English Language Premium in a Marriage Market:
Experimental Evidence from Delhi

A Research Paper presented by:

Tomohisa Miyamoto
(Japan)

in partial fulfilment of the requirements for obtaining the degree of
MASTER OF ARTS IN DEVELOPMENT STUDIES

Major:

Economics of Development
(ECD)

Specialization:

Econometric Analysis of Development Policy

Members of the Examining Committee:

Professor Dr. Arjun Singh Bedi
Professor Dr. Lorenzo Pellegrini

The Hague, The Netherlands
December 2021

Disclaimer:

This document represents part of the author's study programme while at the International Institute of Social Studies. The views stated therein are those of the author and not necessarily those of the Institute.

Inquiries:

International Institute of Social Studies
P.O. Box 29776
2502 LT The Hague
The Netherlands

t: +31 70 426 0460
e: info@iss.nl
w: www.iss.nl
fb: http://www.facebook.com/iss.nl
twitter: @issnl

Location:

Kortenaerkade 12
2518 AX The Hague
The Netherlands

Contents

List of Tables

List of Figures

List of Acronyms

LC	Lower Caste
UC	Upper Caste
SC	Scheduled Caste
IHDS	India Human Development Survey

Acknowledgement

I have received guidance and support from various people in conducting this research. Here, I would like to express my gratitude.

This research was made possible thanks to the precise supervision of Professor Arjun Bedi and Professor Lorenzo Pellegrini. Professor Arjun Bedi taught me the basics of econometrics in an easy-to-understand manner through the coursework at ISS and provided me with detailed support from the idea of the research topic to the realization of the experiment and the writing of the paper. His extensive experience with econometric analysis and experimental approach, as well as his deep understanding of the context of the study made this study alive. I am very grateful to him, and I am very happy that I was able to enrol in the ISS and receive his guidance.

I also offer my sincere gratitude to my second reader, Professor Lorenzo Pellegrini. Over the two seminars, he gave me very critical comments and suggestions that made this paper logical and understandable.

This study was made possible also by a grant from the ECD field research fund at the ISS. I would like to thank Professor Matthias Rieger and Professor Elissaios Papyrakis for accepting this research proposal and advising me on the experimental design.

Additionally, I am grateful for the excellent research assistantship by Smiti. She played an integral role in the data collection and logistics of the field experiment from Delhi, India. She also helped me with my literature search from a sociolinguistics perspective.

I would also like to thank my coursemates who worked hard with me during my master's program at ISS. In particular, I would like to thank Harsha Jain, Deepakshi Singh and Ngan Nguyen for being my discussant and clearing my doubts many times. Also, I could not have gotten through my studies under the severe situation of pandemic without Edwin and Sarah.

The motivation for this research stems from my undergraduate studies on South Asian Studies in Tokyo and my experiences in India and Nepal. In this study, I used the knowledge of Hindi that I acquired during this period. I would like to thank all the people who have helped me during this time.

Finally, it goes without saying that without the full support of my family, I would not have been able to complete my master's program, nor would I have been able to carry out this research. Once again, I would like to express my deepest gratitude for their presence.

Abstract

In India, English has become a status symbol for the wealthy and educated middle class, while at the same time creating a divide between those who speak it and those who do not. This study examines the social returns to English language skills in the marriage market in urban India by conducting a correspondence experiment on an online matrimonial website. I find that indicating strong English language skills in a profile increases the number of people interested in the profile by 38%. Further, the probability of matching with a high-caste man (marry-up) increased by 57.7 percentage points when a low-caste female profile mentions English proficiency. The results suggest that English language proficiency can expand one's choice of potential marriage partners in the marriage market and help overcome caste barriers

Relevance to Development Studies

The development issues that this study addresses are the disparity between English speakers and non-English speakers, and caste inequality. In the Indian context, differences in English proficiency arise mainly from differences in the language of instruction between private and government schools. The educational gap between private and government schools has been widely studied, but the consequences of the difference in the medium language of education have not been fully understood. This study intends to contribute to the formation of an equitable language policy by clarifying the returns to language proficiency and generating suggestions for diminishing economic disparity due to educational disparity. Also, by examining English proficiency as a factor in promoting inter-caste marriage, this study discusses social mobility in the caste hierarchy.

Keywords

English, Return to language skills, Caste, Inter-caste marriage, social mobility, Marriage market, Field experiment

Chapter 1
Introduction

In the 2017 film "Hindi Medium", the last film of the great actor Irrfan Khan, a couple in Delhi disguise themselves as a poor family and struggle to get their daughter admitted to a posh English medium school[1], even moving to a poor neighbourhood (Chaudhary, 2017). The couple wants to avoid at all costs their daughter being enrolled in a government school where Hindi is the language of instruction. In the pivotal scene of the film, the wife exclaims *"She won't learn anything in a government school. She'll be terrified if anyone speaks in English. She will be a misfit in society. English isn't just a language in this country, it's a 'class'!"*

In India as well as in many other countries, English has become an essential skill for better jobs and income. In response, the popularity of English medium private schools is on the rise. Reflecting this situation, being able to speak English has become a status symbol. Prior research has shown that the returns to education in English are higher than the returns to education in local languages(Munshi and Rosenzweig, 2006; Kapur and Chalraborty, 2016), while there is a strong correlation between subjective English proficiency and a person's income(Azam, Chin and Prakash, 2012). However, there is scarce research on non-monetary returns to English language proficiency, and there are no studies which have identified the causal influence of English language skill on any social outcome in the Indian context. In particular, the returns to English language proficiency for women, who are less likely to be engaged in wage labour, is also unknown.

To explore the social aspect of returns to English languages, I will focus on a particular aspect of social return, that is, returns to English language skills in the marriage market. The question this study seeks to answer is: Does English proficiency enable matching with higher status partners? To respond to this question, I conducted a correspondence experiment on a major Indian matrimonial website, which is one of the primary means of finding a marriage partner in the Indian marriage market. The experiment consisted of two stages. In the first phase, I constructed a set of fictitious female profiles with different levels of English proficiency, caste, educational background, and annual income, and observed the number of visitors to the profile and the number of interests received from other profiles as outcomes. In the second phase, I sent interest from the fictitious female profiles to existing (real) male profiles and observed the acceptance of interest as a preference for marriage partners. The experiment was conducted using profiles of middle-class Hindi-speaking people living in Delhi. From the information in the matrimonial profiles, the status of potential marriage candidates was observed in terms of caste, annual income, and educational background. The data collection was carried out in August and September 2021 and involved a total of 384 candidates.

My econometric strategy exploits the exogenous nature of English proficiency and other factors in the fictitious profiles. I first created a profile of a middle-class woman with a standard occupational and family background, and self-introduction, to which I assigned 3 educational backgrounds, 4 annual incomes, and 2 caste status. In addition to the 24 profiles thus created, I prepared 24 profiles that were the same, but on these profiles, I signalled strong English language proficiency. Thus, the profiles with and without English proficiency were on average exactly same with respect to the other attributes, and there was no

[1] They tried to enrol in private schools using the admission quota reserved for the poor, under the Right to Education Act (RTE). This policy is detailed in Chudgar and Creed (2016), Rao (2019), and Joshi (2020).

correlation between any of the assigned attributes and English language skills. Therefore, the difference in the number of visitors and interest received by posting such profiles is indicative of the average causal effect of indicating English proficiency in the profile. Another strategy is the random assignment of existing profiles to which interests were sent. In the experiment, the existing profile receiving the interest was also exogenously assigned to each fictitious profile. In this way, I was able to observe decisions regarding matching between different castes, income groups, and educational level. Unlike in reality, here the nature of the female profile that sends the interest is irrelevant to the nature of the male profile that receives it. Consequently, the difference in interest acceptance rate may be interpreted as an effect of the features manipulated in the fictitious profile, without being biased by the features of the other profiles' attributes.

The results of the experiment showed that English language proficiency translates into a substantial advantage for prospective brides in the marriage market. Female profiles with English proficiency received 20% more visitors and 38% more interest than female profiles without English proficiency. On average, this means that 17 more male profiles were interested in marrying the women's profiles when they included English language skills. English proficiency has the effect of attracting more potential marriage partners and widening the choices faced by prospective brides.

Furthermore, signalling English proficiency increased the probability of matching with high-caste men for low-caste women. Interest acceptance rates were lowest in the caste hypergamy combinations where the female side was low caste, and the male side was high caste. However, the probability of an interest being accepted increased by 57.7 percentage point when English proficiency was mentioned in the female profile. English proficiency may help to undermine caste barriers by increasing the likelihood of inter-caste marriage with high-caste men, compensating for the disadvantages of caste status for low-caste women. On the other hand, for high-caste women, signalling English skills negatively influenced the acceptance rate of interests.

This study relates to two bodies of literature. The first one is studies on returns to a language skill. Studies on the economic returns to several language skills show that knowledge of English and other dominant languages raises one's earnings (English in the US (Chiswick and Miller, 1995; Bleakley and Chin, 2004), UK (Aoki and Santiago, 2015), India (Azam, Chin and Prakash, 2012), Israel (Lang and Siniver, 2006), South Africa (Levinsohn, 2004)Kazakh and Russian in Kazakhstan (Aldashev and Danzer, 2014), English, Russian, French, German, and Arabic in Turkey (Di Paolo and Tansel, 2015)). Also, several studies have found that teaching in a European/foreign language has higher returns in the labour market than teaching in the local language in many contexts (French and Arabic in Morocco (Angrist and Lavy, 1997), English and Spanish in Puerto Rico (Angrist, Chin and Godoy, 2006), English and Marathi (Munshi and Rosenzweig, 2006), English and Bengali in India (Kapur and Chalraborty, 2016), English and Malay in Malaysia (Parinduri and Ong, 2018)). I add to this literature by examining return to language skills in a marriage market. More broadly, this study is situated in policy research on languages. In recent years, there has been a growing body of quantitative research suggesting the importance of language in education or the labour market, that contribute to language policy making (Nair, 2015; Edgerton, 2016; Jain, 2017; Jain, Maitra and Mani, 2019; Ginsburgh and Weber, 2020; Yuki, 2021). The second body of literature is the study of marriages between different groups, especially inter-caste marriages. A recent study by Ray, Roy and Sahai (2020) showed that the incidence of inter-caste marriage was not associated with educational level of the couples, but it is positively associated with the educational level of the groom's mother. Possibly contrary to their result, this study shows that English proficiency of low caste women increases the probability of inter-caste marriage.

The rest of this paper proceeds as follows. In chapter 2, I explain the background context regarding languages, caste, and marriage market that are peculiar to India. Chapter 3 reviews empirical literature on returns to languages, and factors that affect marriage outcomes including inter-caste marriage. In chapter 4, I explain the empirical strategy and hypotheses, including details of the experiment. Chapter 5 provides description of the data collected from the experiment. Chapter 6 reports the return to English language skills, measured in number of views, interests, as well as interest acceptance rates. Chapter 7 summarizes the results and discusses limitations and possible implications for future research, as well as real world implications. Chapter 8 concludes.

Chapter 2
Background

This chapter describes the background context of India related to this study regarding the following three topics: the status of English language in India, caste system, and the marriage market with a high incidence of arranged marriage.

2.1 English and Language Issues in India

In India, a multilingual society, Hindi is the official language at the federal level as stated by the Constitution. However, at the same time, English has been granted the status of associate official language (Sandhu, 2014) ever since India's independence in 1947. At the level of the state governments, some 23 regional languages recognized in the 8th schedule of the constitution are used as official languages, with English and Hindi being the primary means of communication between states and between the states and the federal government.

Since the British colonial period, the language of instruction for higher education has been English, and to date, a majority of higher education is conducted in English. Consequently, English skills have become essential for obtaining white-collar jobs, including in the public sector. In its education policy, the government of India places English language as a part of a three language formula, along with Hindi and a state language (Jain, Maitra and Mani, 2019). Especially through the expansion of education since 1990s and the mushrooming of English medium private schools, the number of people who acquire English through education is expected to have increased. However, English is almost always used as a second or third language [2].

Nevertheless, the percentage of people who speak English fluently is much lower than is often perceived. According to Azam, Chin and Prakash (2012), based on the India Human Development Survey 2005, only 4% among adults aged between 18-65 can converse in fluent English and 14% can converse little in English. The proportion of people who speak some English is higher for men (26%) than women (14%) and in urban areas (35%) than rural areas (14%). English language proficiency is highly associated with higher education. Individuals with a bachelor's degree are able to speak English at a rate of 89%, compared to 56% for those who have completed secondary schooling. As Azam, Chin and Prakash (2012) mention, it is possible to complete secondary school or college education without being able to speak English, as exam may be written in Hindi or other state languages except in science and engineering stream.

2.2 Indian caste system

This section provides some background details about the caste system in India, to the extent that it is needed to provide a context for the analysis presented in this paper.

Caste is a status system closely related with Hinduism. Although some religions in South Asia other than Hinduism also have a caste system, this study discusses caste in Hinduism. Caste is characterized by three core elements: (1)hereditary group based on occupation, (2) highly endogamous group, and (3)hierarchical structure among castes (Bidner and Eswaran, 2015) hence caste itself signals social status. Caste consists of broad groups (*varna*),

[2] In IHDS 2005, only 0.2% answered English as their mother tongue.

and subgroups (*jati*). The varna category consists of *Brahmin* (priests), *Kshatriya* (warrior caste), *Vaisha* (merchant), and *Shudra* (laborer), which are ranked according to the concept of purity, while people belong to the category of *Adi-Shudra* or untouchable, contemporary more often referred to as *Dalit*, is placed below the four varna categories. Within each category, there are numerous (about 4000, according to Munshi (2019)) subgroups (*jati*) based on traditional occupational groups with their own hierarchy. People belong to a same *jati* are often concentrated in one or a few often neighbouring states, but are scattered at the village, town, and city levels within the same region (Munshi and Rosenzweig, 2015).

At the time of India's independence in 1947, the Constitution banned discrimination based on caste. With economic growth, expansion of education, and the rapid increase in non-traditional occupations, the shackles of caste and occupational choice have been relaxed and the correlation between one's caste and economic level has weakened (Munshi and Rosenzweig, 2006; Hnatkovska, Lahiri and Paul, 2012; Asher, Novosad and Rafkin, 2021). The government's affirmative action, which is aimed at making up for historical inequality based on caste, has also played an important role in improving inequalities in education and economic levels among castes. Among caste categories, Shudra and Dalit people in particular are subject to reverse discrimination in the form of affirmative action for Scheduled Castes (SCs). The affirmative action also includes ethnic indigenous group categorized as Scheduled Tribes (STs), and groups that were more recently recognized as economically disadvantaged (Other Backward Castes, OBCs). Castes not included in the affirmative schemes, the most people from *Brahman*, *Kshatriya* and *Vaisha* caste, are often referred as "forward castes".

Despite the weakening link between an individual's caste and socioeconomic status, caste-based discrimination is still prominent in present labour market (Banerjee and Knight, 1985; Kijima, 2006; Madheswaran and Attewell, 2007; Banerjee *et al.*, 2009; Ito, 2009; Madheswaran and Singhari, 2017). Madheswaran and Singhari's (2017) analysis found that wages are 19.4% lower in the public sector and 31.7% lower in the private sector for scheduled castes workers compared to forward caste workers, holding other individual characteristics constant. The same study also points out that the unequal access to occupations account for the wage gap larger than wage gap within an occupation. However, Munshi (2019) argues that this evidence of caste-based discrimination is possibly due to statistical discrimination where caste serves as a proxy for other observed and unobserved socioeconomic characteristics, or due to own-caste preference under competition when upper caste people comprise a larger percentage of decision makers. Thus, Munshi discusses that caste-based discrimination may disappear as socioeconomic differences between caste converge.

While it is appropriate to focus on the subgroups (*jati*) to analyse the function of caste as a network or an insurance, it is more appropriate to use the broader caste classification (*Varna*), which has a clearer hierarchy, to analyse class, hierarchical, and discriminative aspects of caste (Munshi, 2019). Following this practice in the literature, this study analyses caste using two broad categories of caste: Brahmin and Scheduled Castes (SC). In this study, Brahmin caste is used to represent upper class castes (UC), while Scheduled caste represents lower class castes (LC).

