Preferring Future Male Leaders? Evidence from Junior Civil Servants' Hiring and Promotion

Alex Eble and Feng Hu*

PRELIMINARY AND INCOMPLETE
NOT FOR CITATION OR CIRCULATION

January 2025

Abstract

This paper studies how gender preference shapes decisions in junior hiring and early-career promotions. We combine novel data on job advertisements, candidate evaluations, and the promotion history of junior civil servants in China. We find stated male preference in civil service job advertisements at a level far greater even than found contemporaneously in the private sector. This discourages women from applying for these jobs. During the hiring process, female candidates receive lower evaluations, but only in the part of the evaluation where gender identity is disclosed. Consequently, similarly qualified women are less likely than men to be hired as civil servants. Despite outperforming their male counterparts, women in civil service positions are 35-45% less likely to be promoted. These findings illustrate how gender preference influences career trajectories from job advertisements to promotion decisions, helping explain the persistence of gender inequality in leadership roles. We conclude by presenting quasi-experimental evidence that two targeted policies – designed to address gender preference in advertisements and candidate assessment – effectively reduce gender disparities in both candidate evaluations and hiring outcomes.

*Eble: Columbia University and National Bureau of Economic Research (NBER). Email: eble@tc.columbia.edu, Hu: University of Science and Technology Beijing. Email: feng3hu@gmail.com. Eble and Hu are co-first authors and contributed equally; authors are listed alphabetically by surname. We are grateful for helpful input from the audience at 2024 Camphor Economic Circle Annual Conference, 2024 CCER Summer Institute, 2024 China Labor Economists Forum Annual Conference, and seminars at Fudan University and the Weatherhead East Asian Institute at Columbia University. Keywords: gender, leadership, inequality, civil service, promotion, labor market. JEL codes: J16, L14, J62, J71 I23, O30

1 Introduction

Women remain underrepresented among political, scientific, and business leaders around the world despite significant progress toward gender equality in education and labor market outcomes (Wolfers, 2006; Goldin, 2014; Bertrand and Duflo, 2017; Armstrong et al., 2023). Much of the research and media attention has focused on the barriers women face in reaching senior positions (Matsa and Miller, 2011; Zimmerman, 2019; Corbett et al., 2022). However, an often-overlooked aspect of this issue is the early career barriers for women, both in securing junior positions and in pursuing early promotions, which contribute to persistent inequality later in their careers.

In this paper, we examine how gender preference affects women's entry into and progression within the public sector in China. We combine a series of novel, large-scale administrative datasets on job advertisements, job applicants, applicant evaluations, and the promotion history of junior civil servants. We find stated male preference in job advertisements, which discourages women from applying for these jobs. Female job candidates are further penalized in their evaluation, but only in the part of the candidate evaluation process where their gender identity is known to the assessor. This bias leads to a pro-male skew in final hiring decisions. Among those hired, we find that female civil servants demonstrate better performance on a wide range of metrics officially linked to promotion decisions, but are nonetheless 35-45% less likely than men to be promoted to higher-rank positions.

The Chinese context has several important features which facilitate our study. First, China has a large number of civil service employees, totaling about 50 million.[1] Second, there is a well-documented gender gap in China's civil service that widens with rank, culminating with a very small proportion of women in top leadership positions (Chen, 2022; Jiang, Eaton and Kostka, 2023).[2] Third, China has a long history of hiring civil servants through standardized exams. We

[1]The estimate is based on microdata from the 2010 Chinese census. When focusing exclusively on civil servants regulated by the "Law of Civil Servants," which is the group of interest in this paper, the number is reduced to seven million. Source: `https://www.gov.cn/xinwen/2016-05/31/content_5078420.htm`.

[2]Indeed, since its founding in 1949, China has never had a female president, nor has any women served on the Standing Committee of the Political Bureau of the Chinese Communist Party's Central Committee, the country's highest leadership body. Source: `https://www.cnn.com/2017/10/22/asia/china-female-leaders/index.`

obtain data from these exams which allows us to open the "black box" of job evaluations and empirically assess the extent of gender bias in candidate evaluation.[3]

In the first part of our paper, we examine the relationship between the content of job advertisements and application behavior by gender. We analyze job advertisements from the Chinese civil service between 2010 and 2018, covering all provinces. We find stated gender preference in a large proportion of advertisements, despite Chinese law prohibiting such practices. Roughly 23 percent of jobs express a preference for male applicants, and only 3.5 percent express such a preference for women. This contrasts with the nearly equal levels of male and female gender preference observed in job advertisements in China's private sector (Kuhn and Shen, 2013).

We then show that application behavior responds to the gender preference stated in job advertisements, which in turn affects the aptitude of the candidate pool for each job. Jobs expressing preference for males attract fewer candidates overall and, significantly fewer female candidates. In contrast, job advertisements requesting women yield a candidate pool with the highest assessed aptitude, as measured by civil servant exam scores. On the other hand, job advertisements requesting men draw candidates with the lowest average performance.

Second, we show how gender preference appears in candidate evaluations and final hiring decisions. We analyze official administrative data containing candidate evaluations on the types of candidate evaluation: the written civil service exam and the subsequent oral evaluation. On the written exam, in which candidate identity is not disclosed to the assessor, female candidates perform much better than males. For evaluations on the oral exam, in which assessors know each candidate's gender identity, this advantage is almost reversed. This leads to a 4.4 percent greater likelihood of males being hired for a given job than females.

We then probe and interpret these findings. We argue that men's relative advantages in the non-blind oral test come from gender bias in favor of male candidates. We first provide suggestive evidence that the blind and non-blind test scores measure similar skills of candidates. We also

`html`.

[3]China's civil servant exam dates back to the Sui Dynasty, approximately 1,500 years ago. Since the early 1990s, all junior candidates have been required to take standardized exams to qualify for civil servant employment.

present analyses suggesting that our main estimates are not confounded by possible gender-specific factors that may differentially affect the blind and non-blind test scores.

Third, we study the early-career promotion patterns of male and female civil servants hired through this process. We focus on two groups of civil servants: junior civil servants from various departments in a central Chinese province, and junior civil servants in China's judicial system, for whom we have access to agreed-upon performance metrics. Across both samples, our analysis reveals consistent patterns: within the first ten years of employment, a significant gap emerges in the likelihood of promotion and career trajectories between male and female civil servants. We estimate that female civil servants are 35-45% less likely than males to be promoted to higher rank within the first ten years of employment. We provide suggestive evidence that the gender promotion gap is unlikely explained by gender differences in work performance, and cognitive or non-cognitive skills. Instead, the gender promotion gap appears to stem from gender bias in promotion decisions, which rely heavily on subjective evaluations.

