between age 18 and 2014, ii) an indicator for currently residing in a large city, iii) being a minority in the place of residence in 2014 and iv) ethnic fractionalization in the district of residence. As for social norms, the idea is to employ a proxy for the tolerance and openness toward different ethnic groups. This, in turns, could be fostered by increased educational attainment and consequently affecting the propensity to match with a partner from a different ethnicity. Based on available data, we rely on the variable capturing whether the individual places more trust in others from the same ethnicity or not, which has been described in the data section. Therefore, we use the indicator for trusting more individuals from the same ethnicity than others as a proxy for social norms.

In order to justify the adoption of our IV approach while using these alternative variables as outcomes, we focus on the reduced-form equation in which they are directly regressed against the presence of HEIs at the local level. However, it is important to note that all the variables that we consider in the analysis are observed possibly several years after marriage (i.e. in 2010 for fractionalization and in 2014, the survey year, for other variables). Therefore, the results should be taken with caution because these variables could actually reflect "consequences" of interethnic marriages rather than pure mechanisms (i.e. an individual who is married with someone from a different ethnicity could develop more trust toward others from a different ethnic background). While acknowledging this limitation, we remain convinced that analysing the impact of exposure to HEI on these proxies for migration/residential choices and social norms provides suggestive evidence about the relevance of these factors in the underlying causal effect between educational attainment and the formation of interethnic marriages.

5. Results

Table 4 displays the main results from the OLS estimation of equation (1), for each of the three measures of educational attainment (complete results are reported in Table A1 of the Appendix). The estimates are separately obtained for males and females. We estimate the model without control variables (i.e. only including fixed effects for year of birth × province of residence at 18), as well as controlling for own ethnicity, own religion, number of siblings, parental education, and mixed parental ethnicity. In general, education is positively and significantly associated with the probability of being in an ethnically-mixed marriage. Each additional year of increase in years of schooling is associated with an increase in the probability of exogamy of 0.7 and 0.6 percentage points (p.p.) for males and females, respectively. Having attended or completed university is associated with a higher propensity of interethnic marriages as well (around 6–7p.p.).

The inclusion of control variables leads a certain reduction in the coefficients of all measures of educational attainments, more pronounced for females, although their significance remains unchanged. The results regarding control variables are of independent interest and warrant further discussion. As for own ethnicity, people from the

Table 4
OLS Estimations – Dependent Variable: Exogamy.

	Males		Females	
	(1)	(2)	(3)	(4)
Panel A:				
Years of Schooling	0.007***	0.006***	0.006***	0.004***
	(0.001)	(0.001)	(0.002)	(0.002)
R-squared	0.200	0.260	0.204	0.238
Panel B:				
University Attendance	0.075***	0.058***	0.064***	0.047***
	(0.014)	(0.012)	(0.018)	(0.017)
R-squared	0.199	0.259	0.204	0.238
Panel C:				
University Completion	0.072***	0.057***	0.067***	0.049***
	(0.015)	(0.013)	(0.019)	(0.018)
R-squared	0.198	0.259	0.204	0.238
Controls	No	Yes	No	Yes
Observations	6391	6391	6181	6181

Notes: OLS estimations with exogamy as outcome variable (i.e. having a partner with a different ethnicity than the individual). Main regressors: years of schooling (Panel A), university attendance (Panel B), and university completion (Panel C). Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects. Additional control variables: ethnicity, religion, number of siblings, having parents with low education and having parents with different ethnicities.

[17] As noted by an anonymous referee, improved proficiency in Bahasa Indonesia (the national standard language) resulting from higher educational attainment may also serve as a potential mechanism. The adoption of a common language—used in the education system—can reduce linguistic barriers, thereby facilitating the formation of ethnically mixed couples. However, we are unable to directly assess the role of language skills as a mechanism, as the IFLS questionnaire only includes a general question about the language most frequently used in the household. Exploring the impact of education on language proficiency and usage could be a valuable avenue for future research.

Sundanese ethnicity are not more likely to engage in mixed marriages than those from the Javanese ethnicity (the largest ethnic group in Java). However, those from other ethnicities are generally more likely to be married to a partner from other ethnic groups, except for Betawi females. Religion does not seem to play an important role while other variables are controlled for. Specifically, only Hindu males exhibit a lower likelihood of engaging in interethnic marriages compared to their Javanese counterparts. Having low-educated parents is associated with a slightly lower probability of exogamy. Moreover, as expected, parental exogamy is an important predictor of own exogamy, indicating a certain intergenerational pattern in interethnic marriages.

Overall, the results suggest that higher educational attainment is associated with a greater propensity to marry outside one's ethnic group (exogamy) and a lower likelihood of marrying within the same ethnicity (endogamy). However, due to the potential endogeneity of educational attainment in the exogamy equation, the previous results cannot be interpreted in causal terms. Therefore, to obtain plausibly causal estimates, we employ our measure of geographical exposure to HEIs as an instrument for educational attainment. We start with exposure to HEI defined according to the district of residence at 18, considering the number of available institutions within 10 km of the district's centroid. Table 5 reports the results (with and without controls) for the three educational outcomes. The first-stage coefficients are generally positive and highly significant, highlighting the strength of the exposure to HEI as predictor of years of schooling and university attendance/completion. The IV/TSLS estimates of equation (3) confirm that education exerts a positive effect on the probability of exogamy. Specifically, these coefficients represent the (causal) impact of education on the likelihood of interethnic marriages among those who are induced into higher educational attainments due to the presence of HEIs surrounding their place of residence at 18 (i.e. the compliers). The results from the model without control indicate that each additional year of schooling increases the propensity for exogamy by 3.5p.p. among males and 4.6p.p. for females. University education increases the probability of having a partner from a different ethnic background by around 44–52p.p. for males and 71–73p.p. for females. The model with control variables provides similar evidence, generally with slightly lower second-stage coefficients. Finding similar results from the model that includes controls is a first indication in favour of the internal validity of the results.