2.3 Marriage market in India

This section summarizes the meaning and structure of two distinct characteristics of the Indian marriage market: same-caste marriage and arranged marriages. I also discuss methods of searching for marriage partners that make these features possible today.

Marrying someone who belongs to the same caste (*jati*) is an important norm of the caste system (caste endogamy). Caste grouping and sense of belonging to a caste have been reproduced and maintained through adherence to this caste-endogamy norm (Allendorf and

Pandian, 2016; Titzmann, 2018; Ray, Roy and Sahai, 2020). From an economic perspective, marriages within the same community, including caste endogamy, have been explained as a behaviour to maintain the network and gain insurance from it. Rosenzweig and Stark (1989) argue that women's migration through marriage to other villages with different agricultural risks is a mutually insuring contract for both families. They found that when women migrate to a distant village through marriage in rural South India, there is a consumption smoothing effect for both the woman's parents side family and the groom's household. Further, Munshi and Rosenzweig (2009) pointed out that intra-*jati* marriages maintain community networks and play a role in risk-sharing, and that is why groups with higher average income are less likely to marry outside the community. From these findings, intra-caste marriage can be seen as a strategic decision for households.

Caste has traditionally been both an occupational group and a marital group. Although the association between caste and income or its role in determining jobs are much lowered, the practice of caste endogamy remains almost unchanged in strength (Munshi and Rosenzweig, 2006; Banerjee *et al.*, 2013; Munshi, 2019), , and inter-caste marriages are very limited. Only 11% of married couples were intermarried in The National Family Health Survey in 2006 (Allendorf and Pandian, 2016). Ray et al., (2020) found an even lower percentage of inter-caste marriages, 5.82% from the 2011 India Human Development Survey.

The reason for the traditional abstinence from inter-caste marriage is partly explained by the punishment for inter-caste marriage. A person who marries someone from a caste different from the one to which he or she belongs may face persecution and marginalization from the caste to which he or she belongs and may lose the benefits of the community network to which he or she belongs. However, this punishment is gender asymmetric (Bidner and Eswaran, 2015). When a man and a woman from different castes marry, the woman's caste is usually assumed to change to the man's caste. If the male is of a higher caste, the marriage is called caste-hypergamy, meaning that the woman gains a higher caste status through marriage (Ahuja and Ostermann, 2016; Titzmann, 2018).

In contrast to the actual infrequency of the inter-caste marriages, people's perception toward inter-caste marriages is turning more tolerant, and the number of people interested in inter-caste marriage is increasing. Ahuja & Ostermann (2016) cite the 2004 Indian National Election Study, where 60% of 27,000 voters supported a ban of inter-caste marriage, however, the rate was 47% in urban areas. Also, their data compiled from newspaper matrimonial advertisement shows that the number of newspaper marriage advertisements limiting marriage partners to the same caste has decreased from 30% in 1970s to 19% in 2000s. The results of their data collection from matrimonial websites also shows high interests in inter-caste marriage, 70.7% in SCs and 53.9% amongst UC individuals.

Another feature of the Indian marriage market is the high proportion of arranged marriages. As marriage is traditionally performed as a "family to family relationship" to maintain one's caste, the decision-making process is not limited to the person getting married. This is evident in the large number of arranged marriage in India. The term "arranged marriage" refers to a marriage in which parents or other relatives play the primary role in choosing a partner for their children, taking into account social attributes such as caste and family economic conditions (Banerjee *et al.*, 2013). According to Ray et al. (2020), among married couples recorded in IHDS 2011, 73% of married individuals answered their marriage was an arranged marriage by their parents, and 70% of women answered that they first met their marriage partner on the day of marriage. Interestingly, the percentage of arranged marriage is also high for inter-caste marriages at 63%.

In contrast to arranged marriages, in which parents and relatives play a role in the decision-making process, marriages in which the couple decides whom to marry are often referred to as "love marriage". In reality, however, arranged marriage do not mean that the will

of the couple is not respected or that romantic feelings do not exist; it means marriages in which emotional affection is not a prerequisite, and in which "gaining stability from its embedding in the established networks of social and kinship ties" (Polzenhagen and Frey, 2017) is a priority.

Individuals or families often have to search for marriage partners from a limited pool of people who belong to the same religion and same or similar caste, while income, educational background, family background, language, appearance, and personality required to them, as well as which caste or *jati* status are acceptable, vary from family to family. In order to mitigate the costs of such a complex search for a partner, various means of finding a partner have also developed. Information on people of marriageable age is compiled into profile / marriage CV / matrimonial bio, and advertised through relative networks, matchmakers in the caste community, marriage bureaus, newspaper advertisement, or matrimonial website. Newspaper advertisement is a traditional method with a history of more than 300 years (Polzenhagen and Frey, 2017), and matrimonial websites are a tool to carry on that trend.

With more than 20 million users and 150 different matrimonial websites (Trivedi (2014) in Ahuja & Ostermann (2016)), online matrimonial websites are the fastest growing medium of partner finding in the market. Registration is free on most of the websites, which is a similar feature to social networking services. The registrant creates a profile that includes details on his or her caste, educational background, mother language and occupation, lifestyle, and family background. This profile functions as an advertisement for those who are looking for partners on the same website. Users can browse the profiles of other registrants, and if they like a profile, they can request to be contacted. When searching for a profile, users can set criteria such as caste, income or other traits to browse profiles with desired attributes[3]. Currently, there are three major websites, Jeevansathi.com, Bharatmatrimony.com, and Shaadi.com[4]. Some of the smaller websites are targeted to specific regions, castes, or the high income group, which are also in high demand (Titzmann, 2018; Bhandari, 2020). The matrimonial website used for the experiment in this study will be described in detail in chapter 4.

Contrary to some expectations, parents or relatives continue to play an important role in decision-making in finding a marriage partner through online matrimonial websites. Newspaper and magazine articles often point out that such online matchmaking gives the marrying individual the privacy to escape the eyes of family and relatives and conduct a more individualistic matchmaking search (Bhandari, 2020). However, similar to newspaper advertisements, some of the profiles on matrimonial websites are posted by parents, relatives or friends, indicating that people other than the person themselves are also active in the online search for marriage partners[5]. As Bhandari (2020) points out, many of the items in the profile are about family, and often the free text section is also heavily weighted towards statements about family. Furthermore, Titzmann (2018) argues that caste endogamy have been made easier because information about caste, income, and family are clearly marked as items on those websites.

---

[3] At Jeevasathi.com, a paid account costs 38 or 44 USD per three months, and the business's annual revenue increased 17% to Rs 847 million in the fiscal year 2020. (http://www.infoedge.in/images/announcements/IEILAnnualReport201920.pdf)

[4] The numbers of users of each site are not officially announced. Nevertheless, Shaadi.com clams to have a total of 50 million registered users, BharatMatrimony.com claims to be used by 2 million users, and Jeevansathi.com clams to have been used by 20 million users, according to the website.

[5] About 20% of the profiles I identified in the experiment clearly indicated that the profile administrator was someone other than the individual.

Chapter 3
Literature Review

3.1 Language skills and economic welfare

If a language skill is a valuable skill in society and increases the productivity of workers, then having language skills improve one's income or social status. Although the benefits of language skills are often overlooked, there have been studies on the economic benefits of knowing the "right" language in the context of multilingual societies or regarding the economic performance of immigrants in their destinations (Ginsburgh and Weber, 2020).

The major concern in estimating the returns to language skills is the high endogeneity of language skills and income. Chiswick and Miller (1995) theorized language acquisition as a function of economic incentives for acquiring the language, exposure to the destination language, and efficiency in achieving destination languages. In their theoretical framework, they point out that the endogeneity of language and earning arises from the economic incentive: while economic incentives such as wage gaps or cost efficiency of consumption encourage language acquisition, the higher earnings as a result might also makes language acquisition easier. In fact, Chiswick and Miller found high endogeneity between language and earnings in Australia (Chiswick and Miller, 1995), as well as the United States, Canada(Chiswick and Miller, 2003), and Israel (Chiswick, 1998). Azam, Chin and Prakash's (2012) study also indicates similar pattern of high correlation between English language fluency and educational attainment and earnings in India. Furthermore, especially when analysing the relationship between language proficiency and marriage, Jan (2019) argues that the timing of the acquisition of language proficiency can be a source of bias. This is because the causal interpretation between language proficiency and marriage outcome is reversed depending on whether language acquisition occurs before or after marriage.

The literature has mainly employed two ways to address this issue: instrumental variable approach and natural experiment approach. In estimating return to education in a specific language, a policy change in the medium of instruction is exploited as a natural policy experiment. For example, Angrist & Lavy (1997), one of the earliest studies, find a reduction in returns to school education when the medium language of instruction was changed from French to Arabic in Morocco. Following this approach, Chakraborty & Kapur (2016) evaluated the effect of medium of instruction policy change in West Bengal, from English to Bengali. On the other hand, age at arrival of immigrants is commonly used as an IV in studies on immigrants(Bleakley and Chin, 2004; Aoki and Santiago, 2015, 2018). While many studies focus on the official languages of the country, some studies explore the returns to foreign language knowledge.(Lang and Siniver, 2006; Di Paolo and Tansel, 2015).

Since these studies were conducted on different languages in different regions, one cannot expect consistency in their results. However, there are some implications that are common in these findings, that are relevant for the current study. First, the effects of languages are often heterogenous. Levinsohn (2004) showed that the wage returns to English proficiency varied significantly by race in South Africa. The analysis showed that since South Africa's reintegration in international trade in 1993, returns to English increased from almost 0 to 14% among whites by the year 2000, but among blacks the returns were not significant. Secondly, the educational and economic returns to language may be contradictory. Nevertheless, the economic return to a language is in favour of the dominant languages (often foreign, European languages). In general, children can learn the most in their mother tongue (Benson, 2016; UNESCO, 2016), and this was also found to be true in the context of India

(Jain, 2017). However, several studies have shown that economic returns are higher when the language of instruction in schools is a language that is not the mother tongue of many children (English or French) (Angrist and Lavy, 1997; Kapur and Chalraborty, 2016). This suggests that education in one's mother tongue and education in a foreign language affect earnings in different ways. Ramachandran (2012) sees the choice of language of instruction as a trade-off between cognitive development and cultural transmission, and access to wages, job opportunities, and more information. However, when two languages are compared, many aspects other than wages are often overlooked in economic modelling. In this regard, the third implication is a limitation in the literature, that most studies have analysed the relationship between languages and wages and have not shed light on the social and cultural aspects of languages. Few exceptions are studies on the social assimilation of immigrants in host countries. Aoki & Santiago (2015, 2018) point to the relationship between language proficiency in the host country and the socioeconomic variables, such as health status, fertility, and geographic concentration in the ethnic enclaves. Also, Azam, Chin and Prakash (2012) and Munshi & Rosenzweig (2006)finds that English language skills in India play an important role in occupational choice. The use of experimental methods in this literature is rare, and to the author's knowledge, the only two studies are by Edo et al. (2019) and Leech et al. (2019). Their approaches are discussed further in section 3 in chapter 4.

3.2 Marriage market

The main objective of this study is to explore return to English language skills in a marriage market. In particular, this study also focuses on the probability of inter-caste marriage as a measure of social return to English language proficiency. Simultaneously, I estimate the effect of educational levels, income levels and caste on the matching outcomes. While there has been a substantial amount of theoretical and empirical research on the role of education and income in the marriage market, little research has focused on the role of language. In this chapter, I first review the theoretical prediction and empirical evidence on how education and income attributes of individuals affect marriage outcomes, including inter-caste marriage. Then, I discuss how language skills can affect the marriage outcome, using both theoretical frameworks and ethnographic evidence.

3.2.1 Education

If one's educational background influences the characteristics of a marriage partner, this may be understood as a non-economic form of return to education. In societies where women's participation is wage labour is low, the income of the man she marries will largely determine her own economic status after marriage (Banerjee *et al.*, 2013). In the context of the United States, Goldin (2006) suggests that almost half of the return to college education was through marriage for women who graduated between 1945 to 1960. Also, Kaufmann and Messner (2013) found that graduating from an elite university has a positive effect on the quality of one's marriage partner, and this effect is more than twice as large for women as for men in Chile.

However, return to education in the marriage market can also be negative, especially for women. There are several theoretical explanations for this hypothesis. Akerlof and Kranton (2000) argues that deviating from social norm (such as the norm that men should be the primary breadwinners in the family (Fortin, 2005)) can cause economic costs. This means that a highly educated women may suffer from fewer opportunities in the marriage partner. Besides, Furtado (2012) predicts that the effect of education on the marriage market can be positive or negative, depending on the average education level of the two groups of

men and women. Empirically, Fisman *et al.* (2006) conducted a speed dating experiment in New York City to find out what characteristics men and women valued more in each other. Their findings show that men tended not to value their partner's intelligence as much as women did, and that men did not value women's intelligence when it surpassed men's. The results of Neyt, Vandenbulcke and Baert's (2019) experiment in a dating application (Tinder) likewise showed that while women prefer highly educated men, this is not the case for men. However, their study suggests that men do not have a tendency to avert from highly educated women. Thus, although a large part of the literature find that the effect of education varies by gender, the findings about its consequence are mixed.

How can education be associated with the likelihood of marriage between people from different groups? Furtado (2012) argues that education promotes marriage between different communities by increasing people's adaptability to different cultures (cultural adaptability effect). In addition, status exchange theory is used to explain inter-group marriage, especially studies of caste. Social status that involves hierarchy, such as caste, can be viewed as status endowment. In this case, a person from a higher caste can marry a person from a lower caste with a higher educational background in exchange for his or her caste status (Dhar, 2013; Ahuja and Ostermann, 2016; Ray, Roy and Sahai, 2020). According to this theory, the marginal effect of increased educational attainment may be higher for those from low-caste backgrounds than for higher caste individuals (Ray, Roy and Sahai, 2020).

However, findings from empirical studies do not support such an exchange of educational background for caste status. For instance, a study by Banerjee *et al.* (2013) which analysed newspaper matrimonial advertisements and applications in West Bengal found such a strong own-caste preference by ad-placers that the educational background of the applicants hardly mattered. The result suggests that higher-caste women (and their family) choose men of the same caste with a high school diploma over those from other castes with a master's degree. Similar to this finding, Ray, Roy and Sahai (2020) find no significant association between the educational background of either the bride or the groom and the probability of marriage being inter-caste, from the data on actual incidence of marriages in IHDS 2011. Interestingly, in their study, the only education-related variable which was positively correlated with the probability of inter-caste marriage was the educational attainment of the groom's mother.

3.2.2 Income

For the same reasons as education, income can also positively or negatively affect an individual's marriage market outcomes and probability of inter-caste marriage. For income, contrary to education, a large body of evidence provides results consistent with the predictions of status exchange theory. Also, similar to the effect of education, the effect of income varies significantly between men and women, because men and women react differently to the income of the opposite sex.

Regarding the effect of income on the marriage market outcome in general, Ong and Wang (2015), from an experiment in a matching application in China, found that while income of female profiles did not affect the number of visits by male profiles, male profiles were visited by greater number of female profiles as their income increased. Women's profiles tended to visit profiles of men with higher incomes, and the number of visits increased discontinuously at the point where men's income exceeds women's income. Bertrand, Kamenica and Pan's (2015) analysis of actual couples in the US shows a similar discontinuous relationship between the income of the male and female.

In the context of India, existing literature suggests that a higher income of lower caste individuals leads to a higher possibility of mating with an individual from different castes (Dugar, Bhattacharya and Reiley, 2012; Banerjee *et al.*, 2013; Ahuja and Ostermann,

2016; Hortaçsu, Hwang and Mathur, 2019). Based on the results of a field experiment using newspaper advertisements in West Bengal, Dugar, Bhattacharya and Reiley (2012) claim that caste-based discrimination by female marriage market candidate fell as the income of low-caste men increased. Through data collection from matrimonial website in three other major cities, Ahuja and Ostermann (2016) found that among low-caste individuals, those with higher income were more interested in inter-caste marriage, while among high-caste people, on the contrary, those with lower incomes were more interested in interested in inter-caste marriage on matrimonial websites. This result strongly supports status exchange theory by highlighting the asymmetric interest in inter-caste marriage between lower caste and upper caste people. In the dataset of Banerjee *et al.* (2013), advertisement by higher income males attracted more letters from prospective brides and their family, and also, letters with male profile with higher income sent to female advertisement were more likely to be considered by the female side ad-placer.