Finally, we study the impact of two policies that alleviate the influence of gender bias in these processes. The first policy prohibits explicit gender preference in job advertisements. Although implicit preference can still be conveyed – for example, by stressing harsh working conditions or frequent overtime and travel – we find that this policy increases the proportion of women who advance to the final round of assessment and, ultimately, who are hired. The second policy changes the way candidates are assessed, severing potential ties between assessors and candidates, should they be hired. This change effectively neutralizes the male comparative advantage in the evaluation phase where candidates' gender identity is revealed to the assessor. Together, our results show that the gender disparities we reveal are not inevitable and can be mitigated through targeted policies that disrupt or eliminate opportunities for gender bias at different stages of the hiring process.

We contribute to the following three literatures. First, our study advances understanding of gender inequality in the labor market by estimating the forces that create gender disparities in career trajectories across several distinct and important early career events. We provide evidence of these forces in five distinct junctures affecting the trajectory of early career professionals: job

advertisements, application behavior, candidate evaluation, hiring decisions, and promotion decisions. Together, our results show a complete picture of the ecosystem of forces that affect the labor market outcomes of professionals across much of their careers, helping understand the persistent underrepresentation of women in senior white-collar roles spanning business, politics, and science (Ceci and Williams, 2011; Blau and Kahn, 2017; Huang et al., 2020). By studying several of these events together and their impact on the same group of professionals, we build upon and integrate findings from earlier studies that typically focus on just one of these events—for example, job advertisements, hiring, or promotion—individually (Goldin and Rouse, 2000; Kuhn and Shen, 2013; Hospido, Laeven and Lamo, 2022; Mocanu, 2023; Card, Colella and Lalive, 2024).

We also advance this work by providing evidence using administrative data from an extremely large employer in a non-experimental setting. By complementing a literature that has largely relied on field experiments or resume audit methods, we aim to bring additional external validity to this body of work (Bertrand and Mullainathan, 2004; Breda and Hillion, 2016; Bertrand and Duflo, 2017; Cahuc et al., 2019; Kuhn, Shen and Zhang, 2020).[4] Further, we contribute to a small but growing literature that analyzes the contents of job advertisements to reveal gender discrimination in the labor market (Kuhn and Shen, 2013; Neumark, 2018).

Second, we contribute to the literature on hiring decisions by opening the "black box" of gender bias in candidate evaluation and recruitment decisions. We make progress on this problem by exploiting administrative data from China's civil service exam, where candidate aptitude is evaluated using quantitative, comparable measures that are usually proprietary and unavailable to researchers. These evaluations occur twice: once when the candidate's identity is hidden from the evaluator, and once when it is not, enabling us to estimate the extent of evaluator gender bias (Correll et al., 2020; Mocanu, 2023).[5]

The civil service hiring decisions we study are based solely on a weighted average of these

[4] One recent example is (Amer, Craig and Van Effenterre, 2023), which uses field experiments with blind and non-blind tests to show how gender bias affects hiring of computer programmers.

[5] This approach, comparing "blinded" and "unblinded" evaluations by group to estimate bias, is a common tool in the estimation of the incidence and severity of teacher bias related to gender, race, and other social groups (Goldin and Rouse, 2000; Lavy, 2008; Hinnerich, Höglin and Johannesson, 2011; Botelho, Madeira and Rangel, 2015; Breda and Ly, 2015; Ferman and Fontes, 2022).

test scores. This yields a unidimensional measure of aptitude that is comparable across candidates and positions. Our focus on how different evaluation components do and do not allow for gender bias builds on Mocanu (2023), which examines how different screening practices affect gender equity in hiring. Our data also contain roughly 800,000 candidate observations linked to evaluation scores and demographics, allowing us to delve deeper into estimating the incidence and intensity of gender bias in hiring decisions, and to differentiate between statistical and taste-based discrimination.

Finally, we contribute to a growing literature on the economics of employee recruitment in the public sector (Breda and Hillion, 2016; Estrada, 2019; Aneja and Xu, 2022; Dahis, Schiavon and Scot, 2023; Mocanu, 2023). Although much research has focused on gender discrimination in the private sector, few studies have explored this issue within civil service recruitment, possibly due to data limitations. Contrary to expectations, our findings reveal a significant prevalence of gender preference in job advertisements in the public sector. In the candidate evaluation process, our findings relate to Breda and Hillion (2016), which assesses gender bias in teacher hiring evaluations. We expand on this work by examining a broader range of civil servant roles, and by better differentiating between statistical discrimination and taste-based discrimination as explanations for the pattern we uncover. We go further in two important ways: we provide evidence of gender bias in the promotion of civil servants, and we evaluate two policies to restrain these pressures at different stages of a civil servant's career.

The rest of our paper proceeds as follows. Section 2 provides information on China's civil service, its recruitment behavior, and candidates assessment methods. Section 3 describes our data. Section 4 documents the prevalence of gender preference in civil service job and their impact on applicant behavior. Section 5 explores gender bias in candidate evaluations during the hiring process and its influence on final employment outcomes. Section 6 presents gender differences in performance and promotion among hired civil servants. Section 7 evaluates two policies to address gender preference in hiring, and the final section concludes.

2 Background

In this section we present background information on China's civil service and how civil servants are recruited.

2.1 China's Civil Service

China has a large public sector, comprising nearly 50 million employees in either government offices or public institutions such as public schools and hospitals (Burns, 2007; Burns and Xiaoqi, 2010).[6] Within China, civil service positions are widely considered both socially prestigious and professionally stable, relative to work in the private sector (Mangal, 2024). In this paper, we focus on employees in civilian government administration, as opposed to those in other public institutions such as schools and hospitals. According to Ministry of Human Resources and Social Security of China, 7.17 million civil servants worked in government offices in 2015.[7]

China's civil service system, historically rooted in the imperial examination system in place for several centuries in China, has evolved into a highly structured and hierarchical system. The administrative hierarchy not only includes the division of administrative regions but also the internal rank divisions within government agencies. At the top is the national level, followed by provincial/ministerial, prefecture/department, county/division, township/section, and unit levels, with both chiefs and deputy chiefs within each level.[8] Among China's civil servants, only 10% even face the possibility of being promoted to the deputy county/division level or above (Zhao, Zhao and Yang, 2022).

There is a well-documented gender gap in China's civil service that widens with rank, leading

[6]The estimate is based on microdata from China's 2010 census.

[7]Source: `https://www.gov.cn/xinwen/2016-05/31/content\_5078420.htm`.