Generally, the coefficients are higher than those obtained from OLS, which is consistent with a LATE interpretation of the results. Indeed, the compliers in our setting are likely individuals in the upper-middle part of the distribution of the propensity to attain high levels of education. Specifically, these are individuals who would not pursue further education unless they live relatively close to an higher education institution but decide to continue studying due to university expansion (whereas those at the top of the distribution would be always-takers). Indeed, these individuals may be particularly responsive to higher educational attainment in terms of their likelihood of forming partnerships with individuals from different ethnic groups, which could help explain why the IV/TSLS estimates are higher than the OLS estimates. Moreover, an alternative (or complementary) explanation for the higher estimates obtained using IV/TSLS is the presence of measurement error in educational attainment, which likely biases the OLS estimates downward.

Table 5
IV/2SLS Estimations – Dependent Variable: Exogamy.

	Males		Females	
	(1)	(2)	(3)	(4)
Panel A: Years of Schooling				
First Stage	0.235***	0.196***	0.241***	0.186***
	(0.037)	(0.031)	(0.041)	(0.033)
Second Stage	0.035***	0.032***	0.046***	0.047***
	(0.011)	(0.011)	(0.015)	(0.017)
First-Stage F-statistic	39.635	39.778	34.813	31.339
P-Value	0.000	0.000	0.000	0.002
Panel B: University Attendance				
First Stage	0.019***	0.017***	0.015***	0.013***
	(0.003)	(0.003)	(0.003)	(0.003)
Second Stage	0.441***	0.372***	0.727***	0.696***
	(0.147)	(0.139)	(0.223)	(0.247)
First-Stage F-statistic	36.631	30.203	23.906	19.238
P-Value	0.000	0.000	0.000	0.009
Panel C: University Completion				
First Stage	0.016***	0.014***	0.015***	0.013***
	(0.002)	(0.002)	(0.003)	(0.003)
Second Stage	0.516***	0.435***	0.709***	0.668***
	(0.164)	(0.155)	(0.212)	(0.234)
First-Stage F-statistic	40.821	33.469	19.873	16.673
P-Value	0.000	0.000	0.000	0.000
Controls	No	Yes	No	Yes
Observations	6391	6391	6181	6181

Notes: 2SLS estimation with exogamy as outcome variable (i.e. having a partner with a different ethnicity than the individual). Endogenous regressors: years of schooling (Panel A), university attendance (Panel B), and university completion (Panel C). Instrumental variable: number of Higher Education Institutions (HEI) within a 10 km radius from the centroid of the district of residence at age 18. Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects. Additional control variables: ethnicity, religion, number of siblings, having parents with low education and having parents with different ethnicities.

5.1. Results from alternative specifications and robustness checks

To validate our findings, we report the evidence from several sensitivity checks. For simplicity, we report these results for years of schooling only.[18] First, we show the results obtained by adopting different definitions of the radius of exposure for calculating the number of available HEI, which are displayed in Tables 6 (males) and Table 7 (females). In general, the results are virtually identical across all alternatives, including when employing the Kernel Density Weighting based on the distance from the district's centroid and the location of HEIs. However, using a radius of 10 km yields the highest *F*-statistic for the first stage, and thus is our preferred option.

Second, we consider different relevant ages at exposure (Tables 8 and 9). As can be seen, the results are not affected by the choice of age at exposure. The first-stage coefficients remain positive and significant for both males and females, even when defining the number of available HEIs within a 10 km radius based on the district of residence at birth—though slightly reduced. Indeed, this is also an indication that the instrument is not blurred by endogenous residential sorting. Using exposure at 18 years old provides the largest *F*-statistic for males, although employing age 12 as reference to compute exposure seems to be the best option for females. Nevertheless, given the overall stability of the result, we retain 18 as reference age as baseline for both genders.[19] To further discard the possibility that the results are affected by endogenous residential sorting during the educational progress, we also replicate the estimations after retaining only individuals who never changed district of residence from their birth year until they turned 18 (see Table A3 of the Appendix). Again, the results are virtually the same as for the original estimation sample.

Subsequently, we present the sensitivity checks that are aimed at dispelling doubts about the possibility that the number of available HEIs is capturing (potentially endogenous) demand-side factors. The results

[18] The results of robustness checks for other educational attainments are available upon request.

[19] Similar results are obtained when using the distance to the nearest HEI at age 18 as an alternative instrument (see Table A2 in the Appendix). However, this approach yields a lower first-stage *F*-statistic, which is why we retain the number of HEIs surrounding the district of residence as our main instrument.

Table 6
Robustness check – Using different Radii – Males.

	(1)	(2)	(3)	(4)	(5)	(6)
Radii of exposure:	5 km	10 km	15 km	20 km	25 km	Kernel
Panel A: First Stage – Dependent Variable: Years of Schooling						
HEI within X radius at age 18	0.261***	0.235***	0.233***	0.212***	0.195***	0.408***
	(0.066)	(0.037)	(0.038)	(0.038)	(0.042)	(0.080)
First-Stage F-statistic	15.845	39.635	38.384	31.281	21.605	26.228
Panel B: Second Stage – Dependent Variable: Exogamy						
Years of Schooling	0.037***	0.035***	0.036***	0.035***	0.032***	0.038***
	(0.012)	(0.011)	(0.011)	(0.012)	(0.012)	(0.013)
Observations	6391	6391	6391	6391	6391	6391

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

Table 7
Robustness check – Using different Radii – Females.

	(1)	(2)	(3)	(4)	(5)	(6)
Radii of exposure:	5 km	10 km	15 km	20 km	25 km	Kernel
Panel A: First Stage – Dependent Variable: Years of Schooling						
HEI within X radius at age 18	0.295***	0.241***	0.228***	0.220***	0.195***	0.359***
	(0.064)	(0.041)	(0.044)	(0.041)	(0.043)	(0.085)
First-Stage F-statistic	21.422	34.813	27.447	28.505	20.102	17.794
Panel B: Second Stage – Dependent Variable: Exogamy						
Years of Schooling	0.051***	0.046***	0.047***	0.042***	0.033**	0.043**
	(0.016)	(0.015)	(0.013)	(0.013)	(0.013)	(0.017)
Observations	6181	6181	6181	6181	6181	6181

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

Table 8
Robustness check – Using different age at exposure – Males.