Motivated by this relationship between income and inter-caste marriage, Hortaçsu, Hwang and Mathur (2019) analysed the impact of financial incentives on inter-caste marriage and found that the policy of granting a subsidy of Rs. 10000 increased inter-caste marriages by 4% in rural India.

3.2.3 Languages

This study hypothesizes that having English language proficiency will have a positive effect on marriage market outcomes, especially inter-caste marriage, for multiple reasons.

Linguistic Homogamy

First, people who speak English might be more likely to marry people who also have English ability. Speaking the same language makes communication smoother and enables sharing values and norms (Fu, 2007). Sharing the same language would be a prerequisite for adequate communication with one's spouse (and his/her family) in marriage. These factors increase the probability of marriage between people who share a common language (linguistic homogamy) (Jan, 2019). Conversely, when there is no common language between people who engage in transactions, there is the cost beard by the one side learning the other's language (Ginsburgh and Weber, 2020), and such transactions include marriage. This cost associated with learning languages will reduce the probability of marriage between people who do not share a common language. Empirically, Jan (2019) examined the relationship between ethnic language and English proficiency and marital outcomes among East Asians in the US. Their analysis found that ethnic language speakers tend to marry people of the same ethnicity who speak the same ethnic language, however, such tendency declines as English proficiency improves (holding the knowledge of ethnic language constant).

These arguments also apply to variations of the same language as well as non-verbal languages (Lang, 1986; Leech, Irby-Shasanmi and Mitchell, 2019). In South Asian contexts, it is known that people with more than one language ability frequently switch and mix vocabulary and sentences from their first language and English (Canagarajah, 2009), and the English used here depends on the person's English proficiency (Harish, 2014). Thus, having high English proficiency in the context of the Indian marriage market may imply not only that a person can speak English, but also that they speak a certain variation of a language among speakers of the same regional language (e.g. Hindi). As a consequence, in the situation of India where people with higher education and economic levels have higher English proficiency (Azam, Chin and Prakash, 2012), having English proficiency may increase the probability of marrying such people. And this could also be true even among people who share the same regional languages as their mother tongue.

As Means of Status Improvement

When the language spoken differs by a social class within a hierarchical structure, language becomes a powerful marker of social class. Sociologists argue that individuals and groups can improve their social status by changing their behavior or adopting new habits, especially by imitating higher social class groups. For example, Ignatiev (1995) argues that in the U.S. around 1850, Irish people were accepted by the white elite group, so-called WASPs, by adopting WASP's practice of oppression on African Americans. Employing the same value as the elite class, namely discrimination against blacks, boosted the status of the Irish. Regarding the caste system in modern India, Srinivas (1957) points out that low-caste (scheduled caste and scheduled tribe) people have attempted to improve their social status by adopting new names, customs, and practices.

Such adoption of the habits of those higher in the social hierarchy could include languages. Laitin 's (1998) seminal work on language and identity suggests that learning new languages is a means to improve one's social status in some context. His argument is based on field research and surveys in countries such as Latvia and Estonia, which became independent after the collapse of the Soviet Union. There, Russian-speaking residents, who suddenly lost their status as the dominant language speakers, established their status and identity by acquiring a new official language in the country. LaDousa (2007) was quick to remark that this improvement in status through language adoption was occurring in North India through the choice of the medium language of instruction in school education. In an anthropological study of students and others involved in Hindi medium, Urdu medium, and English medium schools in Banaras, he has described, already in 1996, being educated in English was a gatekeeper to a "new middle class". In India under such context, for those who belong to groups where few people speak English (e.g., low-income and low caste), acquiring English may be a way to improve their social status.

Human Capital Theory

English language skill can be seen as a type of human capital which is acquired largely through education (Laitin and Ramachandran, 2016; Ginsburgh and Weber, 2020). The studies on the economic return to language skills also see language skills as a part of human capital, as discussed in section 1 of chapter 3. In this understanding, language skills would have similar effect as education in the marriage market. In this theory, the effect of having English language skill on the marriage market performance can be both negative and positive depending on the social norm or the English language skill level of the opposite sex.

Signaling

While the theoretical reasonings above assume somewhat direct influence of being proficient in English, one concern raised from this assumption is the possibility of signaling effect. Contrary to human capital theory, signaling theory argues that even if education does not have the effect of boosting an individual's knowledge and skill level, the process of obtaining an education signals an individual's ability in the labor market and therefore educated people are paid higher wages (Weiss, 1995). The same may be said about the marriage market. In other words, even if there is no practical value with English language skills in the marriage market, it may still influence the marriage outcome by signaling other aspects of the candidates. For example, similar to the signaling theory of education, English proficiency can be a signal of a person's ability and earning potential. Moreover, in some context, English

language skills might signal certain family background, social and political values, and international mobility[6].

Qualitative Evidence

Beyond these theoretical rationales, there is a considerable amount of qualitative evidence which suggest that in the Indian context, speaking English can be an advantage in the marriage market and in society in general, or that not being able to speak English can be a disadvantage. In results of several interviews conducted by Bhandari (2020) show the influence of English proficiency on the selection of marriage partners. She interviewed men and women between the ages of 25 and 32 who had used matrimonial websites to find a partner. Her sample was drawn from the graduate community of reputedly excellent colleges and from employees of multinational corporations located in Delhi. In one interview, it is implied that English can be attractiveness.

> "Charu moved to Mumbai for two years for work experience, where she met the man she fell in love with—her work colleague. He was smart and educated, with a flair for English language, which impressed Charu."

In another interview, the respondent speaks about their potential groom's mother's English language skills as an indicator of the family background and turned down the match due to the difference. This is in line with the explanation that language can be a signal of family background. It also indicates similar importance of the parent's attribute in their children's marriage as in (Ray, Roy and Sahai, 2020).

> "My mother is a housewife, she is very well-educated and my father has a good job. We have lived in Delhi for almost 20 years and have a certain lifestyle. But he was brought up in a very different environment. […] Like, his mother doesn't speak English but my mother speaks fluent English. His father has a small business and their social circle is very different from ours. Also, they were very kanjoos [miserly]. In my house, my parents don't think twice about spending. […] See, it is not only about money. I am sure they are well-off, but they are the non-urban, Hindi speaking type, and we are the proper Delhi type […] I did not think these things would matter until I encountered this situation and realised that differences come up in little things."

In another case, a man judged a woman's lack of English language fluency as disqualification of her as his marriage partner. It also suggests that English fluency can signal one's cultural background.

> "I was shocked to know that she had never watched FRIENDS. How can this be? She grew up in Delhi not some village. How can she not know of the most popular series of our time. She simply did not know anything! I was really put off. Also, she did not speak very good

[6] The "Indian" marriage market also includes the Indian diaspora from all over the world who are looking for marriage partners of the same caste or ethnic background. One news article reported people studying IELTS to get the opportunity to marry to people from Canada, the UK or the US, and migrate to the destinations (https://www.dw.com/en/ielts-marriages-indias-ideal-bride-is-proficient-in-english/a-53341947). The profile of the matrimonial website used in current study, Jeevansathi.com, has a section to indicate whether the candidate is willing to migrate abroad.

> English. She was a nice girl, don't get me wrong, but I think she was meant to fit a traditional Marwari setting. I was not looking for that. I wanted someone who can relate to me, who has exposure…"

In another interview, a woman called Nitya appreciated that her *rishta* (potential candidate) prefers to speak in English rather their native language, which is same for her.

> "After a few weeks, they went to Assam, and this time the meeting was 'good', said Jayant. Jayant's mother spoke at length to Nitya and was pleased to know that she grew up in different parts of the world—Dubai, Abu Dhabi, and spent the past few years in Mumbai—and therefore was more cosmopolitan than Assamese in her cultural identity. She was glad to know that Nitya preferred to speak English and Hindi than Assamese."

Not only in the marriage market, the perceptions about knowing and not knowing English are widely discussed in other spheres of Indian society. As a consequence of the economic and political functions of the English language in a multilingual country, there is a symbolic value associated with the English language. Especially among those who are considered middle class in various urban areas of India, speaking and not speaking English is in many cases a sensitive issue. For a woman interviewed by Sandhu (2014), "English language was linked with perceptions of being intellectually superior in comparison with Hindi speakers." Also, Faust and Nagar (2001) who conducted research in Lucknow around the year 2000, when English medium education was expanding, describe how not being able to speak English is a strong complex even for a woman with a high level of education and a career with a master's degree. Jayadeva (2019), in their research on English medium education in low-fee private schools in Bangalore, points out that English is preferred as a medium language of instruction at the expense of educational attainment because English is seen as a symbol of the middle class, and it is believed that being able to speak English is transformative for a student's future. The sociologist Pierre Bourdieu considered language not only as a tool of communication but also as a reflection of power relations and positions among its user (Bourdieu, 1977). According to him, languages, dialects, and accents reflect the social status of the users, and he advocated that by speaking a particular language or accent, a speaker can reassert their power and position. In this line, all of this shows a clear awareness of the hierarchical relationship between English and vernacular languages including Hindi in people's perception.

Chapter 4
Research Methodology and Hypothesis

4.1 Experiment overview

This thesis implements a field experiment on a major matrimonial web service in India to to examine the effects of English language skills on the popularity of the profile of a bride's profile and the probability of matching and inter-caste matching. This section details setting of the matrimonial website where the experiment was conducted, the construction of fictitious profiles used in the experiment, the design of the randomized experiment, as well as the data collection procedure.

4.1.1 Matrimonial website

The matrimonial website used for the experiment offers various functions of These include, (i) registering profiles of a person who is looking for a marriage partner, (ii)searching for other profiles according to one's desired criteria[7], browsing through their profiles, and finding potential marriage partners, and (iii) contacting a candidate by sending an "interest", text message, e-mail, phone call. While there are no costs for registering on the website and using the basic functions, it generates revenue from registration fees for premium paid accounts. Premium accounts get priority customer service, unlimited access to other profiles' photos, and direct access to a certain number of accounts' contact information (phone number and email address). Paid users can also send direct messages to other profiles. On Jeevansathi.com[8] (the website used in this study) the primary means to make a contact with other profiles is sending "interests". A visitor to a profile can communicate their intention to contact him/her by clicking on the "Send Interest" button. A user who receives the interest decides whether to "accept" the interest after reviewing the sender's profile. If the interest is "accepted", both parties can talk via text message and exchange contact information. Depending on the communication, a face-to-face meeting will often be arranged by both families. Thus, when interest is accepted, the two parties enter a more time-taking communication process. Therefore, in this experiment, the acceptance of interest is exploited as a revealed preference that reflects the mate preference of each party. Further details of the website including registration process, all information in profiles, and key functions are presented in the Appendix I.

As already discussed, in India today, a majority of the middle class is engaged in "arranged marriages" in the broad sense (see Chapter 2 Section 3). In such a search for a marriage partner that meets the specific needs of both the individuals and the family, a use of matrimonial website including Jeevansathi.com is one of the most common mediums. The users of such websites are at least literate enough to use web pages in English and have the necessary smart phones or PCs to access the website. In a metropolitan area like Delhi, this is the case for the bulk of its population. In addition, many of the website users also use their relative and caste networks, other matrimonial web services, or marriage bureaus in

[7] For example, the search can be narrowed down to males or females who are between 25 and 30 years of age, 160 to 170 cm tall, Hindi, belong to Bhatt (one jati of Brahmin), native speaker of Hindi, reside in Mumbai, have annual income of at least Rs. 3 lakhs, are unmarried, have a bachelor's degree in economics, work in private sector and are non-vegetarian.

[8] *Jeevan Sathi* means life partner in Hindi.

combination at the same time (Titzmann, 2018; Bhandari, 2020). Therefore, it is unlikely that there is a clear and systematic difference in mate preference between the users of such websites and the rest of the population, except for some possible minority[9]. For these reasons, the users of marriage websites may be thought to be a representative sample of urban middle class households seeking marriage.

4.1.2 Experiment details

A. Creating fictitious variables

We constructed fictitious female profiles with all the combinations of variables whose effects we wanted to test. A sample fictitious profile can be seen in Appendix II. In the fictitious profiles, there are 4 items that are the subject of intervention: English proficiency, caste, income, and education level. Fictitious profiles were created using the following process. Other than these traits, all fictitious profiles had the same characteristics.

First, we set up a common set of characteristics for all profiles. These items were set to the characteristics most common to middle class women in Delhi (details are in Appendix II). Different names were given to all profiles, however the names were kept private to all other users on the website. Therefore, the only information regarding a profile's caste status (not specifying subcaste) was transmitted through the item of caste in the profile as discussed below. No profile pictures were included in fictitious profiles, which is common on the website especially for female profiles for privacy reasons[10].

Next, in addition to the common items, all combinations of the four attributes of English proficiency, caste status, income, and education were added to the profiles. The categories for each item are shown in Table 1. First, we assigned two types of castes to the fictitious profiles: upper caste (UC) and lower caste (LC). On the website, if a user selects Hindu in the religion section, one must select a caste (Brahmin, scheduled caste) and one can further specify the subcaste. We did not specify the subcaste to maximize the number of visitors to the profile when other users sort profiles by caste, and to avoid any possible effect by a status of specific subcaste. Second, we assigned three categories of income: lower middle income (annual income of Rs. 1-2 lakhs (2,300 euros)), middle income (Rs. 3-4 lakhs (4,600 euros)), and upper middle income (Rs. 5-7.5 lakhs (6,800 euros))[11]. In the income section of the profile, one can choose annual income from a list of income profiles as shown in Table 1. We simply selected each income bracket in the registration. Third, fictitious profiles were assigned with one of three categories of educational background: high school graduate,

[9] For example, some people might be strictly opposed to arranged marriages and profile-based match making, and those people would not use marriage websites. However, there is no reason to assume that in the middle and upper economic classes, such preferences are systematically associated with income, educational level, or interest in inter-caste marriage.

[10] Initially, in the experiment, male fictitious profiles were also created in the same manner. However, when the male fictitious profiles were posted, it did not receive a measurable amount of response in terms of both views and interests received. Therefore, the analysis only includes the results from female fictitious profiles.

[11] Middle class is a self-identified concept (Bhandari, 2020), and there is no clear definition of its income level. It is also difficult to determine the exact income distribution in India or Delhi due to lack of reliable data. However, for example, the per capita income in Delhi was Rs. 3.54 lakhs in fiscal year 2019-2020 (https://www.hindustantimes.com/cities/delhi-news/delhis-per-capita-income-goes-down-to-3-54-lakh-101615231355878.html), and Shukla (2010) defines middle class income as annual household income of Rs. 2 lakhs to 10 lakhs, while Krishnan and Hatekar (2017) use daily consumption of Rs. 148-742 (annual Rs. 0.54-2.7 lakhs). Thus, our target income group of below Rs 7.5 lakhs can be expected to cover the volume zone of Delhi's population.

Bachelor's degree, or Master's degree and above. Although the platform allows users to choose the level of degree as well as the subject of the degree (B. Ed, MSc in Engineering and so on), to avoid any effect from the subject of the degree, we indicate BA degree or MA degree.

Finally, signalling English language skills in the profiles explicitly is the most important part of the experiment. In fictitious profiles with English language skills, we included English in the section of "languages known" in the profile, in addition to a local language (Hindi). Further, we mentioned in the self-introduction of the profile that the prospective bride has outstanding written and spoken English language skills. In the profiles with no English skills, we do not mention English language skills in either the language section nor the self-introduction. Mother tongue was set as "Hindi-Delhi" in all the profiles, and self-introduction was written in English in all the profiles.