[8]China's civil servant system is structured into six main levels. At the highest level is the nation Level (guo ji), which includes top national leaders such as the President, Premier, Vice Premiers, and State Counselors. Below this is the provincial/ministerial Level (bu ji), which includes provincial governors, vice provincial governors, ministers, and vice ministers. The prefecture/department Level (ting ji) covers prefecture-level city mayors, provincial department directors, and their deputies. The county/division Level (chu ji) includes leaders of county governments and directors within departments. The township/section Level (ke ji) consists of township leaders and chiefs of sections within divisions. At the base is the unit Level (gu ji), responsible for executing specific tasks within sections. Source: `http://www.dltdjw.gov.cn/ywgz/gwygl/202303/t20230309\_3357858.html`.

to very few females in leadership roles (Chen, 2022; Jiang, Eaton and Kostka, 2023). For instance, since the founding of the People's Republic in 1949, China has had no woman serving on the country's central leadership team (the Standing Committee of the Political Bureau of the CPC Central Committee). Looking at leadership one level down, only 4.8% (three out of 62) provincial governors or CPC party secretaries are female in 2024 and 6.1% (223 out of 3,681) of prefecture-level city mayors or party secretaries are female during the period between 2003 and 2020 (Jiang, Eaton and Kostka, 2023).[9] In contrast, in the past few decades, women have accounted for nearly half of junior civil servants.[10]

2.2 Recruitment, evaluation and hiring of junior civil servants

China's civil service exam dates back to the Sui Dynasty (581-618 CE). The imperial civil service used some version of this exam to select government officials for over a millennium, a practice which lasted until the end of the Qing Dynasty (Bai and Jia, 2016). The imperial civil service exam was a meritocratic system used for recruiting government employees, open to all male citizens regardless of social class or background. Because of its focus on literacy and related skills, it also incentivized members of the public to learn to read and write (Bai and Jia, 2016; Xue and Zhang, 2023).

After the founding of the People's Republic of China, a new version of the civil service exam was reintroduced in 1994 with no gender eligibility criteria. In the following years it became extremely competitive and a focus of major public interest and discussion (Burns and Xiaoqi, 2010; Liu, 2023). It is now the only channel through which junior civil servants are hired, with millions of young people taking it each year. In scale it is comparable to the high-stakes Chinese college entrance exam, or *gaokao*: in 2022, more than eight million people attended the civil service exam to compete for about 200,000 positions, as compared to 12 million *gaokao* takers.[11]

China has two types of civil service recruitment process: recruitment for the national civil

[9]Source: `https://www.thepaper.cn/newsDetail\_forward_27957779`.
[10]Source: `https://www.stats.gov.cn/sj/zxfb/202302/t20230203\_1899316.html`.
[11]Source: Zhonggong Education (`https://www.offcn.com`), a large company in China which provides services to help people sit for the civil service exam.

service, which selects employees for roles in the central government, and recruitment for the local civil service, which selects employees for provincial and other lower-level governments. We focus on the local civil service recruitment process. This process focuses on recruiting junior civil servants, and has two stages, similar to civil service exams in other countries (Breda and Hillion, 2016; Greenan et al., 2019; Colonnelli, Prem and Teso, 2020).

Each province runs its own recruitment, but all provinces follow the same two-stage process. In the first stage of the process, each province uses a province-specific standardized written exam that all candidates in that province take. Candidates use their scores on this exam to apply to specific positions, and those with the highest scores are selected to sit for a second-stage, position-specific oral exam. In most cases, there are three candidates who progress to the second stage. In the second stage candidates are again evaluated and given a numeric score. The final performance of applicants is a weighted average of the first stage and second stage exam scores, usually with equal weight given to each. The government then admits the highest-ranked applicants based on their final scores and the number of openings available for a given position.

The first-stage written exam is composed of two sections. The first section is a multiple-choice section assessing the candidate's administrative ability. This covers pre-defined, publicized candidate traits, including basic knowledge, logical reasoning, and quantitative analysis skills. The second section is an essay, which evaluates the candidate's problem-solving ability and written communication skills.[12] Performance on the two parts of the written exam are assessed on a 0-100 point scale and their scores are weighted equally to generate a total written exam score for each candidate.

For every junior civil service position, all citizens are eligible to sit for the first-round written exam, conditional on satisfying job-specific requirements such as gender, age, and education. The number of applications per position varies with demand: there is no limit on the number of applicants per position, and applicants can apply to the same position in subsequent years if their earlier applications are unsuccessful. In a given year, all applicants to civil service jobs in a partic-

[12]See Table 1 for details on the content of each part of the exam.

ular province take the same exam, and their performance is evaluated based on a single standard grading key. Written exam scores are graded by assessors who cannot observe the identity of the candidate, minimizing the risk of bias (including, but not limited to gender bias) in assessment.

The second-round oral exam aims to test whether an applicant's ability or skills are matched with the requirements of the position. It focuses on candidate traits such as analytical ability, interpersonal communication skills, and organization / coordination skills (of Human Resources and of Chongqing Municipality), 2009; of Human Resources and of Hainan Province), 2013). The oral exam usually takes the form of a structured face-to-face interview, in which interviewers ask a candidate a set list of questions in a standardized order and grade performance against a predetermined scoring system.[13] The content of the oral exam is the same for all candidates applying to a given position but can vary across positions (of Human Resources and Security), 2015). All candidates for a given position are evaluated by the same group of examiners. The goal of this test format is to yield second-round scores which are consistent and comparable across candidates for the job in question.

The evaluation committee for the second-round exam usually consists of seven oral examiners who administer all the exams for a given position. The majority are randomly selected from a provincial-level database of examiners, but one or two examiners are often drawn from local high-ranking civil servants related to the job in question (of Human Resources and of Chongqing Municipality), 2009). To ensure fair assessment and to minimize variation in assessment fidelity across examiners, the government requires that examiners assess candidates only on their merits. Further, many measures are in place – such as mandating that no other candidate information be revealed to examiners, randomly assigning assessors to candidates, and mandating that assessors not ask candidates about sensitive background information – to maximize fairness (of Human Resources and Security), 2015). While such requirements are designed to minimize the risk of many types of evaluation bias (such as by class, educational background, or social network), they cannot prevent gender discrimination to the same extent because the gender of each candidate is observable to the

[13] In rare cases, the oral exam also includes a component in which groups of candidates discuss issues without a pre-specified leader, similar to the US Foreign Service exam.

assessor.[14]

3 Data

In this section, we describe our four main data sources: data from publicly posted job advertisements for civil service positions; data on the applications made to these jobs; data on the performance evaluation of applicants to civil service jobs; and the work performance and promotion records of hired civil servants.

3.1 Job advertisements

Our first dataset comprises advertisements for junior civil service positions. We collected these from two main sources: the official websites of provincial governments, and from private companies who sell services to help people prepare to apply for civil service jobs.[15] These data comprise roughly 870,000 job advertisements together recruiting 1.37 million civil servants. These include jobs across all 31 provinces of mainland China over the period from 2010 to 2018.

These job advertisements specify various desired traits of the applicant. The most commonly specified traits relate to gender, education (e.g., degree, major), prior work experience, and other traits (e.g., hukou/household registration status, membership in the Communist Party of China). Some advertisements impose requirements on candidates' skills, such as writing ability, computer skill, self-motivation, attitude towards hard work and long hours, disposition towards service, leadership skill, willingness to work as part of a team, creativity, independence, and ability to adapt to new situations.

3.2 Applications to civil service jobs

Our second data set comprises data from roughly 320,000 applications to civil service jobs. These data come from four provinces (i.e. Gansu, Hubei, Hunan, and Sichuan) where such data were

[14] In recent years, many local governments have made test scores publicly available to reduce the risk of corruption in junior civil servant recruitment. This also suggests that the gender discrimination in the time period we study may be less severe than was the case earlier, given the pressures for fairness put on assessors from these transparency efforts.