	(1)	(2)	(3)	(4)	(5)
Age at exposure:	18	15	12	6	0
Panel A: First Stage – Dependent Variable: Years of Schooling					
HEI within 10 km radius	0.235***	0.216***	0.215***	0.204***	0.187***
	(0.037)	(0.036)	(0.036)	(0.039)	(0.041)
First-Stage F-statistic	39.635	36.148	34.863	27.898	20.540
Panel B: Second Stage – Dependent Variable: Exogamy					
Years of Schooling	0.035***	0.038***	0.037***	0.039***	0.038***
	(0.011)	(0.011)	(0.011)	(0.012)	(0.013)
Observations	6391	6391	6391	6391	6391

Notes: Standard errors (in parentheses) are clustered by district of residence (at the corresponding age). *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for fixed effects by year of birth dummies × province of residence (at the corresponding age).

Table 9
Robustness check – Using different ages at exposure – Females.

	(1)	(2)	(3)	(4)	(5)
Age at exposure:	18	15	12	6	0
Panel A: First Stage – Dependent Variable: Years of Schooling					
HEI within 10 km radius	0.241***	0.230***	0.218***	0.210***	0.193***
	(0.041)	(0.038)	(0.035)	(0.036)	(0.037)
First-Stage F-statistic	34.813	36.518	39.075	34.681	27.709
Panel B: Second Stage – Dependent Variable: Exogamy					
Years of Schooling	0.046***	0.048***	0.048***	0.051***	0.054***
	(0.015)	(0.015)	(0.015)	(0.017)	(0.019)
Observations	6181	6181	6181	6181	6181

Notes: Standard errors (in parentheses) are clustered by district of residence (at the corresponding age). *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for fixed effects by year of birth dummies × province of residence (at the corresponding age).

are reported in Tables 10 and 11. Here we mainly focus on reduced-form equations, although we also display the results for the first-stage and the second-stage for comparison. Column (1) contains the results from the reduced-form equation (4) obtained from the baseline specification of the instrument. As expected, geographical exposure to HEI at age 18 exerts a positive and significant effect on the likelihood of exogamy for both genders, which is in line with the previous IV/TSLS results. In column (2) we repeat the estimations after controlling for the number of HEIs at birth, which would capture potential long-standing trends in the demand for higher education at the local level. Indeed, this additional control has a very small and insignificant point estimate in the reduced form equation, as well as in the second-stage equation. Moreover, the main results remain unchanged. We also obtain qualitatively similar evidence when restricting the sample to individuals born in districts with at least one HEI nearby. This restriction implies comparing districts that were generally similar in terms of pre-existing factors related to the demand for higher education. In column (4) we seek to control for potential recent changes in the demand for higher education across birth cohorts, by controlling for the number of newly established HEIs (i.e. those created since the individual's birth year and the year in which he/she turned 18). Also in this case, the corresponding coefficient is virtually zero and insignificant in the reduced form equation and in the second stage, while the coefficient of years of schooling remains practically unchanged. Finally, as in Currie and Moretti (2003), we control for the number of HEIs surrounding the district of residence at age 25 (column (5)). This additional control variable has a null coefficient both in the reduced form model and in the structural equation, and the coefficient of interest is virtually the same than in our baseline specification. Overall, the findings from these robustness check provide suggesting evidence in favour of the validity of the exclusion restriction assumption for our instrument.

We also obtain reassuring evidence regarding the validity of the instrument from the falsification based on the random assignment of the district of residence at 18 and the replication of 10,000 estimations of the reduced form equation using fake exposure to HEI (permutation test). As displayed in Fig. 1A of the Appendix, the distribution of fake

Table 10
Robustness check for demand-related factors – Males.

	(1)	(2)	(3)	(4)	(5)
Panel A: Reduced Form – Dependent Variable: Exogamy					
HEI within 10 km radius at age 18	0.008***	0.013**	0.008**	0.007*	0.009***
	(0.003)	(0.006)	(0.003)	(0.004)	(0.003)
HEI within 10 km radius at age 0		−0.007			
		(0.006)			
new HEI (0–18) in 10 km radius				0.001	
				(0.005)	
HEI within 10 km radius at age 25					−0.001
					(0.001)
Panel B: First Stage – Dependent Variable: Years of Schooling					
HEI within 10 km radius at age 18	0.235***	0.365***	0.152***	0.166***	0.296***
	(0.037)	(0.064)	(0.049)	(0.052)	(0.042)
HEI within 10 km radius at age 0		−0.184***			
		(0.064)			
new HEI (0–18) in 10 km radius				0.099**	
				(0.045)	
HEI within 10 km radius at age 25					−0.072***
					(0.023)
First-Stage F-statistic	39.635	32.246	9.833	10.203	49.430
Panel C: Second Stage – Dependent Variable: Exogamy					
Years of Schooling	0.035***	0.036**	0.053**	0.044*	0.031***
	(0.011)	(0.017)	(0.024)	(0.025)	(0.009)
HEI within 10 km radius at age 0		−0.000			
		(0.004)			
new HEI (0–18) in 10 km radius				−0.003	
				(0.007)	
HEI within 10 km radius at age 25					0.001
					(0.001)
Observations	6391	6391	2709	6391	6391

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects. Estimations in column (3) are obtained after retaining only individuals who were born in districts with at least one HEI within a radius of 10 km.

Table 11
Robustness check for demand-related factors – Females.