Table 1. Categories of key profile attributes

Attributes	Assigned categories	What appears in the profile
English proficiency	Yes	"Outstanding English language skills, both written and spoken"
	No	-
Caste status	UC	Brahmin
	LC	Scheduled Castes (SCs)
Income	Lower-middle income	Rs. 1-2 lakh
	Middle income	Rs. 3-4 lakh
	Upper-middle income	Rs. 5-7.5 lakh
Education	Less than bachelor's degree	High school
	Bachelor's degree	BA
	Postgraduate degree	MA

B. Data collection

The process of data collection was divided into two stages. In the first stage, fictitious profiles were posted on the matrimonial website, and we recorded the number of visitors to the profile within a certain time (72 hours). When a profile is created on the website, all profiles are browsable for all other existing users. Then, users can check how many other users have viewed the profile. We interpret this number as a measure of popularity of the profile in the online marriage market, following Ong and Wang (2015). They argue that the number of visits to a profile is a credible measurement of mate preference, as it does not require time costs or investment to other users. In our analysis, we simply compare the number of visits to the fictitious profiles with different attributes and determine the effect of each attribute on the number of visits.

The second part of the experiment involves interaction with existing profiles: "Interest" was sent from fictitious profiles to existing profiles and the acceptance rate was recorded (figure 1). At the beginning of this stage, we identified eight existing profiles who meet our criteria per each fictitious profile and recorded their information. Key information such as income, educational background, caste, and English language skills were coded in the same manner as in the case of the fictitious profiles that were created. Next, we randomly assigned the identified profiles to the fictitious profiles. Eight real profiles were assigned to one fictitious profile at a time. Consequently, the "interests" were sent from all the 48 fictitious profiles to the identified profiles assigned to each. Therefore, in total, 384 (48 times 8) profiles

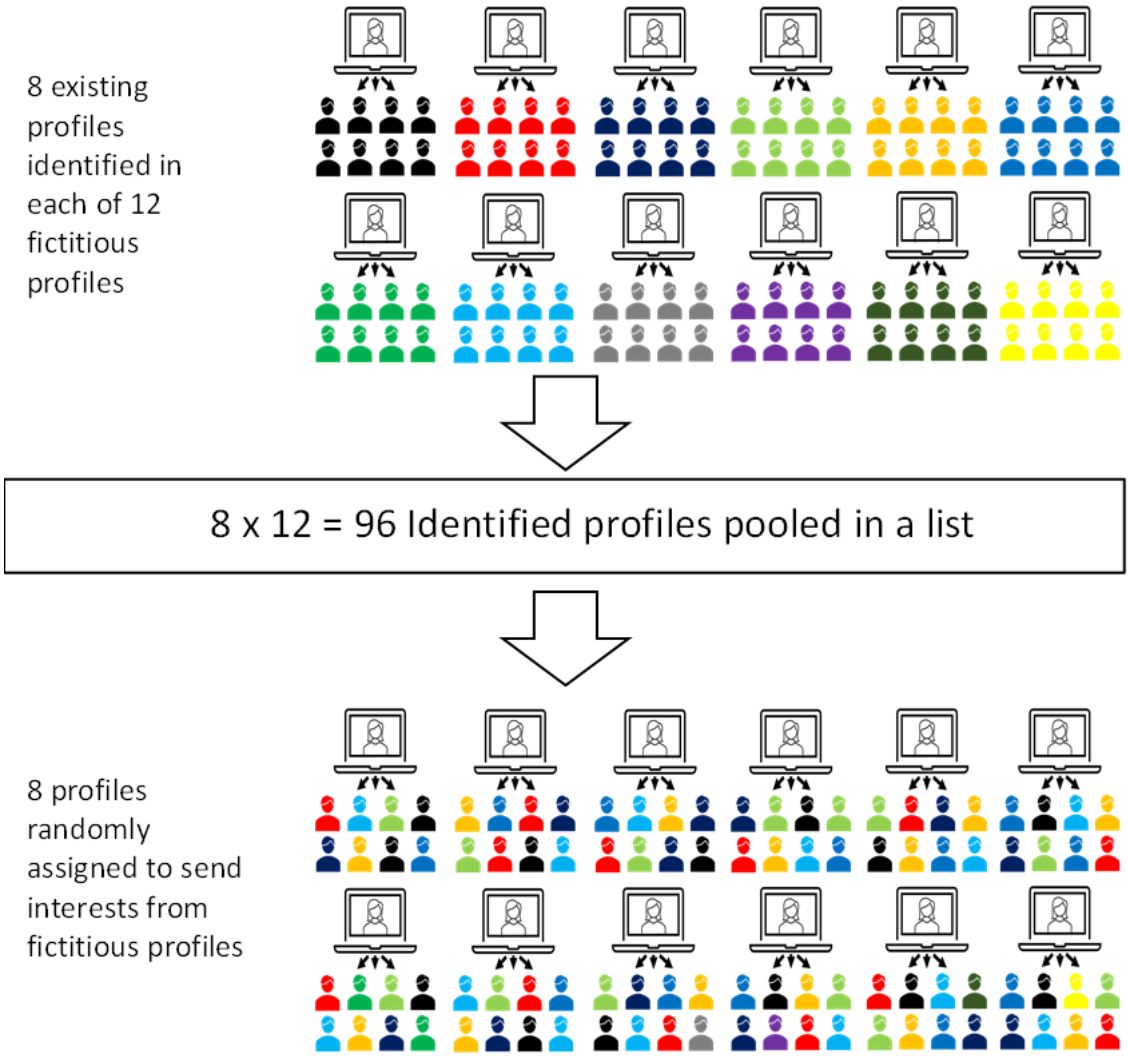
were identified and received interests. As an outcome variable, we recorded if the identified profiles accepted the "interest". If the interest was accepted, it may be interpreted that a match has been made between the fictitious profile and the identified profile.

In this process, male profiles that met our criteria were sampled from the website. This was done using the search function of the website. Reflecting our research focus, the criteria used for the search were: age 25 years and above, Hindu, living in Delhi, annual income below 7.5 lakh, Hindi as mother tongue, and unmarried men. We searched using the same criteria for each of the fictitious profiles and identified the top eight search results. If the same profiles appeared as matches for multiple fictitious profiles, the top eight excluding those profiles that had already been identified were used. The reasons for this process are discussed in greater detail in the next section.

C. Operation

Due to logistical constraints, we repeated identical waves of the experiment 4 times. To register on the matrimonial website, each profile requires a unique phone number. Therefore, we created 12 fictitious profiles with 12 phone numbers in one week. Each wave of the experiment lasted 5 days. In the first 72 hours, we recorded the number of visits to each profile. At the same time during this period, we identified 96 existing profiles that met our criteria and prepared a dataset with coded attributes of these profiles. After 72 hours, we sent "interests" from each fictitious profile to 8 profiles from amongst 96 identified profiles in each wave. Then, we waited for 48 hours to see if the interests were accepted. To avoid any possible biases, each wave began at the same day of a week. By repeating this process for 4 times (4 weeks) with different fictitious profiles, we covered all 48 fictitious profiles.

Figure 1. Process of identifying profiles, random assignment and sending interests



4.3 Why this methodology?

While the experimental approach employed in this study has the advantage of overcoming some critical limitations in existing literature on returns to language, its ethical concerns also need to be fully discussed. The first part of this section discusses the ethical aspect of this approach as well as measures taken to address some concerns, and the second part points out some strengths of the approach.

The experiment in this study is an application of so called correspondence studies that became prevalent after seminal work by Bertrand and Mullainathan (2004) which is increasingly used in the literature of labour market and discrimination (Banerjee *et al.*, 2009; Siddique, 2011; Bedi, Majilla and Rieger, 2018), as well as marriage market (Dugar, Bhattacharya and Reiley, 2012; Ong and Wang, 2015; Ahuja and Ostermann, 2016; Neyt, Vandenbulcke and Baert, 2019). Typically, in a correspondence survey, a set of fictitious CVs that are identical except for one or few attributes in the profile, are sent to job openings. Differences in between the outcome, often call back rates, across CVs is interpreted as influence of the attribute that differed in the fictitious CVs. In the literature on marriage market, profiles on newspaper advertisement, matrimonial website, or dating applications are used instead of CVs, and applications are sent to other profiles. The frequently used measure of

outcome is the acceptance of the application, which implies a match between two profiles to proceed to further steps, such as dating, family meeting or further conversation between the candidates.

One of the ethical concerns arising in this study is the absence of informed consent. As in other studies, in this experiment, existing users of the matrimonial website receiving the interest are involved in the study without being informed about and consenting to the study. Another concern is the privacy of personal information. The information posted on matrimonial website profiles is personal, including an individual's education, income, place of residence, origins, and photos. This information can be viewed by all other users of the site, nevertheless, it is not intended to be viewed by anyone other than the registered users of the site. Lastly, the possible effect of doing the experiment on the overall credibility of the platform (matrimonial website) should also be concerned. Frauds such as fake profiles or profiles containing false information are one of the main problems for the operators of such websites. To prevent such fraud and maintain the quality of the matching platform, the website Jeevansathi.com is taking various measures. For instance, any new user needs to register a unique phone number authenticated with a one-time password. Further, all newly registered profiles must be manually verified by the staff in Jeevansathi.com before it can be browsed by other users. Finally, in some cases, agents from Jeevansathi.com contact profile owners and visit them to verify the authenticity of their profiles in person[12].

I chose the experimental approach even in the light of these concerns for the following reasons. First, the resources required for the users of the website involved in the experiment is minimum, and it is not likely that the involved users are significantly affected by any influence from this experiment. The users of the website receive numerous contacts from other users constantly[13], hence the time and effort spent on checking interest sent from our fictitious profiles is likely small (accepting interest is a one-click process). Also, to minimize the impact on involved users, each user received interest from a fictitious profile only once in the experiment. Furthermore, even when the interest was accepted, no further communication took place, and the account was deleted within 48 hours. In the actual account between users as well, even if an interest is accepted, there is no guarantee that further communication will take place. Also, to minimize the impact on the website, only 12 profiles were registered per week, and the experiment was divided into four weeks. As a result, all fictitious profiles constructed by the research team were authenticated. Therefore, the process of the experiment is natural for existing users and has minimal influence on both users and the credibility of the website itself.

Throughout the experiment, the protection of personal information was emphasized. In all the existing profiles collected in the experiment, the owner's real name was kept private, so the collected data did not include their name. All information about existing profiles collected was kept in a secure location accessible only by the author and was not shared with anyone.

The experimental approach was employed in this study also due to its strategic advantages as well as practical considerations. To begin with, both subjective and objective language skills are difficult to observe, and for this reason, especially second language skills are rarely included as survey items in data. Furthermore, in the case of subjective language skills, there are concerns about the reliability of the variable due to the difficulty of ensuring their standardization. However, in the matrimonial website used in our approach, language

[12] Jeevansathi.com is accessed via a smartphone application by most of its users. Alternatively, Jeevansathi.com also have physical desks in 32 locations mainly in northern and western Indian states, where their clients can visit and register on the website with a consultation to an agent.

[13] For example, the fictitious profiles on average received 55 interests in 72 hours.

skills can be clearly stated as one of the items in the profile. Therefore, the effect of the language proficiency examined in this study is not of a specific proficiency of a language, but it is of indicating proficiency regardless of one's actual language skills.

The second and foremost empirical advantage of the approach is the elimination of endogeneity problem between one's language skills and other socioeconomic variables. As argued in section 3.1 (literature review on returns to language skills), including language skills as an explanatory variable in a regression is prone to endogeneity issues therefore the estimates may do not capture the causal relationship between the language skill and the outcome variable. As in other correspondence studies, this study deals with the endogeneity problem by preparing a fictitious profile and assigning traits randomly such that by construction there is no correlation between an individual's English language skills and other attributes. In this way, arguably, the approach enables the researcher to examine the unbiased causal influence of English language proficiency indicated in a profile on the matching outcome.

One correspondence study that exploits this advantage to examine the effect of language skills is known to the author. Edo, Jacquemet and Yannelis (2019) included French language skills in their fictitious resume in a correspondence study to examine discrimination against migrants in hiring in French labour market. The treatment used in their experiment to signal French language proficiency in the resume were mentioning experience in French tutoring, membership of a reading club, participation in scrabble and crossword competition, or experience as a writer of college newspaper. This rather indirect signalling of language skills intended to emphasize French language proficiency both for non-migrant French native candidates and migrant candidates who were also educated in France in French language. Their analysis found that indicating French language skills reduces discrimination for female migrants but not for males. In some other correspondence studies, language is used as a signal of other traits (such as race) (Leech, Irby-Shasanmi and Mitchell, 2019) or a control variable (Arceo-gomez and Campos-vazquez, 2014).

As a third advantage, by using correspondence study approach, there is very low risk of omitted variable bias in our analysis. This is so because in a correspondence survey, by constructing fictitious resume or profiles, all the information used for decision making by the employer (in our case the male users) is also observed by the researcher (Bertrand and Duflo, 2017). The analysis is also immune to some factors that can greatly affect mate preference and are difficult to control for, such as looks, skin tone, height, weight, or even the background of a profile photo.

Lastly, the current approach was chosen also due to a practical benefit that the approach does not require any physical movement of a research team or in-person contact for the data collection. Under a coronavirus pandemic, safe and risk-free data collection was a top priority as well as a methodological constraint. The data collection took place in August and September 2021, just after a large swath of India, including Delhi, was hit by one of the worlds' most devastating levels of corona virus cases. However, the approach enabled us to collect data online during the entire process without any face-to-face contact by any of research team and the involved people[14].

4.4 Hypothesis and empirical specification

Based on the data collected in the experiment, we conduct an analysis using two types of outcome variables in two approaches. The first is an analysis using a t-test with the number of visitors to the fictitious profiles as the outcome variable, and the second is regression

[14] Restriction on travel and human due to the restrictions on travel and human contact during coronavirus outbreak, the sales of Jeevansathi.com significantly increased in the 2019-2020 fiscal year.

analysis taking acceptance rate of interest sent from the fictitious profile to the identified profiles as the dependent variable. To analyse any possible heterogeneity of the effect, sub-group analysis is also conducted. The hypotheses tested in this analysis are organized as follows.

Hypothesis 1: Profiles of people with English skills are more popular than those without.

We measure the effect of English language skills on popularity in the marriage market by using the number of visitors to a fictitious profile and the number of applications a fictitious profile receives from existing real profiles. Our fictitious profiles include a group of profiles with English proficiency and a group of profiles without, but each with all combinations of educational background, income, and caste. Therefore, the two groups are exactly the same except for the presence or absence of English proficiency, and there is no statistical association between English and the other attributes (see Table 2). Hence the average difference in the popularity of the profiles of those two groups bay be interpreted as the effect of English language skills. We examine the difference between the two groups using a t-test.

Table 2. Summary statistics of fictitious profiles

	Without English	With English
Caste	0.5	0.5
Income		
No income	0.25	0.25
Lower middle income	0.25	0.25
Middle income	0.25	0.25
Upper middle income	0.25	0.25
Education		
High school	0.333	0.333
BA	0.333	0.333
MA	0.333	0.333
Observations	24	24

*The numbers represent the mean value of given variable in each group.

Hypothesis 2: English proficiency increases the probability of being considered as a marriage partner.

The aim of this hypothesis is to examine if indicating English language proficiency increases the probability of a response to an interest sent from the fictitious profile to a real profile or in other words does English language proficiency increase the probability of a match. The model simultaneously assesses the effect of caste status, educational background and income level mentioned in profile on the probability of acceptance. We examine this hypothesis by estimating the following model

$$Accept_{ij} = \alpha + \beta_1 English_i + \beta_2 Caste_i + \beta_3 Income_i + \beta_4 Education_i + \varepsilon_{ij} \quad (1)$$

where $Accept_{ij}$ is a dummy variable equals to 1 if existing profile j accepted "interest" sent from a fictitious profile i, $English_i$ is the main explanatory dummy variable which takes 1

when the fictitious profile i indicates English language fluency. $Caste_i$ is also a binary variable that takes 1 when the profile is of upper caste, and 0 for lower caste profiles. $Income_i$ is a categorical variable denoting income level mentioned in the profile in three categories (0=no income, 1=lower middle income, 2=middle income, and 3=upper middle income). Similarly, $Education_i$ is a categorical variable that represent educational background mentioned in the fictitious profile i (0=Highschool graduation, 1=B.A. degree, and 2=M.A. Degree).