[15] Zhonggong Education,`https://www.offcn.com`, and Huatu Education, `https://www.huatu.com`.

available online and which we then manually collected.[16]

For each application, we have information on the job applied to, as well as candidate name and, in more than half of cases, gender.[17] These data contain applicants to a total of 10,262 job vacancies.[18] Contrary to the popular belief that too many women apply for civil service jobs in China, we find more than half (56.69%) of applicants are male.[19]

3.3 Evaluations of applicants

We also collect two sets of data reporting how applicants to the civil service are evaluated. The first is performance data from the first-round written exam for all observations in our applicant data described in Section 3.2. These data include test scores for both the evaluation of the candidate's administrative ability (a multiple choice test) as well as their writing ability (an essay).

The second round comprises official performance evaluations from the face-to-face interviews. The law mandates that these data be made publicly available. We hand-collected these from both official government websites and the website of private companies which collate these data.[20] These data contain second-round assessments for roughly 730,000 job candidates who applied to jobs between 2010 and 2018 in 21 of China's 31 provinces.[21] These data contain traits of the individual (name, gender, education) and assessment scores, both for the written administrative ability test and essay test, as well as scores from the second round oral assessment.

According to provincial Administrations of Civil Service, the administrative ability test part of the written exam aims to assess a set of candidate skills which are graded objectively, including basic knowledge, language comprehension and expression, logical reasoning, and quantitative

[16] Other provinces generally do not make public the type of data we collect. Local governments are not mandated to publicize the first-round applicant information.

[17] When gender information is unavailable, we implement a machine-learning algorithm, described later in this section, to infer gender information based on candidate name.

[18] One unique feature of our civil servant recruitment data is that employers do not filter applicants by gender in the first round when job advertisements prefer men but are not exclusively limited to men. This suggests that our analysis sample reflects the self-selection of applicants, rather than pre-screening of employers.

[19] Source: `https://www.thepaper.cn/newsDetail_forward_10021801`.

[20] These include Zhonggong Education Co. (www.offcn.com), Huatu Education Co. (www.huatu.com), and others.

[21] Data from the remaining 10 provinces are excluded because local governments publicized the exam data without the necessary information (stated gender or name) to identify gender for the observation

Table 1. Skills tested in different exams

Test type	*Skills assessed*
Administrative (written test)	Basic knowledge, Language comprehension and expression, Logical reasoning, Quantitative analysis skills
Essay (written test)	Comprehensive analysis, Problem-solving and problem-posing, Reading comprehension and written communication, Implementing and organizing, Flexibility
Oral exam	Comprehensive analysis, Problem-solving, Language communication and interpersonal communication, Planning and organizing, Flexibility and emergency response, Creativity, Professional ability, Job-application motivation and job match, Appearance and manners

analysis skills. In the essay part of the written exam, candidates are asked to explain how they would respond to a situation (for instance, an unexpected event) as if they were a government official. This evaluates skills beyond the quantitative and other, more easily measurable skills covered in the administrative ability test. The skills include comprehensive analysis, problem-solving and problem-posing, reading comprehension and written communication, implementing and organizing, and flexibility of dealing with complicated situations. We list the skills assessed in each test in Table 1. Visual inspection reveals very little overlap between the skills assessed in the administrative test and those assessed in the essay.

In the second round, candidates sit for an oral exam. The skills assessed in the oral exam include comprehensive analysis skills, problem-solving skills, communication skills, ability to organize, and ability to deal with emergencies and/or complicated situations. Visual inspection of the stated skills, as well as a review of past test questions, reveal a substantial overlap between the skills assessed on the oral exam and on the essay part of the written exam. Based on this overlap, we argue that the skills measured by the essay section in the first-round test (in which candidate gender identity is concealed) are comparable to those measured in the second-round oral test (in which candidate gender identity is observable to the assessor).

A key variable throughout our analysis is candidate gender. In these data, we have two versions of this variable. We use the candidate's reported gender in the 250,000 cases where it is

available (roughly one third of the total). For the remaining cases, we apply a machine-learning algorithm to infer gender from the candidate's name. We validate this classification by applying the same algorithm to observations where gender is known. The algorithm correctly classifies gender in 83 percent of these observations. Furthermore, it is much more accurate for males than for females.[22] Such misclassification would bias our later results against finding a penalty for females, as female candidates are more likely to be misclassified as males.[23]

In our analysis, the main dependent variables are the raw test scores and their transformation into percentile ranks. We calculate the percentile rank for each candidate within a given position, setting the lowest score to 0 and the highest score to 1 (Breda and Hillion, 2016). Each variable provides distinct information valuable for our analysis. The raw test score represents the observable measure of ability for candidates within a province in a given year. Since all candidates take the same test, these scores allow us to compare the ability of candidates applying for different positions. In contrast, the percentile rank reflects the relative ability of candidates who applied for the same job. This metric is particularly useful when the score dispersion for a particular job is narrower compared to other positions. It also addresses the censoring issue that arises because candidates who proceed to the second-round assessment have already passed the threshold set for advancement. Finally, because candidates with varying abilities apply for different jobs, relative comparisons rather than absolute scores are crucial for employers when making hiring decisions.

3.4 Work performance and promotion of civil servants

Our final collection of data captures official records of job performance and promotion for Chinese civil servants. We collect and use two sets of complementary data in this analysis. The first comprises civil servants sample from Hubei Province, a large province located in central China.[24] We collect promotion information of civil servants from more than 100 local county-level and

[22]The accuracy rate is 93.48% for males and 71.97% for females.

[23]On one hand, women perform better than men in the written exam but not in the oral test, leading to a smaller employment penalty for women when we use a machine-learning algorithm to infer gender. On the other hand, women perform better than men in the civil servant exam but are less likely to be promoted after being hired, leading to a smaller promotion penalty for women with the machine-learned gender.

[24]Due to the significant resources needed to manually collect promotion data, we chose to focus on one province – Hubei – which also has the largest share of observations in our civil servant exam dataset.

prefecture-level government websites in the province, which publicize official personnel changes. We then match these data on promotion with individuals present in our civil servant exam dataset. This yields a match of civil service scores to promotion information for 586 hired civil servants associated with 416 job positions recorded in the civil servant exam dataset.

The promotion information on civil servants from Hubei Province includes their rank at each promotion and time of promotion. These data cover the period from 2013 to 2023. Because these civil servants are relatively young, they are unlikely to be promoted to higher-level ranks: most of them are promoted to low ranks such as vice-unit, unit, vice-section, and section levels. For the other 1,706 civil servants in our data from Hubei Province who were hired at the same time for the same position, but for whom we have no information on promotion, we assume that they have not been promoted during our analysis period.[25]

Our second dataset focuses on civil servants employed in the judicial system in roles that lead to becoming a judge in China's local courts. We focus on these civil servants because we are able to collect fine-grain data on their productivity as a result of transparency legislation enacted in 2014 requires that court verdicts be made publicly available.[26] The judgment files contain names and ranks of judges who are responsible for the verdicts, who constitute our main sample of judges. This effort yields data on roughly 340,000 judges linked to 85 million judgments filed between 2013 and 2021 and downloaded from official government websites.[27] We then use judge names to match these observations to our data on civil servant exam scores. This yields a dataset containing work performance and promotion history for 5,100 court judges linked to roughly one million judgment files. We use these data to explore differences in on-the-job productivity and promotion by gender.