	(1)	(2)	(3)	(4)	(5)
Panel A: Reduced Form – Dependent Variable: Exogamy					
HEI within 10 km radius at age 18	0.011***	0.013**	0.013***	0.009**	0.012***
	(0.003)	(0.006)	(0.004)	(0.004)	(0.003)
HEI within 10 km radius at age 0		−0.004			
		(0.007)			
new HEI (0–18) in 10 km radius				0.003	
				(0.003)	
HEI within 10 km radius at age 25					−0.001
					(0.001)
Panel B: First Stage – Dependent Variable: Years of Schooling					
HEI within 10 km radius at age 18	0.241***	0.363***	0.152***	0.159***	0.276***
	(0.041)	(0.071)	(0.050)	(0.053)	(0.044)
HEI within 10 km radius at age 0		−0.171**			
		(0.073)			
new HEI (0–18) in 10 km radius				0.117***	
				(0.038)	
HEI within 10 km radius at age 25					−0.041**
					(0.020)
First-Stage F-statistic	34.813	26.425	9.323	8.841	39.975
Panel C: Second Stage – Dependent Variable: Exogamy					
Years of Schooling	0.046***	0.037**	0.086***	0.054*	0.043***
	(0.015)	(0.018)	(0.024)	(0.030)	(0.013)
HEI within 10 km radius at age 0		0.003			
		(0.005)			
new HEI (0–18) in 10 km radius				−0.003	
				(0.007)	
HEI within 10 km radius at age 25					0.001
					(0.001)
Observations	6181	6181	2675	6181	6181

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects. Estimations in column (3) are obtained after retaining only individuals who were born in districts with at least one HEI within a radius of 10 km.

reduced form coefficients is centred around zero and the real reduced-form coefficients are clearly outside its mass. Moreover, we report the results of the overidentified model that includes as instruments dummies for exposure to HEI (Table A4 of the Appendix). Although both the first-stage *F*-statistic and the second-stage coefficient of years of schooling are slightly lower than in the baseline, the results are qualitatively the same. Most importantly, the Hansen J-test for overidentification provides evidence in favour of the null hypothesis that the instruments can be excluded from the second stage, indicating that geographical exposure to HEI seems not to be directly related to exogamy.

Finally, again following Currie and Moretti (2003), we estimate the model considering two different instruments, which are based on exposure to public and private HEIs, respectively.[20] As shown in Table 12, the overall results are very similar when considering exposure to the two types of institutions. The first-stage coefficients are somewhat lower for private HEIs, while the second stage coefficients are slightly

[20] Following the suggestion of an anonymous referee, we also checked the results obtained including exposure to public and private HEIs as separate instruments. The estimated coefficients for years of schooling remained virtually identical to the baseline results, and the Hansen J test for overidentification yielded a very high *p*-value, indicating that the linear combination of both instruments is unlikely to be correlated with the structural equation's error term. However, this approach resulted in a weaker first stage, due to the correlation between the number of public and private HEIs surrounding the district of residence at age 18. These results are not reported but are available upon request.

Table 12
Separate exposure to public and private HEI.

	Males		Females	
	Public HEI	Private HEI	Public HEI	Private HEI
Panel A: First Stage — Dependent Variable: Years of Schooling				
HEI within 10 km radius at age 18	0.299***	0.250***	0.285***	0.240***
	(0.066)	(0.040)	(0.073)	(0.046)
First-Stage F-statistic	20.730	38.482	15.339	27.744
Panel B: Second Stage — Dependent Variable: Exogamy				
Years of Schooling	0.028***	0.040***	0.043**	0.052***
	(0.011)	(0.012)	(0.019)	(0.017)
Observations	6391	6391	6181	6181

Notes: Standard errors (in parentheses) are clustered by district of residence at age 18. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

higher. However, the stability of the results is again reassuring and suggest that the presence of HEIs is not capturing anticipated changes in the demand for higher education, or other local-level unobserved factors that could be directly related to the propensity to form interethnic marriages. Overall, these results suggest that our instrument is not capturing spurious effects that are due to changing trends in the local demand for higher education, supporting the underlying assumption of its exogeneity. As final robustness checks, we also replicate the estimations after excluding individuals who got married before completing education (Table A5 of the Appendix), as well as while retaining in the estimation sample only individuals who report the same ethnicity in 2014 (IFLS 5) as in 2007 (IFLS 4) and appear in both waves of the survey (Table A6 of the Appendix). For both robustness checks, the results are virtually identical to the baseline.

5.2. Evidence about heterogeneous effects and potential mechanisms

The evidence obtained so far indicates that higher educational attainment increases the likelihood of having a partner from a different ethnic background. Moreover, the set of sensitivity checks point to the strong stability of the results, and that they can be plausibly interpreted as causal evidence. The next step consists in understanding whether the effect of education on exogamy is heterogeneous according to the individual's and parental characteristics, and what could be the potential mechanisms that underlie the causal chain between HEI expansion, education and interethnic marriages. As for the first objective, Table 13 displays the results of the estimation of IV/TSLS with heterogeneous coefficients, in which we interacted years of schooling (and the instrument) with i) own ethnicity,[21] ii) the dummy for parental education and iii) the dummy for having parents with mixed ethnic background. The results indicate that the effect of education on exogamy does not significantly depend on parental education and having ethnically-mixed parents. However, the impact of schooling on the likelihood of having a partner from a different ethnic background is lower for individuals with Javanese ethnicity (the largest ethnic group in Java) than for those belonging to other ethnic groups, for whom we detect a larger effect of education on the propensity to interethnic marriage. This result points out that increased educational attainment induced by HEI expansion can reduce the segregation of ethnic minorities.

Concerning the analysis of potential mechanisms, we focus on reduced-form estimations that directly relate exposure to HEI surrounding the district of residence at 18 and the different variables that we consider, given data availability. Although we acknowledge that these variables are not ideal for this purpose, because they are observed possibly several years after marriage, we are still convinced that they deserve a certain attention and could highlight interesting patterns regarding potentially relevant channels. The results, reported in Table 14, indicate that exposure to HEI has a positive impact on the probability of changing place of residence between age 18 and 2014 (column (1)). Moreover, it also exerts a positive on the probability of moving to a large city (column (2)), where several ethnicities are more likely to coexist. Consistently, being exposed to more HEI at age 18 also increases ethnic fractionalization in the district of residence in 2010 (column (3)), although there is no impact on the probability of being an ethnic minority in the community of residence at the time of the survey (column (4)). This evidence indeed suggests that the relevant channel could be migration towards larger agglomerations, where the chances of matching with a person from a different ethnicity are higher, rather than constraints in the marriage market due to residing in enclaves with a very limited number of inhabitants from one's own ethnic group. Finally, we also obtain suggestive evidence regarding the role of changes in social norms. Specifically, individuals exposed to a higher number of HEI during their adolescence are less likely to trust (relatively) more others from the same ethnic group than their counterparts with a different ethnic background. This result highlights the relevance of higher education opportunities in shaping tolerance and trust towards other ethnicities, which could be one of the possible channels through which educational attainments favour the formation of interethnic marriages.