In the constructed set of fictitious profiles, there is no systematic association between English proficiency and other variables, since the set of fictitious profile that indicate English language proficiency and profiles that does not mention English skills are identical except for English skills attributes. Also, the real profile to which the interest was sent was randomly assigned. Hence, the coefficient β_1 of the independent variable $English_i$ may be interpreted as an effect of indicating English language skills in a matrimonial profile on the acceptance rate of interests sent from the profile.

Hypothesis 3: Having English language proficiency increases the probability of inter-caste marriage.

Further, we hypothesize that having English language skills enables more matching with the potential spouses from a caste different from your own. This is the core hypothesis to answer the research question: Does having English proficiency increase social mobility in the marriage market? We examine this hypothesis by estimating the following model

$$Intercastematch_{ij} = \alpha + \beta_1 English_i + \beta_2 Caste_i + \beta_3 Income_i + \beta_4 Education_i + \varepsilon_{ij} \quad (3a)$$

where the variable $Intercastematch_{ij}$ denotes 1 if interest is sent to a profile of a different caste and it is accepted.

To examine the influence of each variable on matching in a way that increases social mobility, we use as an alternative dependent variable the acceptance rate when interest is sent from a LC profile to a UC profile ("marry-up" combination of male and female caste status). This can be examined by the model

$$Marryupmatch_{ij} = \alpha + \beta_1 English_i + \beta_2 Income_i + \beta_4 Education_i + \varepsilon_{ij} \quad (3b)$$

where $Marryupmatch_{ij}$ is a dummy variable takes 1 if an interest was sent from a LC profile i to a UC profile j and accepted.

Chapter 5
Data and Descriptive Statistics

5.1 Descriptive analysis of candidates

In the data collection, 384 existing male profiles were identified and received an interest from our fictitious profiles. Table 3 presents the descriptive statistics of identified profiles as well as the results of comparing LC profiles with UC profiles.

Table 3 presents summary statistics for the real profiles identified in the experiment. The mean age of the sample is 28 years old, which is roughly the average age of marriage for men in urban areas. The male in the sample is well educated, as the majority of the sample (54%) has bachelor's degree and about 28 % have a postgraduate degree. Male income is distributed across income groups, while the income group of Rs. 5-7.5 lakhs is the largest, accounting for 31% of the sample.

Some information regarding languages, career, and management of the profile that might be relevant to this study are also extracted from the profiles. Since Hindi language was set as a mother tongue in the search criteria, there are only 5% of profiles whose mother tongue is not Hindi. Those exceptions were the native speakers of Haryanvi or Punjabi language, that are close to Hindi language. Next, about 30% of the sample specified English as a language they spoke. However, filling in the "lifestyle" field, which includes this item of language spoken, is not mandatory and was left blank in many profiles, so the reliability of this information is limited. Nevertheless, it can be estimated that at least 30% of the sample speaks some form of English. Further, in 11% of the profiles, self-introduction field was filled out in Hindi (no profiles were filled out in languages other than Hindi and English)[15]. In addition, about 22% of the profiles specified that someone other than the candidate in the profile is administer of the profile.

The sample shows large inter-caste differences in some attributes, although LC profiles and UC profiles are samples under same criteria. Of all the profiles identified, 169 indicated LC status and 215 indicated UC status. The difference is significant for educational and income attributes. The percentage of people with less than a college degree is more than twice as high for LC profiles than in UC, and the percentage of postgraduate educated is 9% higher for UC profiles. Also, in terms of income, the percentage of people with lower income levels (Rs. 0-2 lakhs) is higher in LC profiles, while the percentage of higher income level (more than Rs. 4 lakhs) is significantly higher in UC profiles. This indicates that there is a clear association between caste status and other socioeconomic statuses in the male profiles.

[15] Many of the self-introduction written in Hindi were written with the Roman alphabet. And a few were written with Devanagari script.

Table 3. Summary Statistics of Identified Profiles

	All				LC	UC	
	Mean	SD	Min	Max	Mean	Mean	p-value
Age	28.758	2.469	25	40	28.811	28.716	0.711
Education	1.117	0.666	0	2	0.988	1.219	0.001***
lower than bachelor's degree	0.169	0.375	0	1	0.249	0.107	0.000***
Bachelor's degree	0.544	0.499	0	1	0.515	0.567	0.305
Master's degree and above	0.286	0.453	0	1	0.237	0.326	0.056*
Income							
Rs. 0-1 lakh	0.047	0.212	0	1	0.053	0.042	0.601
Rs. 1-2 lakh	0.089	0.284	0	1	0.142	0.047	0.001***
Rs. 2-3 lakh	0.154	0.361	0	1	0.160	0.149	0.769
Rs. 3-4 lakh	0.177	0.382	0	1	0.201	0.158	0.274
Rs. 4-5 lakh	0.169	0.375	0	1	0.112	0.214	0.008***
Rs. 5-7.5 lakh	0.313	0.464	0	1	0.254	0.358	0.030**
Languages							
Mother language other than Hindi	0.052	0.222	0	1	0.053	0.051	0.927
English	0.297	0.457	0	1	0.284	0.307	0.626
Profile written in Hindi	0.112	0.316	0	1	0.118	0.107	0.727
Government job	0.133	0.340	0	1	0.142	0.126	0.639
Profile managed by others	0.223	0.416	0	1	0.165	0.268	0.017**
Accept	0.302	0.460	0	1	0.302	0.302	0.991
Observations		384			169	215	

Notes: p-value shows statistical significance of t-test comparing mean between LC profiles and UC profiles.

5.2 Randomization check

In the data obtained from the experiment, the intervention group (existing profiles that received interest from a fictitious profile with English proficiency) and the control group (profiles received interest from a fictitious profile without English proficiency) are homogeneous. As Table 4 shows, there was no statistically significant difference in the key variables between the intervention group and the control group[16].

[16] Among the variables, only the difference in "Profile written in Hindi" is statistically significant at 1% level of significance between the intervention and control group. However, this is due to the small sample size, as only 11% of all profiles fall under this criterion.

Table 4. Randomization Check

	Without English (N=192)	With English (N=192)	p-value
Age	28.703	28.813	0.665
	(0.179)	(0.177)	
Education			
Lower than bachelor's degree	0.135	0.203	0.077
	(0.025)	(0.029)	
Bachelor's degree	0.552	0.536	0.759
	(0.036)	(0.036)	
Master's degree or above	0.313	0.260	0.260
	(0.034)	(0.032)	
Income			
No income	0.052	0.052	1.000
	(0.016)	(0.016)	
Rs. 0-1 Lakh	0.031	0.063	0.148
	(0.013)	(0.018)	
Rs. 1-2 Lakh	0.063	0.115	0.073*
	(0.018)	(0.023)	
Rs. 2-3 Lakh	0.125	0.182	0.120
	(0.024)	(0.028)	
Rs. 3-4 Lakh	0.188	0.167	0.594
	(0.028)	(0.027)	
Rs. 4-5 Lakh	0.172	0.167	0.892
	(0.027)	(0.027)	
Rs. 5-7.5 Lakh	0.370	0.255	0.015**
	(0.035)	(0.032)	
Languages			
Mother tongue not Hindi	0.036	0.068	0.169
	(0.014)	(0.018)	
English	0.333	0.260	0.118
	(0.034)	(0.032)	
Profile written in Hindi	0.156	0.068	0.006***
	(0.026)	(0.018)	
Government Sector Jobs	0.130	0.135	0.881
	(0.024)	0.025	
Profile managed by others	0.240	0.204	0.416
	(0.031)	(0.03)	

Notes: "Without English" is the group that received interest from a fictitious profile that did not specify English proficiency, and "With English" is the group that received interest from a fictitious profile that did specify English ability. p-value shows statistical significance of t-test comparing mean between two groups.

Chapter 6
Results

This chapter presents the results of analyses according to the order of the hypotheses presented in the previous chapter. The interpretation of the results is discussed in the Chapter 7.

6.1 Stage 1 Results

Hypothesis 1: Profiles of people with English skills are more popular than those without.

The first stage of the experiment recorded the number of views and the number of interests that each fictitious profiles received within 72 hours of registration. Overall, the fictitious profiles received 196 views and 55 interests on average, which means that 28.6% of the viewers of the profile sent interests (Table 5).

The fictitious profiles that specified English language proficiency received more views, interests, interests per view compared to the profiles that did not specify English language proficiency (Table 6). The results of the t-test show that the differences are statistically significant at significance level of 5% for the number of views, 1% for the number of interests received, and 10% for interests per view, respectively. On average, the profiles with English language proficiency attracted 36 (20%) more viewers, received 17 (38%) more interests from existing users. These differences are substantial compared to the differences between the outcome of upper caste profiles and lower caste profiles (UC profiles received 21(11%) more visitors and 22 (52%) more interests on average, see table 7).

Table 5. Stage 1 result (overall)

	Obs.	Mean	Std. dev.	Min	Max
Views	48	196.104	55.932	51	309
Interests received	48	55.396	21.785	21	115
Interests per view	48	0.286	0.080	0.181	0.608

Table 6. Stage 1 result by English proficiency

	Without English	With English	p-value
Views	178.125	214.083	0.024**
	(8.495)	(12.888)	
Interests received	46.458	64.333	0.003***
	(2.591)	(5.172)	
Interests per view	0.265	0.307	0.065*
	(0.011)	(0.019)	
Observations	24	24	

*p-value shows statistical significance of t-test.

Table 7. Stage 1 result by caste status

	LC	UC	p-value
Views	185.625	206.583	0.197
	(10.479)	(12.125)	
Interests received	44.083	66.708	0.000***
	(2.839)	(4.607)	
Interests per view	0.249	0.322	0.001***
	(0.018)	(0.01)	
Observations	24	24	

*p-value shows statistical significance of t-test.

While the first stage of the experiment used number of views and interests received by the fictitious profiles, the analysis lacks some important information, and there is a possible source of bias. As already discussed, the number of profile views and the number of interests received are valid measures of a profile's popularity. However, on the website, users cannot fully know from whom they received a visit. In other words, even if there is a difference in the popularity of a profile, it is impossible to know whom it is popular with.

Second, the algorithms by which websites recommend profiles for users to browse can be a source of bias. The basic algorithm is to show users the profiles that have the most matches to the items they are looking for in a potential marriage candidate[17]. For example, even if a user does not specify the caste of the marriage partner, the algorithm will automatically recommend profiles of the same or similar caste. In addition, while the most recently logged-in user's profile appears higher in the search results on the websites used in the experiment, there may be other reasons why a profile is more likely to appear in the search results[18]. In addition, the website has a function that displays the users who are currently logged in and match your preferences for a marriage partner. Therefore, the number of views and interests can be influenced by how many other users are logged in when the research team logged in to the fictitious profiles in the experiment.

To address these concerns in the experiment, all the fictitious profiles had the same desired characteristics for a marriage partner, and the research team logged in at the same time on the same day each week to register fictitious profiles. Nevertheless, it is impossible to know everything about these algorithms for the researcher. Therefore, we cannot exclude the possibility that factors other than the profile attributes may have influenced the outcome variables in the first stage of the experiment.

6.2 Stage 2 Results

As mentioned above, there are several shortcomings in the stage 1 outcome measures (number of views and interests received). In particular, the fact that the characteristics of the visitors are not known, is an important issue for the purpose of this study. The motivation for this study is to examine whether English proficiency helps upward mobility of social strata such as caste in marriage market. For this purpose, (i) knowing the information of the decision makers regarding marriage and (ii) including profile combinations other than existing profile preferences are essential for the analysis.

[17] If a user does not edit the characteristics, he or she is looking for in a marriage partner, this field will be automatically configured based on the user's profile.

[18] For example, it seems that the paid account profiles are more likely to appear higher in the search, and some of them are "featured" to be appear on the top in every search.

The second stage of the experiment addresses this need by identifying existing profiles, sending interests, and using their acceptance as the outcome variable. This method augments the first stage analysis in three ways. First, information about the decision maker can be captured with accuracy from the profile that receives the interest[19]. Second, by sending interest to randomly selected profiles from among the profiles once identified, the researcher can observe decisions regarding different combinations of caste and income levels[20]. Third, accepting an interest is a much clearer disclosure of mate preference than visiting a profile, which may proceed to further process towards marriage.

In the following, a descriptive analysis is first performed on the interest acceptance rate. Next, the results of the regression analysis will be presented for hypothesis 2, 3, and 4. Lastly, the heterogeneity of the effect is discussed based on a subgroup analysis.

6.2.1 Outcome differences according to English proficiency signals

Overall, 384 interests were sent to identified male existing profiles from 48 fictitious female profiles. Table 8 provides the acceptance rate of interest by English language proficiency (in columns) and other fictitious profile attributes (in row). Among all interests, 116 out of 318 (30.2%) of interests were accepted.

Most importantly, there is overall no difference in acceptance rate by English language proficiency. However, it makes a major difference for some subgroups. First, the largest effect is in the case of the lower caste profile. The acceptance rate for lower caste profiles is as low as 25%, however, the rate is more than 10% higher (30.2%) when the profile indicates English language skills compared with LC profiles that do not specify English proficiency (19.8%). Contrary, acceptance rate becomes lower when English proficiency was indicated for UC profiles. While the overall acceptance rate is higher for UC profiles than average (35.4%), the effect of UC status is almost negated by indicating English, hence the rate is higher without English proficiency (40.6%).

The effect of English is heterogenous for profiles with different levels of education. For profiles with BA indicated as her educational background, indicating English language proficiency increases acceptance rate by 6.2%. The acceptance rate is slightly higher for profiles with high school education and English proficiency than high school and without English. Contrary, for profiles with MA degree, the acceptance rate declines by nearly 8% when they mention English proficiency.

Similar heterogenous effect of English proficiency can be observed in different acceptance rate by income levels. The acceptance rate is overall lower for profiles with lower middle-income profiles. However, the acceptance rate is substantially higher by more than 10% when they indicate English proficiency (27%). On the other hand, the acceptance rate is lower when English skills are mentioned for middle-income and upper middle-income profiles. Overall, the result suggests that indicating English proficiency has a positive effect

[19]In many correspondence studies, it is difficult to construct a dataset regarding the decision making parties (such as organizations, or human resource cells) as it often relies on information in varying formats from job sites, company websites, etc (Duflo and Banerjee, 2017; Edo, Jacquemet and Yannelis, 2019).

[20] For example, if interests were sent according to the search results in each fictitious profile, most of the combinations of female and male profiles might be of the same caste, similar educational and income level. In our experiment, on the other hand, there is no association between the characteristics of the interest sender and the receiver. Although this situation is unlikely in practice, it is useful for determining the interest in inter-caste marriages. The experiment by Ahuja and Ostermann (2016) also employs this strategy to analyse the determinants of interest in inter-caste marriage.

on the acceptance rate for relatively lower status female profiles but has a negative effect for higher status profiles.

Table 9 shows the difference in the interest acceptance rate of inter-caste combinations according by English proficiency. In addition to the overall acceptance rate, there is no difference in the acceptance rate by English proficiency in the inter-caste combinations. However, only for the marry-up combination, the interests send from the fictitious profile with English proficiency were likely to be accepted. The acceptance rate of interest sent from LC women to UC men is very low, averaging 5.7%. Nevertheless, when English proficiency is indicated, the acceptance rate increased largely increased by 116 percentage points compared to the case where it is not indicated. This difference is statistically significant at 10% significance level.