Using these data, we create a series of variables capturing on-the-job performance for the

[25]We believe this assumption is unlikely to lead to estimation bias in our analysis because the availability of public promotion information is unlikely to vary systematically by gender.

[26]To promote judicial transparency and fairness, China's national-level central court (the Supreme People's Court) has required all local courts to publicize both contemporary and historical verdicts on its official website (China Judgments Online, `https://wenshu.court.gov.cn/`) since 2014.

[27]The total number of records over this period is roughly 130 million (Cao, Liu and Zhou, 2023; Liu et al., 2022). The remaining files are beyond our reach because the websites of certain localities implement anti-scraping measures.

judges in our sample. Our performance indicators follow official policy recommendations for how to assess the performance of judges in China's judicial system.[28] These comprise five measures. The first is the number of cases on which the judge served in a given year. The second is the average time taken to close a case in a given year, which is a rough measure of the efficiency with which the judge processed cases. The third is the number of words of reasoning in the judgment file issued by the judge, which is a rough measure of the quality of a judicial decision (Li, 2008; Cao, Liu and Zhou, 2023). We also construct the following two distinct measures of the quality of a decision: the proportion of first-instance cases (that is, new cases, rather than appeals) in a given year which were appealed; and the proportion of first-instance cases in which the judge's original decision was reversed by a higher court (Cao, Liu and Zhou, 2023; Liu et al., 2022).

We can also extract data on the civil servant's rank from the judgment files we study, and we use changes in this variable over time to create promotion records for each civil servant in our data from local courts. There are three possible ranks these civil servants can inhabit. The lowest rank is "court clerk," and civil servants in this rank handle administrative tasks and keep records of the proceedings during each trial. The next higher rank is "judge assistant." Civil servants in this rank assist judges in their judicial work. The highest rank in our data is "judge," a role in which the civil servant runs the court and issues judicial decisions.[29] We use these to generate two related measures: the incidence of promotion in a judge's career, and the time taken from being hired to promotion.

4 Job Advertisements and Application

In this section, we first document gender preference in civil service job advertisements over time. We then show how applicants respond to the gender preference in job advertisements. Finally,

[28] We use the *Guiding Opinions on Strengthening and Improving the Assessment of Judges* issued by the Supreme People's Court, which emphasizes both "quantity" of cases and "quality" of judicial decisions. Source: `https://www.court.gov.cn/zixun-xiangqing-330041.html`.

[29] Court clerks and judge assistants work under the guidance of judges. Before a law reforming the court system was passed in 2016, there was a rank of "assistant judge," which could also act as the role of a judge in times of need and issue judicial decisions. After 2016, the rank of assistant judge was abolished. In our analysis, we count both assistant judges and judges as members of the judge category.

we explore how the gender preference in job advertisements is associated with candidates' performance in the civil service exam and their chances of final employment.

4.1 Gender preference in civil service job advertisements

Some civil service job advertisements impose requirements on applicant gender. This occurs despite the fact that it is illegal to explicitly state gender preference in job advertisements in China. While such explicit discrimination is also present in some private sector job advertisements (Kuhn and Shen, 2013), we find that it is far more prevalent in job advertisements issued by government.[30]

We describe this pattern in Figure 1. We find that 22.77% of civil service job advertisements in our sample either explicitly request only male applicants, or state a preference for male applicants. In contrast, only 3.53% of civil service job advertisements explicitly request only female applicants and almost no civil service job advertisements state a preference for female applicants.[31]

The levels of gender preference in civil service job advertisements are both more common and more skewed towards males than found in the private sector. For instance, by analyzing 1.05 million private-sector job advertisements in China, Kuhn and Shen (2013) find that only 5.5 percent of private-sector job advertisements state a preference for men and a roughly similar proportion of job advertisements state a preference for women. While the proportion of civil service job advertisements explicitly requesting male applicants declined during the 2010-2018 period, the proportion of job advertisements stating a preference for male applicants rose over this same window.[32]

We also uncover a related phenomenon: some governments publish job advertisements separately targeting men and women for identical positions.[33] Within these advertisements, roughly two-thirds specify that only males are eligible, while the other third specify that only females are eligible. This behavior suggests that these governments attempt to explicitly set the proportion of

[30] In the past few decades, China has issued several laws and regulations prohibiting gender discrimination in employment. These are particularly relevant to gender-targeted job ads; see Appendix 2 in Kuhn, Shen and Zhang (2020) for a review of related laws or regulations in China.

[31] Figure A1 of the Appendix presents word clouds of job titles in job advertisements preferring men or women.

[32] Our job advertisements analysis suggests that gender preference may be related to requested academic fields. As Figure A2 of the Appendix shows, compared with job advertisements preferring women, those preferring men are more likely to request graduates from natural sciences or engineering fields.

[33] More than 27,000 job advertisements appear to fit this pattern.

Figure 1. Gender preference in civil service job advertisements over time

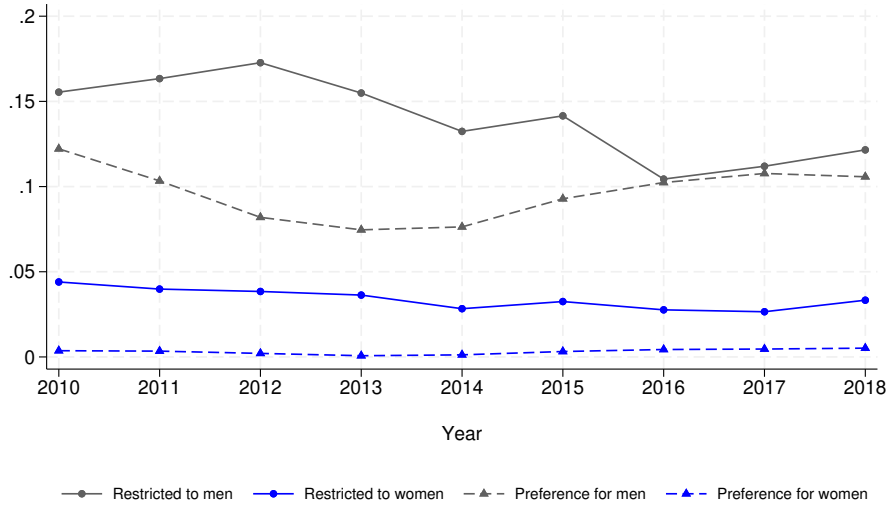


Figure 1 notes: this figure shows the proportions of four types of gender preference in civil service job advertisements during the 2010-2018 period.

male and females who are recruited, again favoring male applicants at a ratio of roughly two to one. It is also consistent with anecdotal accounts that government employers fear that too many women and too few men are hired via the civil service exam.[34]

4.2 How applicants respond to gender preference in job advertisements

Next, we explore whether the presence of stated gender preference or restrictions in job advertisements affects the traits of those who apply to jobs. We show male shares of first-round applicants in Figure 2, broken down by the following four types of gender preference stated in jobs: jobs explicitly restricting eligibility to men, jobs explicitly restricting eligibility to women, jobs expressing a preference for male applicants, and jobs without gender preference. We find that almost no women apply to jobs explicitly restricting eligibility to men, and vice versa, because the civil service enforces these strict restrictions on applications. When job advertisements state a preference for men, men account for about two thirds of candidates in the application. In contrast, when job advertise-

[34]Source: `https://www.thepaper.cn/newsDetail_forward_10021801`, accessed on March 5, 2024.