6. Conclusions

We investigated the effect of educational attainment on the formation of interethnic marriages in Indonesia, exploiting the expansion of Higher Education Institutions that took place in the country from the last half of the 20th century. We focused on the island of Java, the most populated island of the country, and where its capital (Jakarta) is located. The empirical analysis was carried out using data from the 2014 wave of the Indonesian Family Life Survey, combined with administrative information about the year of establishment and the exact location of HEIs that offer undergraduate degrees across this island. The main outcome variable is the probability of having a partner from a different ethnic background than one's own, i.e. exogamy. As for educational attainment, we considered three main measures: years of completed schooling, college attendance and college completion. To address the issue of endogeneity of education, we exploited variation by year of birth and district of residence at age 18 in geographical exposure to HEI in an Instrumental Variable framework.

The results indicate that education has a positive impact on the propensity to form an ethnically-mixed couple with somewhat higher point estimates for females than for males, although not statistically different. Specifically, each additional year of schooling increases the likelihood of exogamy by 3.5p.p. for males and 4.6p.p. for females, while the effects of college attendance/completion range between 44–52 p.p. and 71–73 p.p. for males and females, respectively (considering the baseline model without control variables). These results remain largely unchanged across different specifications and are robust to various sensitivity checks, providing supporting evidence for the validity of the Instrumental Variable approach and its underlying assumptions. We do not find evidence of heterogeneous effects of schooling on the propensity to form an interethnic marriage according to parental education or mixed parental ethnicity. However, the effect of education on exogamy is lower for individuals belonging to the largest ethnic group (Javanese) than their counterparts with other ethnic backgrounds. This evidence highlights the relevance of education as a tool to reduce segregation of ethnic minorities. Finally, the analysis of potential mechanisms reveals that migration/residential choices and changes in social norms are likely channels through which the expansion

[21] Here we grouped together Sundanese, Madurese, Betawi and other ethnicities due to their low number of observations, and used a dummy for belonging to the Javanese ethnicity.

Table 13
IV/TSLS with Heterogeneous Effects.

	Males			Females		
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Schooling	0.064**	0.038***	0.033***	0.072***	0.052***	0.042***
	(0.026)	(0.012)	(0.010)	(0.023)	(0.017)	(0.014)
Years of Schooling X Javanese	−0.039*			−0.036*		
	(0.023)			(0.022)		
Years of Schooling X Low Parental Education		−0.003			−0.010	
		(0.032)			(0.019)	
Years of Schooling X Ethnically-Mixed Parents			−0.033			−0.017
			(0.024)			(0.030)
First-Stage F-statistic	17.015	20.966	18.895	19.317	15.347	16.810
Observations	6391	6391	6391	6181	6181	6181

Notes: Standard errors (in parentheses) are clustered by district of residence at age 18. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects. The regression reported in columns (1) and (4) include as control a dummy for being Javanese (versus other ethnicities). The regression reported in columns (2) and (5) include as control a dummy for having low-educated parents. The regression reported in columns (3) and (6) include as control a dummy for having ethnically-mixed parents.

Table 14
Potential Mechanisms.

Dependent Variable:	Migrated (18–2014)	Migrated to Large Cities (18–2014)	Fractionalization (2010)	Being a minority (2014)	Trust Own Ethnicity (2014)
	(1)	(2)	(3)	(4)	(5)
Panel A: Males					
HEI within 10 km radius at age 18	0.020***	0.006***	0.045***	0.006	−0.019***
	(0.004)	(0.002)	(0.014)	(0.005)	(0.003)
Observations	6391	6391	6391	6391	4515
Panel B: Females					
HEI within 10 radius at age 18	0.019***	0.007***	0.046***	0.007	−0.024***
	(0.003)	(0.002)	(0.014)	(0.004)	(0.003)
Observations	6181	6181	6181	6181	4872

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

of higher education could foster the likelihood of interethnic marriage. Specifically, geographical exposure to HEI increases the propensity to migrate and reside in large cities (which are characterized by a higher degree of ethnic diversity), where ethnically-mixed marriages are more likely. Moreover, individuals exposed to a higher number of HEI during their adolescence are more likely to trust individuals from a different ethnic background. This result highlights the potential role of higher education opportunities on changing social norms and favouring inter-ethnic tolerance and social integration.

From the policy perspective, the results reported in this paper suggest that fostering the formation of human capital through an increase in opportunities for higher education driven by the expansion of the college education infrastructure is likely to be beneficial for several reasons. This is because a wider presence of HEI across the territory, leading to higher educational attainments, could generate positive impacts not only at the individual level in terms of earnings potential and labour market outcomes, but also in health status and other socioeconomic outcomes. Indeed, the increase in education driven by the expansion of HEI can foster changes in social norms that are likely to break existing ethnic-related barriers, promote a sense of unity and reduce ethnic segregation in multi-ethnic societies.

CRediT authorship contribution statement

Antonio Di Paolo: Writing – review & editing, Writing – original draft, Methodology, Formal analysis. **Khalifany-Ash Shidiqi:** Writing – review & editing, Writing – original draft, Methodology, Data curation.

Funding

The research presented in this paper did not benefit from any research funding.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

we acknowledge the support provided by Romi Bhakti Hartarto and Wahyu Tri Wibowo for their support with the data about Higher Education Institutions in Indonesia. Participants at the IESEG seminar, UniZar semina, EALE, ESPE, AIEL and GLO-Berlin Conferences provided useful comments. The usual disclaimers apply.

Statements and declarations

Data Availability Statement: the data used in this manuscript are freely available. The empirical analysis is based on information from the Indonesian Family Life Survey database (see here: https://www.rand.org/well-being/social-and-behavioral-policy/data/FLS/IFLS.html) and an external database constructed by the authors about the location and year of establishment of Higher Education Institution in the Java Island.