Table 8. Acceptance rate by fictitious profile attributes with and without English language proficiency

		English proficiency		All	Obs.
		With English	Without English		
Caste	LC	**30.2%**	19.8%	25%	192
	UC	30.2%	**40.6%**	35.4%	192
Education	HS	**29.7%**	28.1%	28.9%	128
	BA	**37.5%**	31.3%	34.4%	128
	MA	23.4%	**31.3%**	27.3%	128
Income	No income	29.2%	29.2%	29.2%	96
	Lower middle	**27.1%**	16.7%	21.9%	96
	Middle	33.3%	**37.5%**	35.4%	96
	Upper middle	31.3%	**37.5%**	34.4%	96
All		30.2%	30.2%	30.2%	
Obs.		192	192		384

Note: Of the profiles with and without English proficiency, the one with the higher acceptance rate is indicated in bold letters.

Table 9. Acceptance rate and English language proficiency

	All	Without English	With English	p-value	Obs.
Accept	0.302	0.302	0.302	1.000	384
		(0.033)	(0.033)		
Acceptance of inter-caste interest	0.122	0.12	0.125	0.877	193
		(0.023)	(0.024)		
Acceptance of interest from LC to UC	0.057	0.036	0.078	0.079*	108
		(0.014)	(0.019)		
Obs.	384	192	192		

6.2.2 Main regression results

If the above difference in acceptance rate is due to the effect of indicating English proficiency in the profile, then the regression analysis even after controlling for the attributes of the profiles should also yield the same results as shown above. In this section, hypothesis 2 is tested again using a probit with the acceptance of interest as the dependent variable. Hypotheses 3 is then tested in the same way.

Hypothesis 2: English proficiency increases the probability of being considered as a marriage partner.

Table 10 presents estimates of several probit specifications. Specification (1) and (2) presents the results of a model where the probability of acceptance is taken as the dependent variable and attributes in male and female profiles are independent variables. The coefficient of English proficiency in female profile is not statistically significant in neither specification. Hence, consistent with the descriptive statistics in Table 8 and 9, the results here suggest that, on average, indicating English proficiency does not affect the probability of acceptance. Thus, the estimates do not support the second hypothesis.

Table 10. Profile attributes and probability of acceptance/inter-caste match

	(1) Accept	(2) Accept	(3) Inter-caste match	(4) Inter-caste match
Female mentions English skills	0.00712	0.0141	0.0423	0.0549
	(0.05)	(0.10)	(0.25)	(0.30)
Female caste	0.303*	0.339*	0.0648	0.0786
	(2.23)	(2.37)	(0.39)	(0.44)
Female income 1-2 Lakh	-0.231	-0.374	-0.417	-0.515*
	(-1.17)	(-1.79)	(-1.73)	(-1.99)
Female income 3-4 Lakh	0.179	0.0880	-0.292	-0.351
	(0.95)	(0.45)	(-1.26)	(-1.45)
Female income 5-7.5 Lakh	0.149	0.142	-0.139	-0.150
	(0.78)	(0.71)	(-0.63)	(-0.63)
Female BA	0.152	0.147	0.301	0.214
	(0.92)	(0.83)	(1.48)	(0.97)
Female MA	-0.0551	-0.0136	0.0149	-0.0499
	(-0.33)	(-0.08)	(0.07)	(-0.22)
Male mentions English skills		-0.00901		-0.158
		(-0.06)		(-0.74)
Male caste		0.0147		-0.160
		(0.10)		(-0.84)
Male income 0-1 Lakh		0.710		-0.223
		(1.36)		(-0.35)
Male income 1-2 Lakh		0.645		0.146
		(1.37)		(0.29)
Male income 2-3 Lakh		0.643		-0.310
		(1.44)		(-0.62)
Male income 3-4 Lakh		1.317**		0.659
		(3.02)		(1.47)
Male income 4-5 Lakh		0.841		0.328
		(1.89)		(0.70)
Male income 5-7.5 Lakh		0.671		0.149
		(1.55)		(0.33)
Male BA		0.267		0.173
		(1.13)		(0.58)
Male MA		0.117		-0.0709
		(0.43)		(-0.21)
Observations	384	384	384	384

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Hypothesis 3: Having English language proficiency increases the probability of inter-caste marriage.

Specification (3) and (4) in Table 10 shows the results of similar model, where the probability of interest sent to existing profile belong to different caste is accepted (inter-caste match). The coefficients of English skills are not significant here as well.

Overall, these results are consistent with the previous analysis of comparison of acceptance rates. In these regression analyses, profile attributes that have statistically significant effect on the probability of matching are female caste and male income. Acceptance rate of interests sent from upper caste profiles are 30 to 33 percentage point more likely to be accepted compared with those by lower caste profiles. About male side attributes, male profiles with income of 3 to 4 lakhs are 131 percentage point more likely to accept interest compared with male profiles with no income. However, these effects are not significant for the probability of inter-caste match.

6.3.3 Asymmetric effect by caste

The result of the analysis in 6.2.1 showed that the effect of indicating English proficiency is possibly different for different subgroup, especially by caste status. We now examine caste-specific effects of English proficiency, by using regression analysis on sub-sample of each caste. Table 11 present probit model estimates on the results of interests sent from upper caste profiles, and table 12 shows the same results for the lower-caste subsample.

The coefficients of English skills in UC subsample are negative, although statistically insignificant. Contrary to the hypothesis, this result suggests that indicating English proficiency can lower the probability of matching for UC female profiles (Table 11). The result is similar even after controlling for male-side profile attributes.

Table 11. Result of subgroup analysis (UC profiles)

	(1) Accept	(2) Accept	(3) Inter-caste match
Female mentions English skills	-0.290	-0.259	-0.415
	(-1.53)	(-1.20)	(-1.37)
Female income 1-2 Lakh	-0.294	-0.613*	-1.298**
	(-1.08)	(-2.02)	(-2.88)
Female income 3-4 Lakh	0.0586	-0.0752	-0.867*
	(0.22)	(-0.26)	(-2.21)
Female income 5-7.5 Lakh	0.230	0.149	-0.329
	(0.87)	(0.52)	(-0.93)
Female BA	0.476*	0.485	0.0919
	(2.06)	(1.86)	(0.25)
Female MA	0.139	0.130	-0.428
	(0.59)	(0.50)	(-1.12)
Male mentions English skills		0.0506	-0.354
		(0.21)	(-0.98)
Male caste		0.434*	
		(1.97)	
Male income 0-1 Lakh		0.776	
		(1.13)	
Male income 1-2 Lakh		0.542	-0.449
		(0.85)	(-0.62)
Male income 2-3 Lakh		0.744	-0.215
		(1.15)	(-0.29)
Male income 3-4 Lakh		1.586*	1.097
		(2.52)	(1.58)
Male income 4-5 Lakh		0.871	-0.244
		(1.37)	(-0.33)
Male income 5-7.5 Lakh		0.681	-0.161
		(1.11)	(-0.23)
Male BA		-0.0124	-0.310
		(-0.04)	(-0.69)
Male MA		-0.106	-0.399
		(-0.28)	(-0.77)
Observations	192	192	180

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

For LC profiles, the effect of indicating English skills turns out to be positive. Especially in specification (3) in Table 12, on the probability of inter-caste match for LC profiles, the effect of English skills is positive and statistically significant at 10% significance level. This result is about the cases where the interest sent from the LC female profile to the UC male profile is accepted, i.e., a combination that is a "marry-up match" for the female fictitious profile. The result suggests that indicating English proficiency in the profile increases the probability of matching by 57.7 percentage point for such combination of male and female caste status. This result is consistent with the results of the analysis of descriptive statistics presented in Table 8. Therefore, the results of the analysis partially confirm Hypothesis 3: Having English language proficiency increases the probability of inter-caste marriage.

Table 12. Result of subgroup analysis (LC profiles)

	(1) Accept	(2) Accept	(3) Inter-caste match
Female mentions English skills	0.334	0.382	0.577*
	(1.67)	(1.73)	(2.02)
Female income 1-2 Lakh	-0.147	-0.180	-0.178
	(-0.50)	(-0.57)	(-0.46)
Female income 3-4 Lakh	0.320	0.300	0.118
	(1.16)	(1.01)	(0.32)
Female income 5-7.5 Lakh	0.0757	0.135	0.0109
	(0.27)	(0.44)	(0.03)
Female BA	-0.198	-0.124	0.0510
	(-0.82)	(-0.46)	(0.15)
Female MA	-0.248	-0.169	-0.0585
	(-1.02)	(-0.65)	(-0.17)
Male mentions English skills		-0.0821	-0.353
		(-0.32)	(-1.00)
Male caste		-0.435	
		(-1.95)	
Male income 0-1 Lakh		0.604	0.719
		(0.70)	(0.78)
Male income 1-2 Lakh		0.735	0.518
		(1.06)	(0.67)
Male income 2-3 Lakh		0.326	-0.870
		(0.53)	(-1.13)
Male income 3-4 Lakh		1.010	0.174
		(1.72)	(0.27)
Male income 4-5 Lakh		0.856	0.540
		(1.42)	(0.85)
Male income 5-7.5 Lakh		0.608	0.225
		(1.02)	(0.35)
Male BA		0.598	0.890
		(1.57)	(1.53)
Male MA		0.336	0.373
		(0.78)	(0.58)
Observations	192	192	192

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Chapter 7
Discussion

This chapter discuss the results of the study. The first section interprets the results in comparison with the previous studies. The second section assembles implications from the findings. The third section discuss the limitations of the current study and possible improvement, and the fourth section suggests some tasks for the future research.

7.1 Interpretation of the results

First, indicating English proficiency in female profiles on a matrimonial website increased the number of visitors to the profile and the number of interests received significantly. Profiles with English proficiency received 36 (20%) more visitors and 17 (38%) more interests than profiles without an English signal. The lower caste profiles received fewer visitors and less interest than the upper caste profile, however, the difference was reduced when English proficiency was specified in LC profiles.

Next, we sent interest from fictitious profiles to a randomly assigned existing profile and examined their acceptance rate. The results showed that there was no difference in the acceptance rate of interest between the profiles with and without English proficiency. However, only when the interest was sent from the LC profiles to a profile of a higher caste (caste hypergamy), English ability was found to increase the acceptance rate. The interests for a LC profile that mentioned English proficiency was 57.7 percentage points higher compared to a profile that did not mention such proficiency. On the other hand, for UC profiles, indicating English proficiency was detrimental.

The results of stage 2 of the experiment are not necessarily inconsistent with the results of stage 1. In stage 1, the number of profile views and the number of interests received increased when English proficiency was mentioned, while it is impossible to know by whom these responses were made. Therefore, it is possible that the popularity if the fictitious profiles with English proficiency increased due to the response of some male profiles that particularly evaluated English skills positively. On the other hand, in stage 2, after controlling for information from male profiles, the description of English proficiency worked positively only for LC women.

The most important result of the analysis is that indicating English proficiency in LC women's profiles increases the acceptance rate by UC men, compensating for the disadvantage caused by LC status. In other words, indicating English proficiency increases the probability of upward caste marriage. There are a number of possible reasons why the effect of English proficiency varies by caste status in the profile. The first interpretation is, as discussed in one of the theories, that English is used as a method of status improvement. The prerequisite for such an interpretation is that being high caste implies a higher socioeconomic background, and at the same time English is perceived as the language of such a relatively higher social class. At the same time, the low-caste status of women is a sign to men that the socioeconomic status of women is also relatively low. In such a situation, for low-caste people, emphasizing their proficiency in English, the language of "high social class and high caste" at the same time, becomes a way to elevate their status by adopting parts of their language, lifestyle, and culture. These assumptions are beyond the scope of the empirical finding of this study. However, they are consistent with the qualitative research on English perceptions already presented.

The second interpretation is that English proficiency was valued as human capital in the marriage market, however, the returns were obscure where the skill was not scarce. The premise of this interpretation is that English proficiency is more common among UC people who have relatively better socioeconomic backgrounds, while it is relatively scarce among LC women who are associated with lower socioeconomic status. In this case, this interpretation could explain why the returns to English proficiency in the marriage market are positive only for LC women, but not for UC women. In short, if it is more common for UC women to speak English, speaking English is not particularly advantageous for them. This interpretation is consistent with some findings on the return to English skills in the labour market. A study in India by Azam, Chin and Prakash (2012) found lower returns to English proficiency among those in groups with higher proportions of English speakers (younger generations or in some regions where English has historically been more prevalent) compared to other groups. Also, in a study by Casale and Posel (2011), the returns to English proficiency for white population, where a higher proportion of people have English proficiency, are lower than the return for the black population. Furthermore, the fact that the return to human capital in the form of English proficiency was higher for LC women than for UC women is consistent with the prediction of social exchange theory that I introduced in Section 2 of Chapter 3. Applying this theory, the return on English language skills may have been higher for LC women, because they could match with UC men by compensating for their low caste status in exchange for their high language skills. The study by Ray, Roy and Sahai (2020) found that the probability of inter-caste marriage was unrelated to the education level of the couple, hence they did not find differences in returns to education by caste as predicted by social exchange theory. The current result may contrast with their findings in that the human capital of the marrying individual may affect the likelihood of inter-caste marriage.

The third interpretation is that both English proficiency and caste status are strong signals of a candidate's family background in the marriage market. Information about the cultural and economic background of the family, especially about the parents, is an important factor influencing the selection of a marriage partner (Ray, Roy and Sahai, 2020), even in an online matrimonial market (Bhandari, 2020). However, in the experiment of this study, the setting of family background in the fictitious profiles was kept to a minimum. This was to prevent such factors from influencing the result regarding English skills. However, since the male side prefers women with a "good" family background, and if UC status is a sufficient signal of having a "good" family background, then for LC women, English proficiency may compensate for the lack of high caste status providing a signal of having a "good" family background. Therefore, it is possible that English proficiency only worked in favour of lower-caste women but not for high caste women.

It is more difficult to explain why indicating English proficiency had a negative association with the acceptance rate for some people. The results indicate that stating English proficiency may lower acceptance rate for UC women. However, it is unlikely that those who are more likely to speak English avoid women who are proficient in English. Therefore, indicating English proficiency in the experiment could have signalled differently for some people. For example, for people for whom English skills is more common (such as UC people), emphasizing English skills in the profile might have conversely suggested the person's lack of English skills.

7.2 Implications of the study

The implications of the findings of this study for the society and policy are as follows. First, listing English proficiency in the profile increased the number of visitors to the profile and

the number of interests received. This means that the profile owner has more options for marriage partners, which is a social return to language skills that may itself increase their welfare. This implies that acquiring English proficiency can lead to more options for marriage partners.

In the study, we also found that indicating English proficiency enables caste-hypergamy for lower caste women. This suggest that acquiring English language can increase upward social mobility for lower caste women in terms of caste, therefore, it can weaken caste disparity. This effect is a causal effect, meaning it holds positive after controlling for educational and income attributes of the profile. Therefore, the findings imply that learning English language to get a better marriage partner is a rational behaviour. From a policy perspective, encouraging low-caste women to learn English can diminish caste barriers and close the caste gap.

On the other hand, the results of this experiment also suggest that English may act as a signal for other factors such as caste and family background. Complementarities between caste, social class and English language proficiency need to be reduced as much as possible to prevent social division between caste and economic classes on the basis of language. To this end, it would be effective to ensure that children with socioeconomic background have sufficient opportunities to acquire English through education. On the contrary, the spread of English language skills only among high-status population will lead to increased inequality, as Chakraborty and Kapur (2016) also point out. The results of this study suggest that such a relationship between English and inequality is closely linked not only to economic inequality but also to social inequality through marriage.

Finally, the fact that English proficiency leads to social benefits also means that not knowing English can be a disadvantage. As long as Hindi is a primary official language at the national level and public education is provided in Hindi or other regional languages, users of those languages should not be discriminated for not speaking English. Hence the government should ensure that the language rights of speakers of other languages are guaranteed in the provision of public services, public education, recruitment of civil servants, and the media.