Figure 2. Applicants' responses to gender preference in civil service job ads

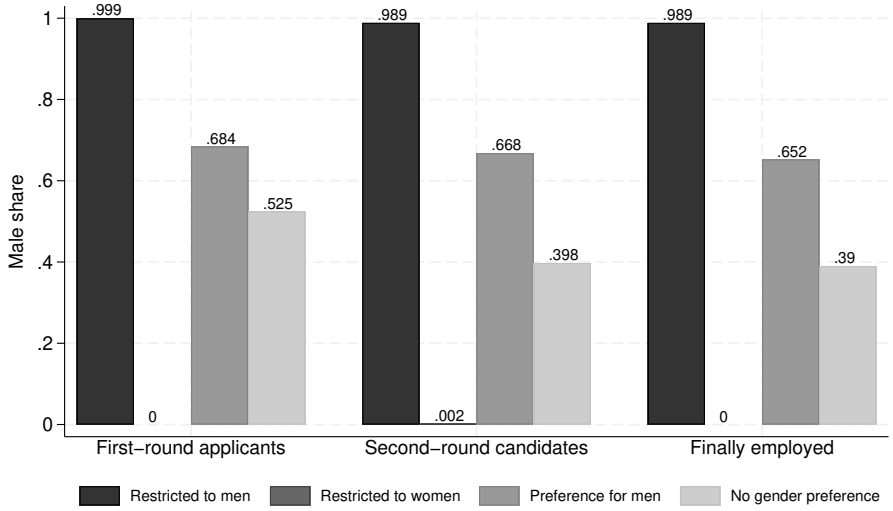


Figure 2 notes: this figure shows applicants' responses to gender preferences in civil service job ads.

ments do not state a gender preference, the gender composition of applicants is more balanced. We find a similar pattern among second-round candidates and the finally employed people, except for a lower male share when job advertisements have no gender preference. In short, these results suggest applicants respond to the presence of both explicit and implicit gender preference in job advertisements (Kuhn, Shen and Zhang, 2020; Flory et al., 2021).

We estimate applicants' responses to gender preference in job advertisements using the following equation:

$$(1) \quad Y_{ict} = \alpha_0 + \alpha_1 * ReqMen_{ict} + \alpha_2 * ReqWomen_{ict} + \alpha_3 * PrefMen_{ict} + \alpha_4 * X_{ict} + \lambda_{ct} + \varepsilon_{ict}$$

This equation shows us how the outcome variable – in all cases, some measure of applicant characteristics – responds to stated gender preference, as compared to jobs that have no stated preference. Our outcome variable, Y_{ict} is measured at the level of job i based in city c, advertised in year t. Our outcome variables include the total number of male applicants, the share of applicants who are male, and average applicant performance on the component parts of the written civil service exam, as well as their overall scores.

Table 2. Applicant responses to gender preference in job ads

	No. of applicants		Prop. of male applicants			
	(1)	(2)	(3)	(4)	(5)	(6)
Requesting men	-41.261***	-49.779***	0.515***	0.525***	0.421***	0.447***
	(5.969)	(6.791)	(0.026)	(0.032)	(0.013)	(0.018)
Requesting women	-21.448*	-9.756	-0.498***	-0.431***	-0.378***	-0.328***
	(12.473)	(10.809)	(0.035)	(0.034)	(0.028)	(0.027)
Preferring men	-54.649**	-30.984	0.172*	0.178**	0.085**	0.078**
	(24.851)	(20.203)	(0.086)	(0.064)	(0.036)	(0.030)
N	6144	6141	1778	1775	6144	6141
R^2	0.090	0.418	0.425	0.522	0.454	0.548
Controls	No	Yes	No	Yes	No	Yes
Machine-learned gender	–	–	No	No	Yes	Yes

Table 2 notes: this table presents estimates of α_1, α_2, and α_3 from Equation 1. The dependent variable is number of applicants or proportion of male applicants for a job position in a given year. We control for job-related variables in columns (2), (4) and (6), including number of slots of the job position, education requirements, major requirements, a dummy variable for CPC (Communist Party of China) membership, job experience, local hukou (household registration status), and cognitive or non-cognitive skill requirements. We use actual gender information in columns (3) and (4) and machined-learned gender information based on applicants' names in columns (5) and (6). Standard errors are clustered at the city-year level. *** < 0.01, ** < 0.05, * < 0.1.

Our main independent variables are indicator variables capturing whether a job ad explicitly restricts eligibility to either men ($ReqMen_{ict}$, not allowing women) or women ($ReqWomen_{ict}$, not allowing men), or states a preference for men ($PrefMen_{ict}$, still allowing women). We also control for stated job requirements related to education, major, years of experience, hukou (household registration location) status, Communist Party membership, and skills. We also control for the number of vacancies in a given advertisement and include city-by-year fixed effects (λ_{ct}). Standard errors are clustered at the city-year level. We present our results in Table 2.

We find that the average number of applications to jobs explicitly requesting women is larger than that to jobs requesting or preferring men. We also find that applicants respond to gender preference in job advertisements as expected, even after controlling for job requirements, suggesting applicants' compliance with employers' gender requests. We find such applicants' compliance behavior even when job advertisements only hint that jobs are more appropriate for men.

Table 3. Applicants' first-round performance relating to gender preference in job ads

	Admin test scores		Essay scores		Written exam scores	
	(1)	(2)	(3)	(4)	(5)	(6)
Requesting men	-0.616	-0.522	-1.041**	-1.491***	-0.829**	-1.007*
	(0.447)	(0.517)	(0.398)	(0.519)	(0.407)	(0.510)
Requesting women	1.692***	1.898***	3.218***	2.579***	2.455***	2.239***
	(0.391)	(0.682)	(0.426)	(0.521)	(0.385)	(0.567)
Preferring men	0.904	0.298	-1.398**	-1.119*	-0.247	-0.410
	(0.804)	(0.726)	(0.618)	(0.577)	(0.681)	(0.635)
N	6144	6141	6144	6141	6144	6141
R^2	0.566	0.603	0.612	0.629	0.572	0.597
Controls	No	Yes	No	Yes	No	Yes

T able 3 notes: this table presents estimates of α_1, α_2, and α_3 from Equation 1. The dependent variable is applicants' administrative test scores, essay scores, or total written exam scores for a job position in a given year. We control for job-related variables in columns (2), (4) and (6), including number of slots of the job position, education requirements, major requirements, a dummy variable for CPC (Communist Party of China) membership, job experience, local hukou (household registration status), and cognitive or non-cognitive skill requirements. We use actual gender information in columns (3) and (4) and machined-learned gender information based on applicants' names in columns (5) and (6). Standard errors are clustered at the city-year level. *** < 0.01, ** < 0.05, * < 0.1.