Researchers who wish to access the IFLS dataset must register and accept the conditions for downloading the data: https://www.rand.org/well-being/social-and-behavioral-policy/data/FLS/IFLS/access.html.

The authors will make their own-constructed dataset on the location and year of establishment of Higher Education Institutions on Java Island, as well as the do-file for potential replication analysis, available to other researchers through the open Research Data Repository (RDR.

available here https://cora.csuc.cat/en/rdr-research-data-repository/).

Appendix 1

Table A1a
Complete OLS Results – Males.

	Years of Education		HE Attendance		HE Completion	
	(1)	(2)	(3)	(4)	(5)	(6)
Estimate	0.007***	0.006***	0.076***	0.058***	0.074***	0.058***
	(0.001)	(0.001)	(0.015)	(0.013)	(0.016)	(0.014)
Ethnicity						
I(Javanese)	*Reference Category*					
I(Sundanese)		0.010		0.011		0.011
		(0.033)		(0.034)		(0.034)
I(Maduranese)		0.082***		0.077**		0.077**
		(0.030)		(0.030)		(0.029)
I(Betawi)		0.102**		0.103**		0.103**
		(0.048)		(0.049)		(0.049)
I(Other Ethnicities)		0.329***		0.330***		0.332***
		(0.050)		(0.050)		(0.050)
Religion						
I(Islam)	*Reference Category*					
I(Christian)		−0.023		−0.021		−0.018
		(0.028)		(0.027)		(0.028)
I(Hindu)		−0.263*		−0.264*		−0.254*
		(0.147)		(0.144)		(0.136)
I(Other Religions)		0.183		0.164		0.162
		(0.137)		(0.135)		(0.135)
Number of Siblings		−0.003		−0.002		−0.002
		(0.002)		(0.002)		(0.002)
Low Parental Education		−0.015		−0.036***		0.036***
		(0.012)		(0.013)		(0.013)
Ethnically-Mixed Parents		0.211***		0.211***		0.211***
		(0.019)		(0.019)		(0.019)
R-squared	0.166	0.227	0.164	0.225	0.163	0.225
Controls	No	Yes	No	Yes	No	Yes
Observations	6391	6391	6391	6391	6391	6391

Notes: OLS estimations with exogamy as outcome variable (i.e. having a partner with a different ethnicity than the individual). Main regressors: years of schooling (Panel A), university attendance (Panel B), and university completion (Panel C). Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

Table A1b
Complete OLS Results – Females.

	Years of Education		HE Attendance		HE Completion	
	(1)	(2)	(3)	(4)	(5)	(6)
Estimate	0.006***	0.004***	0.064***	0.049***	0.066***	0.050***
	(0.002)	(0.002)	(0.019)	(0.018)	(0.020)	(0.019)
Ethnicity						
I(Javanese)	*Reference Category*					
I(Sundanese)		0.020		0.021		0.021
		(0.032)		(0.032)		(0.032)
I(Maduranese)		0.056*		0.049*		0.050*
		(0.030)		(0.029)		(0.029)
I(Betawi)		0.045		0.046		0.045
		(0.046)		(0.046)		(0.046)
I(Other Ethnicities)		0.261***		0.261***		0.260***
		(0.048)		(0.048)		(0.048)
Religion						
I(Islam)	*Reference Category*					
I(Christian)		0.020		0.021		0.021
		(0.025)		(0.025)		(0.025)
I(Hindu)		−0.096		−0.100		−0.102
		(0.222)		(0.221)		(0.221)
I(Other Religions)		−0.037		−0.044		−0.044

(*continued on next page*)

Table A1b (*continued*)

	Years of Education		HE Attendance		HE Completion	
	(1)	(2)	(3)	(4)	(5)	(6)
		(0.176)		(0.175)		(0.175)
Number of Siblings		0.001		0.002		0.002
		(0.002)		(0.002)		(0.002)
Low Parental Education		−0.023**		−0.036***		−0.036***
		(0.010)		(0.009)		(0.009)
Ethnically-Mixed Parents		0.186***		0.187***		0.187***
		(0.024)		(0.024)		(0.024)
R-squared	0.171	0.207	0.170	0.207	0.170	0.207
Controls	No	Yes	No	Yes	No	Yes
Observations	6181	6181	6181	6181	6181	6181

Notes: OLS estimations with exogamy as outcome variable (i.e. having a partner with a different ethnicity than the individual). Main regressors: years of schooling (Panel A), university attendance (Panel B), and university completion (Panel C). Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

Table A2
Robustness check – minimum distance as alternative instrument.

	Males		Females	
Panel A: First Stage — Dependent Variable: Years of Schooling				
HEI within 10 km radius at age 18	0.235***		0.241***	
	(0.037)		(0.041)	
distance to the nearest HEI at age 18		−0.031***		−0.027***
		(0.008)		(0.009)
First-Stage F-statistic	39.635	13.186	34.813	8.522
Panel B: Second Stage — Dependent Variable: Exogamy				
Years of Schooling	0.035***	0.034***	0.046***	0.036**
	(0.011)	(0.010)	(0.015)	(0.015)
Observations	6391	6391	6181	6181

Notes: Standard errors (in parentheses) are clustered by district of residence at age 18. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

Table A3
Robustness check – excluding individuals who changed district of residence (0–18).

	Males		Females	
	Baseline	Never Move (0—18)	Baseline	Never Move (0—18)
Panel A: First Stage — Dependent Variable: Years of Schooling				
HEI within 10 km radius at age 18	0.235***	0.231***	0.241***	0.246***
	(0.037)	(0.039)	(0.041)	(0.040)
First-Stage F-statistic	39.635	35.143	34.813	37.693
Panel B: Second Stage — Dependent Variable: Exogamy				
Years of Schooling	0.035***	0.034***	0.046***	0.041***
	(0.011)	(0.011)	(0.015)	(0.014)
Observations	6391	6257	6181	6066

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

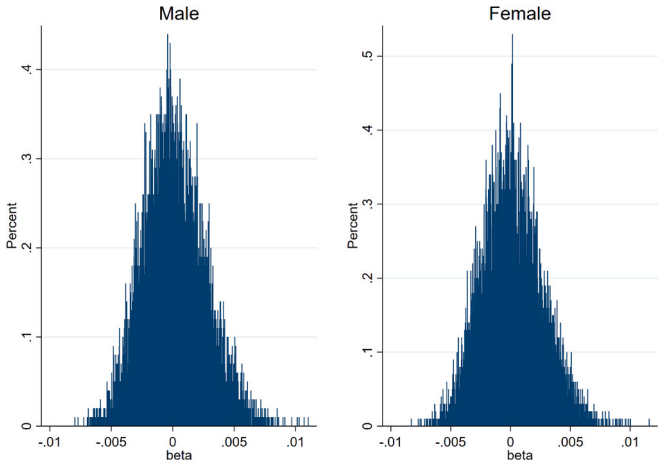

Fig. 1a. Fake reduced form coefficient – permutation test with random assignment of districts of residence at 18.