7.3 Limitations and possible improvements

This section presents limitations of this study, the improvements that should be made in future studies, and the remaining research issues. The first limitation is on the validity of the English language proficiency signal. The signal used in the experiment was mentioning "outstanding English language proficiency, both written and spoken" in the self-introduction in the profiles, which does not refer to a specific level of language proficiency. Therefore, as discussed in the previous section, the meaning received from the signal may differ from person to person. Furthermore, in the marriage market, the linguistic ability of prospective candidates is a factor that is judged not only from their profiles, but also in the subsequent process of arranging meetings between themselves or their families. Hence, the ability to converse in English is likely to have a greater impact on the selection of marriage partners. However, stating English language proficiency in one's profile is also explicit enough to proxy one's ability to converse in English. In addition, it was not possible to have variation in the level of English ability of the candidates in this experiment. Thus, it was impossible to determine what level of English proficiency would be advantageous in the marriage market, while the effect can vary depending on the level of English.

To overcome this challenge, in future studies, English ability can be signalled by audio information. Specifically, an experiment such as speed dating (as in Fisman *et al.* (2006)), in which several candidates talk to each other face-to-face and make decision about each other as potential marriage partners, could be considered. However, face-to-face

conversation can lead to the transmission of information from the candidate's appearance and demeanour in addition to language ability, which may have an endogenous relationship with language skills. Therefore, in order to rigorously estimate the returns to linguistic ability, experimental methods such as signalling with recorded data or speed dating with speech alone are preferred. As an example of such means, in the experiment by Leech, Irby-Shasanmi and Mitchell (2019), the racial category of people who inquire about a hospital is signalled by speaking English with and without a black accent over the phone.

The second limitation is that the external validity of this study may be limited by geographic area and income strata. The population covered in this study is the Hindi speaking population living in the Delhi NCR area with an annual income of up to Rs. 7.5 lakh. Hence, the result of this study may be different from rural areas as well as for other regions due to different attitudes towards inter-caste marriage and different perceptions towards English language. Also, this study only includes Hindi-speaking Hindus, and does not include inter-marriage between different religions or matching people with different mother tongues (often synonymous to regional background). However, in the marriage market of such combinations, English will have a greater influence as the common language of the couple. Analysis on other regions and more extensive research, including the possibility of inter-regional and international marriages, is a subject for future study.

The third limiting factor is that the results of the experiment did not allow us to analyse the possible effects of English proficiency on male candidates. This is because, the experiment did not yield a measurable response to the male fictitious profiles from the existing female profiles. It is clear from previous studies that much of the return on men's English proficiency would have accrued as a labour wage. However, whether there are gender differences in social returns to language, such as in the marriage market, needs to be tested in future research.

Finally, it goes without saying that the returns to language in the marriage market are only one aspect of social return to the knowledge of a language. Laitin and Ramachandran, (2016) point out that language proficiency can be a gatekeeper not only to human capital, but also to access to healthcare knowledge, political participation, and even self-esteem. In the context of India, as suggested by the qualitative evidence reviewed in this paper, English would be a requirement for access to cultural capital and knowledge resource beyond higher education. The social returns to other various languages, and the presence or absence of language fluency as a discriminatory factor, should attract more academic and practical attention to help redress inequalities among speakers of different languages.

Chapter 8
Conclusion

Languages itself is an important aspect of human capital, a primary means of accumulating human capital, and also an inherent part of our identity. Knowledge of a dominant language, such as English, not only leads to getting a white-collar job and higher wages, but it also gives people access to far more knowledge and culture, and allows them to meet different people. On the other hand, if the acquisition of such a language is restricted to the economically advantaged, as in the case in many countries, then the growing importance of that language will lead to a widening gap between those who know it and those who do not. In India, English proficiency is now a powerful signal that points to a person's educational background, cultural and economic background of the family, and even social class.

Despite this situation, non-economic returns to language proficiency have rarely been studied as a development issue in low- and middle-income countries, specially returns to women's language skills. To supplement this gap in the literature, this study examined the returns to English language proficiency in an online marriage market in Delhi, India. To overcome the problem of endogeneity and other socioeconomic variables and to signal language proficiency uniformly, the study used experimental methods collected data through a major matrimonial website.

The results show that in the marriage market, English proficiency provides women with a wider choice of men to marry. Profiles with English proficiency attracted 20% more visitors and 38% more contacts from real male profiles than profiles without English fluency. Furthermore, when interests were sent from the constructed fictitious (female) profiles to randomly assigned real (male) profiles, the study found that stating English proficiency substantially and statistically significantly increased the acceptance rate for low-caste women sending interests to high-caste men by 57.7 percentage point. This size of the effect is large, given the very low rate of actual inter-caste marriages and the fact than this combination is between Brahmin caste men, which is the highest caste, and scheduled caste, which is the lowest caste category. This result suggests that knowledge of English language has the effect of increasing the probability of matching in the marriage market, compensating for the disadvantage of a low caste status. On the other hand, for high-caste women, indicating English proficiency in women's profiles had negative effects. This result suggests that strong linkages are assumed by participants in the marriage market between caste, economic level, education level, and English proficiency.

Word Count: 17,273

References

Ahuja, A. and Ostermann, S. L. (2016) 'Crossing Caste Boundaries in the Modern Indian Marriage Market', *Studies in Comparative International Development*, 51(3), pp. 365–387. doi: 10.1007/s12116-015-9178-2.

Akerlof, G. A. and Kranton, R. E. (2000) 'Economics and Identity', *The Quarterly Journal of Economics*, 115(3), pp. 715–753.

Aldashev, A. and Danzer, A. M. (2014) *Economic Returns to Speaking the Right Language(s)? Evidence from Kazakhstan's Shift in State Language and Language of Instruction*.

Allendorf, K. and Pandian, R. K. (2016) 'The Decline of Arranged Marriage? Marital Change and Continuity in India', *Population and Development Review*, 42(3), pp. 435–464. doi: 10.1111/j.1728-4457.2016.00149.x.

Angrist, J., Chin, A. and Godoy, R. (2006) 'Is Spanish-Only Schooling Responsible for the Puerto Rican Language Gap?', *Working Paper 12005, National Bureau of Economic Research*, 33(January).

Angrist, J. D. and Lavy, V. (1997) 'The effect of a change in language of instruction on the returns to schooling in Morocco', *Journal of Labor Economics*, 15(1), pp. 48–76. doi: 10.1086/209856.

Aoki, Y. and Santiago, L. (2015) *Education , Health and Fertility of UK Immigrants : The Role of English Language Skills*. 9498. Bonn.

Aoki, Y. and Santiago, L. (2018) *Deprivation, Segregation, and Socioeconomic Class of UK Immigrants: Does English Proficiency Matter?*

Arceo-gomez, B. E. O. and Campos-vazquez, R. M. (2014) 'Race and Marriage in the Labor Market : A Discrimination Correspondence Study in a Developing Country', *American Economic Review: Papers & Proceedings*, 104(5), pp. 376–380.

Asher, S., Novosad, P. and Rafkin, C. (2021) *Intergenerational Mobility in India: New Methods and Estimates Across Time, Space, and Communities*.

Azam, M., Chin, A. and Prakash, N. (2012) 'The Returns to English-Language Skills in India', *Economic Development and Cultural Change*, 61(2), pp. 335–367. doi: 10.1086/668277.

Banerjee, A. *et al.* (2009) 'Labor market discrimination in Delhi: Evidence from a field experiment', *Journal of Comparative Economics*, 37(1), pp. 14–27. doi: 10.1016/j.jce.2008.09.002.

Banerjee, A. *et al.* (2013) 'Marry for what? Caste and mate selection in modern india', *American Economic Journal: Microeconomics*, 5(2), pp. 33–72. doi: 10.1257/mic.5.2.33.

Banerjee, B. and Knight, J. B. (1985) 'Caste Discrimination in the Indian Urban Labour Market', *Journal of Development Economics*, 17, pp. 277–307.

Bedi, A., Majilla, T. and Rieger, M. (2018) *Gender Norms and the Motherhood Penalty: Experimental Evidence from India*. Available at: www.iza.org (Accessed: 15 May 2020).

Benson, C. (2016) *Addressing language of instruction issues in education: recommendations for documenting progress; Background paper prepared for the 2016 Global education monitoring report, Education for people and planet: creating sustainable futures for all; 2016*.

Bertrand, M., Kamenica, E. and Pan, J. (2015) 'Gender Identity and Relative Income Within Households', *The Quarterly Journal of Economics*, 130(2), pp. 571–614. doi: 10.1093/qje/qjv001.Advance.

Bertrand, M. and Mullainathan, S. (2004) 'Are Emily and Greg More Employable than Lakisha and Jamal ? A Field Experiment on Labor Market Discrimination', *American Economic Review*, 94(4), pp. 991–1013.

Bhandari, P. (2020) *Matchmaking in middle class India: Beyond arranged and love marriage*, *Matchmaking in Middle Class India: Beyond Arranged and Love Marriage*. doi: 10.1007/978-981-15-1599-6.

Bidner, C. and Eswaran, M. (2015) 'A gender-based theory of the origin of the caste system of India', *Journal of Development Economics*, 114, pp. 142–158. doi: 10.1016/j.jdeveco.2014.12.006.

Bleakley, H. and Chin, A. (2004) 'Language Skills and Earnings: Evidence from Childhood Immigrants', *The Review of Economics and Statistics*, 86(May), pp. 481–496.

Bourdieu, P. (1977) 'The economics of linguistic exchanges', *Social science information*, 16(6), pp. 645–668.

Canagarajah, S. (2009) 'The plurilingual tradition and the English language in South Asia', *AILA Review*, 22(January 2009), pp. 5–22. doi: 10.1075/aila.22.02can.

Casale, D. and Posel, D. (2011) 'English language proficiency and earnings in a developing country: The case of South Africa', *Journal of Socio-Economics*, 40(4), pp. 385–393. doi: 10.1016/j.socec.2011.04.009.

Chaudhary, S. (2017) *Hindi Medium*. India: AA Films, Zee Studio.

Chiswick, B. R. (1998) 'Hebrew language usage: Determinants and effects on earnings among immigrants in Israel', *Journal of Population Economics*, 11(2), pp. 253–271. doi: 10.1007/s001480050068.

Chiswick, B. R. and Miller, P. W. (1995) 'The Endogeneity between Language and Earnings: International Analyses', *Source: Journal of Labor Economics*, 13(2), pp. 246–288.

Chiswick, B. R. and Miller, P. W. (2003) 'The complementarity of language and other human capital: Immigrant earnings in Canada', *Economics of Education Review*, 22(5), pp. 469–480. doi: 10.1016/S0272-7757(03)00037-2.

Chudgar, A. and Creed, B. (2016) 'How are private school enrolment patterns changing across Indian districts with a growth in private school availability?', *Oxford Review of Education*, 42(5), pp. 543–560. doi: 10.1080/03054985.2016.1217688.

Dallin, A. (2000) 'Review Reviewed Work (s): Identity in Formation : The Russian-Speaking Populations in the Near Abroad', *The Journal of Modern History*, 72(2), pp. 583–585. Available at: https://www.jstor.org/stable/10.1086/316042.

Dhar, R. L. (2013) 'Intercaste Marriage : A Study from the Indian Context', *Marriage & Family Review*, 49(1), pp. 1–25. doi: 10.1080/01494929.2012.714720.

Duflo, E. and Banerjee, A. V. (2017) *Handbook of Economic Field Experiments*. 1st edn. Edited by E. Duflo and A. V. Banerjee. North Holland: Elsevier. Available at: https://www.elsevier.com/books/handbook-of-field-experiments/duflo/978-0-444-63324-8.

Dugar, S., Bhattacharya, H. and Reiley, D. (2012) 'Can't buy me love? A field experiment exploring the trade-off between income and caste-status in an indian matrimonial market', *Economic Inquiry*, 50(2), pp. 534–550. doi: 10.1111/j.1465-7295.2011.00398.x.

Edgerton, N. (2016) 'Wage Returns to English Skills : A Comparison of Local and Multinational Banks in Thailand', *Southeast Asian Journal of Economics*, 4(June), pp. 113–140.

Edo, A., Jacquemet, N. and Yannelis, C. (2019) 'Language skills and homophilous hiring discrimination: Evidence from gender and racially differentiated applications', *Review of Economics of the Household*, 17(1), pp. 349–376. doi: 10.1007/s11150-017-9391-z.

Faust, D. and Nagar, R. (2001) 'Politics of Development in Postcolonial India: English Medium Education and Social Fracturing', *Economic & Political Weekly*, 36(30), pp. 2878–2883.

Fisman, R. *et al.* (2006) 'Gender differences in mate selection: Evidence from a speed dating experiment', *Quarterly Journal of Economics*, 121(2), pp. 673–697. doi: 10.1162/qjec.2006.121.2.673.

Fortin, N. M. (2005) 'Gender role attitudes and the labour-market outcomes of women across OECD countries', *Oxford Review of Economic Policy*, 21(3), pp. 416–438. doi: 10.1093/oxrep/gri024.

Fu, X. (2007) 'Inter-racial marriage and family socio-economic status: A study among Whites, Filipinos, Japanese, and Hawaiians in Hawaii', *Journal of Comparative Family Studies*, 38(4). doi: 10.3138/jcfs.38.4.533.

Furtado, D. (2012) 'Human capital and interethnic marriage decisions', *Economic Inquiry*, 50(1), pp. 82–93. doi: 10.1111/j.1465-7295.2010.00345.x.

Ginsburgh, V. and Weber, S. (2020) 'The Economics of Language', *Journal of Economic Literature*, 58(2), pp. 348–404. doi: 10.1257/jel.20191316.

Goldin, C. (2006) 'The Quiet Revolution That Transformed Women's Employment , Education , and Family', *American Economic Review*, 96(2), pp. 1–21. doi: 10.1257/000282806777212350.

Harish, S. (2014) 'Social strategy use and language learning contexts: A case study of Malayalee undergraduate students in India', *System*, 43(1), pp. 64–73. doi: 10.1016/j.system.2014.01.002.

Hnatkovska, V., Lahiri, A. and Paul, S. (2012) 'Castes and Labor Mobility', *American Economic Journal: Applied Economics*, 4(2), pp. 274–307.

Hortaçsu, A., Hwang, S. I. M. and Mathur, D. (2019) 'Monetary incentives on inter-caste marriages in India: Theory and evidence', *Journal of Development Economics*, 141(October 2018), p. 102371. doi: 10.1016/j.jdeveco.2019.102371.

Ignatiev, N. (1995) *How the Irish became White*. New York: Routledge.

Ito, T. (2009) 'Caste discrimination and transaction costs in the labor market : Evidence from rural North India', *Journal of Development Economics*, 88(2), pp. 292–300. doi: 10.1016/j.jdeveco.2008.06.002.

Jain, T. (2017) 'Common Tongue: The Impact of Language on Educational Outcomes', *Journal of Economic History*, 77(2), pp. 473–510. doi: 10.1017/S0022050717000481.

Jain, T., Maitra, P. and Mani, S. (2019) 'Barriers to skill acquisition : Evidence from English training in India', *World Development*, 114, pp. 314–325. doi: 10.1016/j.worlddev.2018.10.011.

Jan, J. S. (2019) 'Ethnic Language Retention and East Asian Endogamy in the United States', *Marriage and Family Review*, 55(1), pp. 59–77. doi: 10.1080/01494929.2018.1458003.

Jayadeva, S. (2019) 'English-medium: Schooling, social mobility, and inequality in Bangalore, India', *Anthropology & Education Quarterly*, 50, pp. 151–169. doi: 10.1111/aeq.12287.

Joshi, R. (2020) 'Can social integration in schools be mandated: Evidence from the Right to Education Act in India', *International Journal of Educational Development*, 77(December 2019), p. 102228. doi: 10.1016/j.ijedudev.2020.102228.

Kapur, S. and Chalraborty, T. (2016) 'English language premium: Evidence from a policy experiment in India', *Economics of Education Review*, 50, pp. 1–16. doi: 10.1016/j.econedurev.2015.10.004.

Kaufmann, K. M. and Messner, M. (2013) 'Returns to Elite Higher Education in the Marriage Market : Evidence from Chile'.

Kijima, Y. (2006) 'Caste and Tribe Inequality : Evidence from India , 1983 – 1999', *Economic Development and Cultural Change*, (2001), pp. 1983–1999.

Krishnan, S. and Hatekar, N. (2017) 'Rise of the New Middle Class in India and', *Economic and Political Weekly*, 52(22), pp. 40–48.