4.3 The relationship between gender preference in job advertisements and aptitude of the applicant pool

4.3.1 The first-round written exam

We next examine how the measured aptitude of applicants to civil service positions varies with the presence of gender preference in civil service job advertisements. We present a series of condition correlations between these two variables in Table 3. We find that applicants to jobs requesting women have better performance on the first-round written exam, compared with those to jobs without gender preference. In contrast, applicants to jobs requesting men have the lowest performance; the average first-round scores of applicants to jobs stating a prefernece for men but still allowing female applicants fall in between. These findings suggest that applicant quality also responds to gender preference in civil service job advertisements.

One concern with interpreting these relationships is that gender preference stated in job ad-

vertisements could reflect unobservable job characteristics and not gender discrimination against women per se. For instance, jobs restricting eligibility to men could have higher safety risks – for example, being located in unsafe and/or remote regions, or otherwise be (perceived as) unsuitable for women. To address this concern, we restrict our attention to those jobs in which multiple people are recruited for the same position in the same employer via multiple distinct advertisements, some of which limit eligibility to male applicants, and others which limit eligibility to female applicants. We reproduce the same analysis from Table 3 only using this sample, and report results in Table A1. Here we find very similar patterns of stated preference for men in advertisements where we can definitively rule out such suitability concerns.

4.3.2 The second-round face-to-face interview and final employment

In this subsection, we explore whether the male share among second-round candidates is consistent with gender preference in job ads. We also compare candidates' average performance of the second-round face-to-face interview across jobs with various gender preferences.

Our analysis proceeds in similar fashion to the results in previous section, estimating ordinary least squares regressions to characterize the differences in candidate gender and aptitude assessment by the gender preference expressed in the job advertisement. We present this analysis in Table 4.

We find that, as in the first round of applications and candidate assessment, the candidates who make it to the second round vary by stated preference. As anticipated, jobs that restrict eligibility to men or express preference for male applicants bring in more second-round male candidates than other jobs (columns 1 and 2). In the remaining columns, we find that candidate aptitude varies with these requests as well. In both jobs that restrict eligibility to men and state a preference for men, candidates have lower aptitude scores - on the written part of the exam (column 3), on the oral part of the exam (column 4), and on the overall assessment score (column 5). Conversely, in jobs which limit eligibility to females, second-round candidates do no worse on the written exam than second-round candidates in jobs without a gender preference, and do better on oral exams and

Table 4. Applicants' second-round performance relating to gender preference in job ads

	Prop. of male applicants		Written scores	Oral scores	Total scores
	(1)	(2)	(3)	(4)	(5)
Requesting men	0.585***	0.509***	-3.408***	-0.760***	-1.750***
	(0.016)	(0.018)	(0.213)	(0.126)	(0.138)
Requesting women	-0.375***	-0.384***	-0.294	0.553***	0.642***
	(0.015)	(0.019)	(0.314)	(0.203)	(0.232)
Preferring men	0.262***	0.181***	-1.268***	-0.276***	-0.832***
	(0.022)	(0.019)	(0.155)	(0.073)	(0.096)
N	14248	23975	24033	24033	24033
R^2	0.517	0.399	0.648	0.351	0.485
Machine-learned gender	No	Yes	–	–	–

Table 4 notes: this table presents estimates of α_1, α_2, and α_3 from Equation 1. The sample is limited to those who attend the second-round oral test. The dependent variable is proportion of male applicants, written test scores, oral test scores, or total exam scores for a job position in a given year. We control for job-related variables in all columns, including number of slots of the job position, education requirements, major requirements, a dummy variable for CPC (Communist Party of China) membership, job experience, local hukou (household registration status), and cognitive or non-cognitive skill requirements. We use actual gender information in columns (1) and machined-learned gender information based on applicants' names in columns (2). Standard errors are clustered at the city-year level. *** < 0.01, ** < 0.05, * < 0.1.

overall.[35]

Next, we analyze these relationships among the individuals who are finally hired. We present our results in Table 5. In Columns 1 and 2, we show that, as previously observed, gender restrictions and preferences influence the gender composition of hires. Specifically, we find that the lower aptitude measures persist for both jobs that restrict eligibility to males and jobs that express preference for males, even at the hiring stage. For jobs expressing preference for females, the advantage seen in Table 4 disappears. As a whole, the results suggest that women may be disadvantaged in hiring when job advertisements prefer men rather than women.

To sum up, we find that gender preference is common in China's civil service job ads, with 22.77% of them preferring men and only 3.53% preferring women. We also find that applicants

[35] As we have done with first-round applications, we also limit the analysis to the job advertisements that separately request men and women but are from the same position of the same employer in the second round. As shown in Table A2, we find results similar to those in Table 4, suggesting that the concern about unobservable confounding factors may be minor in the second-round estimation results.

Table 5. Final employment relating to gender preference in job ads

	Test score		Percentile rank	
	(1)	(2)	(3)	(4)
Male	0.011		0.009	
	(0.043)		(0.008)	
Essay score	0.031***	0.032***		
	(0.001)	(0.001)		
Male * Essay score	0.000			
	(0.001)			
Male (ML)		0.018		0.013**
		(0.031)		(0.006)
Male (ML) * Essay score		0.000		
		(0.000)		
Essay percentile rank			0.319***	0.323***
			(0.011)	(0.009)
Male * Essay percentile rank			0.036***	
			(0.013)	
Male (ML) * Essay percentile rank				0.031***
				(0.009)
N	106596	180005	106596	180005
R^2	0.125	0.122	0.112	0.107
Machine-learned gender	No	Yes	No	Yes

Table 5 notes: this table presents estimates of α_1, α_2, and α_3 from Equation 1. The sample is limited to those who are finally employed through the civil service exam. The dependent variable is proportion of male applicants, written test scores, oral test scores, or total exam scores for a job position in a given year. We control for job-related variables in all columns, including number of slots of the job position, education requirements, major requirements, a dummy variable for CPC (Communist Party of China) membership, job experience, local hukou (household registration status), and cognitive or non-cognitive skill requirements. We use actual gender information in column (1) and machined-learned gender information based on applicants' names in column (2). Standard errors are clustered at the city-year level. *** < 0.01, ** < 0.05, * < 0.1.

comply with employers' gender requests and such compliance behaviors also exist in the second-round face-to-face interview and final employment outcomes. Finally, we find that candidates (and final employees) of jobs requesting women have the best performance on job interviews measured by test scores. In contrast, candidates of jobs requesting men have the lowest performance while candidates of jobs preferring men (still allowing women) have the performance in between. The above findings suggest that women are disadvantaged in the civil service recruiting process because they are given fewer chances and have to compete with their high-ability female peers.