Table A4
Overidentified IV/TSLS with dummies for the number of HEI.

	Males		Females	
	(1)	(2)	(3)	(4)
Panel A: First Stage — Dependent Variable: Years of Schooling				
HEI within 10 km radius	0.235***		0.241***	
	(0.037)		(0.041)	
I(HEI within 10 km radius at age 18 = 0)		*reference category*		*reference category*
I(HEI within 10 km radius at age 18 = 1)		1.817***		1.690***
		(0.355)		(0.482)
I(HEI within 10 km radius at age 18 = 2)		0.433		0.512
		(0.568)		(0.601)
I(HEI within 10 km radius at age 18 = 3)		0.769		1.192
		(0.527)		(0.828)
I(HEI within 10 km radius at age 18 = 4)		2.017***		1.966***
		(0.396)		(0.419)
I(HEI within 10 km radius at age 18 = 5)		3.522***		1.871**
		(0.487)		(0.858)
I(HEI within 10 km radius at age 18 = 6)		1.349*		1.346
		(0.697)		(0.888)
I(HEI within 10 km radius at age 18 = 7)		2.472***		1.910***
		(0.372)		(0.364)
I(HEI within 10 km radius at age 18 = 8)		2.176***		2.666***
		(0.554)		(0.694)
I(HEI within 10 km radius at age 18 = 9)		1.329***		1.930***
		(0.397)		(0.460)
I(HEI within 10 km radius at age 18 ≥ 10)		2.194***		2.312***
		(0.315)		(0.324)
First-Stage F-statistic	39.635	13.199	34.813	7.996
P-Value(1st-Stage F-statistic)	0.000	0.000	0.000	0.000
Panel B: Second Stage — Dependent Variable: Exogamy				
Years of Schooling	0.035***	0.027***	0.046***	0.032***
	(0.011)	(0.008)	(0.015)	(0.011)
P-Value(Hansen J statistic)		0.417		0.387
Observations	6391	6391	6181	6181

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects.

Table A5
Robustness check – removing individuals who married before completing education.

	Males		Females	
	Baseline	Married after completing education	Baseline	Married after completing education
	(1)	(2)	(3)	(4)
Panel A: First Stage — Dependent Variable: Years of Schooling				
HEI within 10 radius at age 18	0.235***	0.238***	0.241***	0.253***
	(0.037)	(0.037)	(0.041)	(0.040)
First-Stage F-statistic	39.635	40.832	34.813	39.498
Panel B: Second Stage — Dependent Variable: Exogamy				
Years of Schooling	0.035***	0.035***	0.046***	0.042***
	(0.011)	(0.011)	(0.015)	(0.014)
Observations	6391	6355	6181	5982

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects. Estimates in columns (2) and (4) are obtained after excluding individuals who married before the year in which they completed education (= year of birth + 6 + years of schooling).

Table A6
Robustness check – removing individuals who changed ethnicity between 2007 and 2014.

	Male		Female	
	Baseline	Same Ethnicity	Baseline	Same Ethnicity
	(1)	(2)	(3)	(4)
Panel A: First Stage — Dependent Variable: Years of Schooling				
HEI within 10 radius at age 18	0.235***	0.225***	0.241***	0.206***
	(0.037)	(0.035)	(0.041)	(0.038)
First-Stage F-statistic	39.635	41.187	34.813	29.593
Panel B: Second Stage — Dependent Variable: Exogamy				
Years of Schooling	0.035***	0.036***	0.046***	0.046***
	(0.011)	(0.012)	(0.015)	(0.015)
Observations	6391	4461	6181	4563

Notes: Standard errors (in parentheses) are clustered by district of residence at 18. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. All regressions control for year of birth dummies × province of residence (at age 18) fixed effects. Estimates reported in columns (2) and (4) are obtained after retaining only individuals who report the same ethnicity in 2014 than in 2007 and are interviewed in both waves of IFLS.

Data availability

The authors do not have permission to share data.

References

Alesina, A., Giuliano, P., & Reich, B. (2021). Nation-building and education. *The Econometrics Journal, 131*(638), 2273–2303.

Allendorf, K., & Thornton, A. (2015). Caste and choice: The influence of developmental idealism on marriage behavior. *The American Journal of Sociology, 121*(1), 243–287.

Bandiera, O., Mohnen, M., Rasul, I., & Viarengo, M. (2019). Nation-building through compulsory schooling during the age of mass migration. *The Econometrics Journal, 129*(617), 62–109.

Bandyopadhyay, S., & Green, E. (2021). Explaining inter-ethnic marriage in Sub-Saharan Africa. *Journal of International Development, 33*(4), 627–643.

Barro, R. J. (2001). Human capital and growth. *The American Economic Review, 91*(2), 12–17.

Bazzi, S., Gaduh, A., Rothenberg, A. D., & Wong, M. (2019). Unity in diversity? how intergroup contact can foster nation building. *The American Economic Review, 109*(11), 3978–4025.

Belskaya, V., Sabirianova Peter, K., & Posso, C. M. (2020). Heterogeneity in the effect of college expansion policy on wages: Evidence from the russian labor market. *Journal of Human Capital, 14*(1), 84–121.

Buchori, M., & Malik, A. (2004). The evolution of higher education in Indonesia. In P. G. Altbach (Ed.), *Asian universities: Historical perspective and contemporary challenges*. Baltimore: The Johns Hopkins University Press.