LaDousa, C. (2007) 'Of Nation and State Language, school, and the reproduction of disparity in a North Indian City', *Anthropological Quarterly, 80(4), 925-959.*, 80(4), pp. 925–959.

Laitin, D. D. (1998) *Identity in formation: The Russian-speaking populations in the near abroad*. Cornell University Press.

Laitin, D. D. and Ramachandran, R. (2016) 'Language Policy and Human Development', *American Political Science Review*, 110(3), pp. 457–480. doi: 10.1017/s0003055416000265.

Lang, K. (1986) 'A Language Theory of Discrimination', *The Quarterly Journal of Economics*, 101(2), pp. 363–382.

Lang, K. and Siniver, E. (2006) *The Return to English in a Non-English Speaking Country: Russian Immigrants and Native Israelis in Israel*. 12464. Cambridge.

Leech, T. G. J., Irby-Shasanmi, A. and Mitchell, A. L. (2019) '"Are you accepting new patients?" A pilot field experiment on telephone-based gatekeeping and Black patients' access to pediatric care', *Health Services Research*, 54(January), pp. 234–242. doi: 10.1111/1475-6773.13089.

Levinsohn, J. (2004) *Globalization and the Returns to Speaking English in South Africa*. 10985. Cambridge, MA. Available at: http://eprints.uanl.mx/5481/1/1020149995.PDF.

Madheswaran, S. and Attewell, P. (2007) 'Caste Discrimination in the Indian Urban Labour Market : Evidence from the National Sample Survey', *Economic and Political Weekly*, 42(41), pp. 4146–4153.

Madheswaran, S. and Singhari, S. (2017) 'Social exclusion and caste discrimination in public and private sectors in India : A decomposition analysis', *The Indian Journal of Labour Economics*, 59(2), pp. 175–201. doi: 10.1007/s41027-017-0053-8.

Munshi, K. (2019) 'Caste and the Indian Economy', *Journal of Economic Literature*, 57(4), pp. 781–834. Available at: https://doi.org/10.1257/jel.20171307%0ACaste.

Munshi, K. and Rosenzweig, M. (2006) 'Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy', *American Economic Review*, 96(4), pp. 1225–1252. doi: 10.1257/aer.96.4.1225.

Munshi, K. and Rosenzweig, M. (2009) *Why is Mobility in India So Low? Social Insurance, inequality, and Growth*, *NBER Working Paper Series*. 14850. Cambridge, MA. Available at: http://www.nber.org/papers/w14850.

Munshi, K. and Rosenzweig, M. (2015) *Insiders and Outsiders: Local Ethnic Politics and Public Goods Provision*. 21720. Cambridge, MA. Available at: http://www.nber.org/papers/w21720.

Nair, S. K. (2015) *Does Medium of Instruction Affect Learning Outcomes?-Evidence Using Young Lives Longitudinal Data of Andhra Pradesh, India*.

Neyt, B., Vandenbulcke, S. and Baert, S. (2019) 'Are men intimidated by highly educated women? Undercover on Tinder', *Economics of Education Review*, 73(June), p. 101914. doi: 10.1016/j.econedurev.2019.101914.

Ong, D. and Wang, J. (2015) 'Income attraction: An online dating field experiment', *Journal of Economic Behavior and Organization*, 111, pp. 13–22. doi: 10.1016/j.jebo.2014.12.011.

Di Paolo, A. and Tansel, A. (2015) 'Returns to Foreign Language Skills in a Developing Country: The Case of Turkey', *Journal of Development Studies*, 51(4). doi: 10.1080/00220388.2015.1019482.

Parinduri, R. A. and Ong, K. (2018) *The effects of mediums of instruction on educational-and labor market outcomes: Evidence from Malaysia*.

Polzenhagen, F. and Frey, S. (2017) 'Are Marriages Made in Heaven? A Cultural-Linguistic Case Study on Indian-English Matrimonials', in Sharifian, F. (ed.) *Advances in Cultural Linguistics*. Springer, pp. 573–605.

Ramachandran, R. (2012) 'Medium of Instruction Policies and Human Capital Formation: Theory and Evidence from Africa', *SSRN Electronic Journal*. doi: 10.2139/ssrn.2001331.

Rao, G. (2019) 'Familiarity Does Not Breed Contempt: Generosity, Discrimination, and Diversity in Delhi Schools', *American Economic Review*, 109(3), pp. 774–809. doi: 10.1257/aer.20180044.

Ray, T., Roy, A. and Sahai, K. (2020) 'Whose education matters? An analysis of inter caste marriages in India', *Journal of Economic Behavior and Organization*, 176, pp. 619–633. doi: 10.1016/j.jebo.2020.02.011.

Rosenzweig, M. R. and Stark, O. (1989) 'Consumption Smoothing, Migration, and Marriage: Evidence from Rural India', *Journal of Political Economy*, 97(4), pp. 905–926. Available at: https://www.jstor.org/stable/1832196.

Sandhu, P. (2014) 'The interactional and narrative construction of normative and resistant discourses about hindi and english', *Applied Linguistics*, 35(1), pp. 29–47. doi: 10.1093/applin/ams075.

Shukla, R. (2010) *How India earns, spends and saves: unmasking the real India*. SAGE Publications India.

Siddique, Z. (2011) 'Evidence on Caste Based Discrimination ☆', *Labour Economics*, 18, pp. S146–S159. doi: 10.1016/j.labeco.2011.07.002.

Srinivas, M. N. (1957) 'Caste in Modern India', *The Journal of Asian Studies*, 16(4), pp. 529–548.

Titzmann, F. (2018) 'Hindu religious identification in India's online matrimonial market', in Zeiler, X. (ed.) *Digital Hinduism*. 1st Editio. Routledge, pp. 35–50. Available at: https://www.taylorfrancis.com/chapters/edit/10.4324/9781315107523-3/hindu-religious-identification-india-online-matrimonial-market-fritzi-marie-titzmann.

UNESCO (2016) *If you don't understand, how can you learn?* Paris.

Yuki, K. (2021) 'Language education and economic outcomes in a bilingual society', *SSRN Electronic Journal*, (February). doi: 10.2139/ssrn.3846985.

Appendix

Appendix I. Details of Matrimonial Website

a. Registration Process

Registration processes include the process of identification / authentication and entering personal information. Most of the personal information required at this stage is mandatory for all profiles. For major items with choices, the choices are summarized in Table A.

① Sign Up

New registrants to the website must first enter and select the following information.

- Email address
- Mobile phone number
- Password
- Create profile for: (Choose from) Self, Son, Daughter, Brother, Sister, Other. When other is selected, further specify from relative/Friend or client (for marriage bureaus).
- Gender (Male or Female)

② Profile Details

In the next page, users are required to enter key characteristics of the candidate.

- Groom/Bride's Name
- Whether or not to show the name to other accounts
- Date of Birth
- Mother tongue
- Religion
- Caste
- Whether or not to be open to marry for all castes ("Caste no bar")
- Manglik status
- Whether if you prefer horoscope match
- Marital Status
- Height

③ Career details

- Country of residence
- State of residence
- City of residence
- Pin code of the address
- Highest degree
- Sector of employment
- Occupation
- Annual income
- Self-introduction
  It is free-text and must be at least 50 words.

④ Authentication process

The registrant enters a one-time password of four-digit number sent by SMS to the

mobile phone number they have entered to confirm their identity.

⑤ Lifestyle & Family (this section can be skipped)
- Family type (Joint family, nuclear family, others)
- Father's occupation
- Mother's occupation
- Number of brothers and how many of them are married
- Number of sisters and how many of them are married
- Place of residence of the family
- Native City
- Contact address
- Free description of the family

b. Further Information in the Profiles

Once a user entered the above information, the profile is complete and awaiting verification by the administrator. Verification usually takes place within a few hours. Most of the information except for the date of birth, caste, and gender can be edited afterwards. Further, few more details such as followings can be added to profile after registration.

⑥ Education & Career
- Name of school, undergraduate and post graduate college
- Name of the employer
- Interest in settling abroad (Yes, No, Undecided)
-

⑦ Family details
- Gothra
- Family value (Orthodox, Conservative, Moderate, Liberal)
-

⑧ Habits
- Dietary Habits (vegetarian, non-vegetarian, Jain, Eggetarian)
- Drinking Habits (Yes, No, Occasionally)
- Smoking Habits (Yes, No, Occasionally)
- Open to Pets? (Yes, No)
- Own a House? (Yes, No)
- Own a Car? (Yes, No)
- Languages I speak (Indian languages + Some European languages such as English, French, German, Portuguese, as well as Asian languages including Arabic, Pushto, Persian, Japanese, and Mandarin)
- Blood Group
- HIV+ (Yes, No)
- Thalassemia (Major, Minor, No)
- Challenged (Physically-from birth, Physically Due to accident, Mentally-from birth, Mentally -due to accident, further choose nature of handicap for physical handicaps)

⑨ Your likes
- Hobbies
- Interests
- Favorite Music

- Favorite Book
- Favorite Read
- Dress Style
- TV Shows
- Favorite Movies
- Movies
- Sports
- Cuisine
- Food I cook
- Vacation Destination

Table A. Items with choices

Category	Available choices	
Age	A number between 18 to 70 years old.	
Height	Choice of 4 feet (122 cm) to 7 feet (213 cm) or more in inches	
Religion and Caste	Hindu, Muslim, Sikh, Christian, Buddhist, Jain, Parsi, Jewish, Bahai, Other	Choose from more than 500 subcastes (Group of subcastes: Aggarwal, Arya Vysya, Bania, Banik, Brahmin, Brahmin Andhra, Brahmin Bengali, Brahmin Dravida, Brahmin Gaur, Brahmin Gujarati, Brahmin Kannada, Brahmin Kanyakubj, Brahmin Karnataka, Brahmin Maharastra, Brahmin Rajasthani, Brahmin Saraswat, Brahmin Utkal, Chettiar, Kayastha, Khatri, Kapu, Kshatriya, Maratham Marwari, Mudaliar, Nair, Patel, Patil, Rajput, Scheduled Caste, Scheduled Tribe, Sindhi, Teli, Vaishnav, Vellalar)
	Muslim	Select "sect" from Shia or Sunni, Select Caste from 47 categories
	Sikh	Select one from 19 caste categories
	Christian	Select one from 30 "sect" categories
	Buddhist	(No choice of subgroup available)
	Jain	Select from 3 caste categories
	Parsi	(No choice of subgroup available)
	Jewish	(No choice of subgroup available)
	Bahai	(No choice of subgroup available)
	Other	(No choice of subgroup available)
Mother Tongue	Hindi, Hindi-Delhi, Hindi-MP/CG, Hindi-UP/UK, Punjabi, Hindhi-Bihar/Jharkhand, Hindi-Rajasthan, Haryanvi, Himachali, Kashmiri, Sindhi, Urdu, Marathi, Gujarati, Kutchi, Konkani, Tamil, Telugu, Kannada, Malayalam, Tulu, Bengali, Oriya, Assamese, Sikkim/Nepali, English	
Country	Select one from all countries	
City / State	(When India is selected for country)	
Income	No Income, Rs. 0-1 Lakh, Rs. 1-2 Lakh, Rs. 2-3 Lakh, Rs. 3-4 Lakh, Rs. 4-5 Lakh, Rs. 5-7.5 Lakh, Rs. 7.5^10 Lakh, Rs. 10-15 Lakh, Rs. 15-20 Lakh, Rs. 25-35 Lakh, Rs. 35-50 Lakh, Rs. 50-70 Lakh, Rs. 70 Lakh-1 Crore, Above Rs. 1 Crore	
Marital Status	Never Married, Awaiting Divorce, Divorced, Widowed, Annulled	

Have Children	"No", "Yes, living together", "Yes, living separately"	
Astro	Manglik Status	Non Manglik, Manglik, Angshik (partial manglik)
	Horoscope	Date of Birth, Time of births

Appendix II. Example of fictitious profiles

All fictitious profiles contained the same information in all categories except English proficiency, caste, educational background, and income.

In all the profiles, gender was set to female, profile manager was the individual, mother tongue was Delhi Hindi, religion was Hindu, sub-caste was left blank, marriage status was set to Never Married, height was set to 5 feet, which is the average height of Indian women, and age was set to 23 years, which is the average age of marriage for women in urban India. It was set to Her place of residence and origin is Delhi, and she is working in the private sector, particularly in banking. Her family is nuclear and lives in Delhi, her father is in the private sector, her mother is a housewife, and she has one married brother. Her preferred marriage partner is a Hindu, but she does not specify her caste (caste no bar). Her lifestyle includes occasional drinking, and he does not smoke. His date of birth and zip code were randomly assigned from the list. For the self-introductions, several variations of the standard content were created, and one of these was randomly selected to be included in each profile. An example of a fictitious profile registered on a website is as shown below.

Jeevansathi Be found HOME MATCHES INBOX SEARCH UPGRADE HELP 99+

Change Cover Photo

Upload Photos From

Computer Facebook

You can set Photo Privacy

Preview Album

Add details to your profile

Upload Photos: +26%

Write About You & Family: +15%

Add Family Details: +7%

Add Horoscope Details: +7%

38%

Profile completion

Last Edited on 29th Oct, 2021

Profile Views 0

Critical Fields - Can be edited only once in lifetime Edit

Age
25 (2nd Apr 1996)

Marital Status
Never Married

Basic Details Edit

Full Name - Geeta Paswan

Your profile verification is pending... Get Verified NOW

Height
5' 0"

Religion
Hindu

Mother Tongue
Hindi-Delhi

Annual Income
Rs. 3 - 4 Lakh

Location
New Delhi

Caste, Sect
Hindu: Scheduled Caste , Not filled in

Profile managed by
Her profile is managed by Self

About me Under Screening Edit

As a baking professional, I love my job and love to deal with financial issues and to study/learn new things. I am an easy-going, flexible, and energetic person, also a cricket enthusiast. I like to travel to new places or watch movies on my days off.

The most important thing I look for in a partner is to respect each other's freedom and live together helping each other and our families to be happy.

Outstanding English language skills, both written and spoken.

Describe yourself in 5 words

Not filled in

Contact Details Edit

Email id Under Screening
[redacted] Verify

Alternate Email id
Not filled in

Mobile No.
[redacted] Verified

Alternate No.
Not filled in

Landline No.
Not filled in

Suitable time to call
Not filled in

Contact Address
Not filled in

Parent's Address
Not filled in

Horoscope Details Edit

Create Horoscope

Place of Birth
Not filled in

About My Family Under Screening

My family is a small nuclear family well settled in Delhi.
I have one brother who is already married

Education

Not filled in

Occupation

Not filled in

Apr 02, 1996

Time of Birth
Not filled in

Horoscope match is not necessary

Sun sign
Not filled in

Rashi/Moon Sign
Not filled in

Nakshatra
Not filled in

Manglik
Non-Manglik

Horoscope Privacy
Not filled in

Education & Career

Edit

Highest Education
B.A

School Name
Not filled in

UG Degree
B.A

PG Degree
Not Applicable

UG College
Not filled in

PG College
Not Applicable

Other UG Degree
Not filled in

Other PG Degree
Not Applicable

Employed In
Private Sector

Occupation
Banking Professional

Organization Name
Not filled in

Annual Income
Rs. 3 - 4 Lakh

Interested in settling abroad?

Family Details

Edit

Mother's Occupation
Housewife

Father's Occupation
Service - Private

Sister(s)
0 sister

Brother(s)
1 brother of which 1 married

Gothra
Not filled in

Gothra (maternal)
Not filled in

Family Status
Not filled in

Family Income
Not filled in

Family Type
Nuclear Family

Family Values
Not filled in

Family based out of
New Delhi, Delhi

Living with parents?

Lifestyle

Edit

Habits
Dietary Habits?, Drinks occasionally, Doesn't smoke

Assets
House - No, Car - No

Languages Known
English, Hindi

Blood Group
Not filled in

Special Cases
Challenged - None, , Thalassemia - No, HIV+ - No

(Source: profile created by the author on the website)