Buttenheim, A. M., & Nobles, J. (2009). Ethnic diversity, traditional norms, and marriage behaviour in Indonesia. *Population Studies, 63*(3), 277–294.

Carneiro, P., Liu, K., & Salvanes, K. G. (2023). The supply of skill and endogenous technical change: Evidence from a college expansion reform. *Journal of the European Economic Association, 21*(1), 48–92.

Corvalan, A., & Vargas, M. (2015). Segregation and conflict: An empirical analysis. *Journal of Development Economics, 116*, 212–222.

Crespin-Boucaud, J. (2020). Interethnic and interfaith marriages in sub-Saharan Africa. *World Development, 125*, Article 104668.

Currie, J., & Moretti, E. (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *Quarterly Journal of Economics, 118*(4), 1495–1532.

Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *The American Economic Review, 91*(4), 795–813.

Esteban, J., Mayoral, L., & Ray, D. (2012). Ethnicity and conflict: An empirical study. *The American Economic Review, 102*(4), 1310–1342.

Fiorini, M., & Stevens, K. (2021). Scrutinizing the monotonicity assumption in IV and fuzzy RD designs. *Oxford Bulletin of Economics and Statistics, 83*(6), 1475–1526.

Furtado, D., & Song, T. (2022). Interethnic marriages. In *Handbook of Labor, Human Resources and Population Economics* (pp. 1–19). Cham: Springer International Publishing.

Ida Bagus, M. (2008). Mixed marriages in jembrana, bali: Mediation and fragmentation of citizenship and identity in the post-bomb(s) bali world. *Asia Pacific Journal of Anthropology, 9*(4), 346–362.

Kamhöfer, D. A., Schmitz, H., & Westphal, M. (2019). Heterogeneity in marginal non-monetary returns to higher education. *Journal of the European Economic Association, 17*(1), 205–244.

Kukić, L. (2023). The last Yugoslavs: Ethnic diversity and national identity. *Explorations in Economic History, 88*, Article 101504.

Kyui, N. (2016). Expansion of higher education, employment and wages: Evidence from the Russian transition. *Labour Economics, 39*, 68–87.

Jagnani, M., & Khanna, G. (2020). The effects of elite public colleges on primary and secondary schooling markets in India. *Journal of Development Economics, 146*, Article 102512.

Mukminin, A., Habibi, A., Prasojo, L. D., Idi, A., & Hamidah, A. (2019). Curriculum reform in Indonesia: Moving from an exclusive to inclusive curriculum. *Center for Educational Policy Studies Journal, 9*(2), 53–72.

Ngo, J., & Meek, L. (2019). Higher education governance and reforms in indonesia: Are the matrices of autonomy appropriate? *Journal of International and Comparative Education, 8*(1), 17–26.

Pannen, P. (2018). Higher Education Systems and institutions, Indonesia. In A. Amaral (Ed.), *Encyclopedia of International Higher Education Systems and Institutions* (pp. 1143–1151). Dordrecht: Springer.

Parker, L., Hoon, C. Y., & Raihani. (2014). Young people's attitudes towards inter-ethnic and inter-religious socializing, courtship and marriage in Indonesia. *South East Asia Research, 22*(4), 467–486.

PDSP Kemdikbud (2013) Statistik Perguruan Tinggi (PT) 2012/2013. Available at: https://publikasi.data.kemdikbud.go.id/upload/file/isi_11800868-A25A-4153-BBBB-C416C54BCA5F_.pdf.

Rademakers, R., & van Hoorn, A. (2021). Ethnic switching: Longitudinal evidence on prevalence, correlates, and implications for measuring ethnic segregation. *Journal of Development Economics, 152*, Article 102694.

Ray, T., Chaudhuri, A. R., & Sahai, K. (2020). Whose education matters? an analysis of inter caste marriages in India. *Journal of Economic Behavior & Organization, 176*, 619–633.

Rohner, D., and Saia, A. (2019). Education and Conflict: Evidence from a Policy Experiment in Indonesia. CEPR Discussion Paper No. 13509. https://cepr.org/publications/dp13509.

Roth, C., & Sumarto, S. (2015). Does education increase interethnic and interreligious tolerance? Evidence from a natural experiment. MPRA Paper No. 64558. https://mpra.ub.uni-muenchen.de/64558/.

Sanjaya, M. R., Chuah, S. H., Feeny, S., & Hoffmann, R. (2023). The impact of cultural heterogeneity on violence in Indonesia: Fractionalisation versus polarization. *Applied Economics, 55*(16), 1790–1806.

Statistics Indonesia (2010) *Sensus Penduduk 2010*. Available at: https://sensus.bps.go.id/main/index/sp2010.

Strauss, J., Witoelar, F. and Sikoki, B. (2016) The Fifth Wave of the Indonesia Family Life Survey: Overview and Field Report: Volume 1, WR-1143/1-NIA/NICHD. RAND Working Papers.

Utomo, A. J. (2019). Love in the melting pot : Ethnic intermarriage in Jakarta. *Journal of Ethnic and Migration Studies, 46*(14), 1–18.

Utomo, A., & McDonald, P. (2016). Who marries whom? Ethnicity and marriage pairing patterns in Indonesia. *Asian Population Studies, 12*(1), 28–49.

Utomo, A. J., & McDonald, P. F. (2021). Internal migration, group size, and ethnic endogamy in Indonesia. *Geographical Research, 59*(1), 56–77.

Valero, A., & Van Reenen, J. (2019). The economic impact of universities: Evidence from across the globe. *Economics of Education Review, 68*, 53–67.

Westphal, M., Kamhöfer, D. A., & Schmitz, H. (2022). Marginal college wage premiums under selection into employment. *The Economic Journal, 132*(646), 2231–2272.

Welch, A. R. (2007). Blurred vision?: Public and private higher education in Indonesia. *Higher Education, 54*(5), 665–687.

World Bank. (2023). *Mean age at first marriage*. Gender Data Portal. Available at: https://genderdata.worldbank.org/indicators/sp-dyn-smam/.