5 Hiring

In this section, we study how gender preference manifests in the civil service hiring process. First, We characterize how gender preference manifests in the assessment of candidates. Second, we explore alternative explanations for gender differences in candidate assessment. Finally, we show how gender preference affects final employment decisions and discuss two competing types of gender bias: statistical discrimination and taste-based discrimination.

5.1 Gender preference in candidate assessment

In this subsection, we exploit several features of the civil service examination process to unpack the ways in which gender preference exerts itself in candidate assessments throughout the hiring process. First, as shown in Subsection 3.3, the skills assessed in the first-round essay test are comparable to those assessed in the second-round oral test. Second, candidates' gender identity is shielded from the assessor in the first-round blind assessment, while candidates' gender identity is known to the assessor in the second-round non-blind assessment. Our analysis compares the two assessments.

We implement this comparison with the following regression specification:

$$Y_{ipcto} = \alpha_0 + \alpha_1 * Male_i + \alpha_2 * Oral_{ipcto} + \alpha_3 * Male_i * Oral_{ipcto} + \lambda_p + \varepsilon_{ipcto} \quad (2)$$

Our dependent variable, Y_{ipcto} is a candidate's raw test score or percentile rank among applicants to the job. This variable is measured at the level of candidate i who applied for position p in city

c in year t, and for exam o (either the essay or oral test). The key explanatory variable is the interaction between the person being male and the indicator for the oral test, which we interpret as suggestive evidence of gender preference in the assessment of these candidates for the civil service. We include fixed effects at the job level (that is, we include a separate fixed effect for each advertised position) to account for potential sorting across positions. We cluster our standard errors at the city-by-year level.

The differences-in-differences nature of the estimation of the above equation implies that individual and position fixed effects are absorbed if they have the same effects on the blind essay scores and the non-blind oral test scores (Lavy, 2008). On the other hand, evaluators are randomly assigned to candidates of the same job positions in the non-blind oral exam. This suggests that evaluation bias is unlikely to come from the match between evaluators and job positions.

Under the following two assumptions, the comparison of the two assessments should yield suggestive evidence of gender bias. The first assumption is that the stated skills tested in the exam are largely comparable in the blind essay and the non-blind oral tests, as discussed in Section 3.3. The second is that no gender-specific confounding factors systematically affect the blind and non-blind test scores differently. An example of such a confounder would be if women perform worse than men under larger pressure in the oral test than in the written exam (Gneezy, Niederle and Rustichini, 2003). We engage with these assumptions, alternative explanations, and further suggestive evidence in the following empirical analysis.

We show our main estimates in Table 6. While females outperform males in the essay test, where candidate identity is concealed, that advantage reverses when candidates' gender identity is observable in the face interview though candidates' other information is unavailable to evaluators. This holds both for raw test score results and percentile ranks. According to the first column of Table 6, male candidates' oral test scores are 2.77 higher than those of female candidates, completely reversing their disadvantage in essay test scores. The rank results in the third column also show that male candidates have a large advantage over females in the oral test, given the difference in their blind essay test scores. Being male increases candidates' relative percentile

Table 6. Gender preference in the civil service oral test

	Test score		Percentile rank	
	(1)	(2)	(3)	(4)
Male	-2.535***		-0.121***	
	(0.174)		(0.003)	
Oral test	15.420***	15.228***	-0.009***	-0.009***
	(0.663)	(0.531)	(0.002)	(0.001)
Male * Oral test	2.769***		0.051***	
	(0.345)		(0.003)	
Male (ML)		-1.698***		-0.091***
		(0.164)		(0.003)
Male (ML) * Oral test		1.732***		0.046***
		(0.325)		(0.002)
N	267838	540332	266074	536790
R^2	0.744	0.669	0.023	0.019
Machine-learned gender	No	Yes	No	Yes

Table 6 notes: this table presents estimates of α_1, α_2, and α_3 from Equation 2. The sample is limited to those who have both essay and oral test scores. The dependent variable is test score or percentile rank of an individual who took the written essay or oral test in a given year. We control for job position fixed effects, which absorb all job-related variables. We use actual gender information in columns (1) and (3) and machined-learned gender information based on applicants' names in columns (2) and (4). Standard errors are clustered at the city-year level. *** < 0.01, ** < 0.05, * < 0.1.

rank in the non-blind oral test by 5.1%. The estimation based on the machine-learned gender information yields similar results.

5.2 Exploring explanations

In this subsection, we explore different explanations for men's advantages in the non-blind oral test, relative to the blind written exam. First, we examine whether men's relative advantages in the oral test can be attributed to gender differences in skills that are observable to evaluators but not to researchers. Second, we provide suggestive evidence on gender bias against women in the oral test by focusing on those whose oral test scores are narrowly below and above the threshold score.

5.3 Comparability of skills assessed in the essay and oral test

5.3.1 Gender differences in unobservable skills

The main challenge to our arguments for gender bias is that men's relative advantages in the oral test may be due to some unobservable measures of competence used by evaluators of the non-blind oral test, but not reflected in the blind essay test. If the differences in measured skills between the blind and non-blind tests are gender-specific, then our main estimates cannot be interpreted as gender bias in favor of men, but due to these skill differences (Lavy, 2008; Botelho, Madeira and Rangel, 2015; Breda and Ly, 2015; Exley and Nielsen, 2024).[36] We identify a set of skills that may load more on males than females (leadership, handling of unexpected events, and creativity) and, conversely, on females than males (reading comprehension and communication) (Breda and Napp, 2019; Born, Ranehill and Sandberg, 2022). We re-estimate men's relative advantages in the oral test after dropping jobs that specify the need for these skills. We present five estimates, dropping jobs that mention the need for (1) interpersonal communication skills (F), (2) leadership (M), (3) adaptability (M), (4) creativity (M), and (5) all four of these skills. We present these results in Figure 3.

We find that our estimates are all largely unchanged. In Panel (a), we present estimates for test scores using both observations where gender is stated and observations where gender is inferred using the ML method. Our estimates are similar to their analog in the first two columns of Table 6. The estimates using the percentile rank in Panel (b) are also similar to the final two columns of Table 6. As a whole, the results suggest that possible gender differences in skills valued in the oral test cannot explain men's large advantages.

To further capture possible gender differences in cognitive or non-congitive skills valued in the oral test, we complement the Equation 2 by adding individual education level and its interaction with the oral test dummy because education may be closely correlated with both cognitive and non-cognitive ability of individuals (Heckman and Rubinstein, 2001). The individual education level

[36] For instance, men may be better at handling the higher pressure of interpersonal assessments like these interviews (Gneezy, Niederle and Rustichini, 2003).

Figure 3. Exploring explanations: gender differences in specific skills

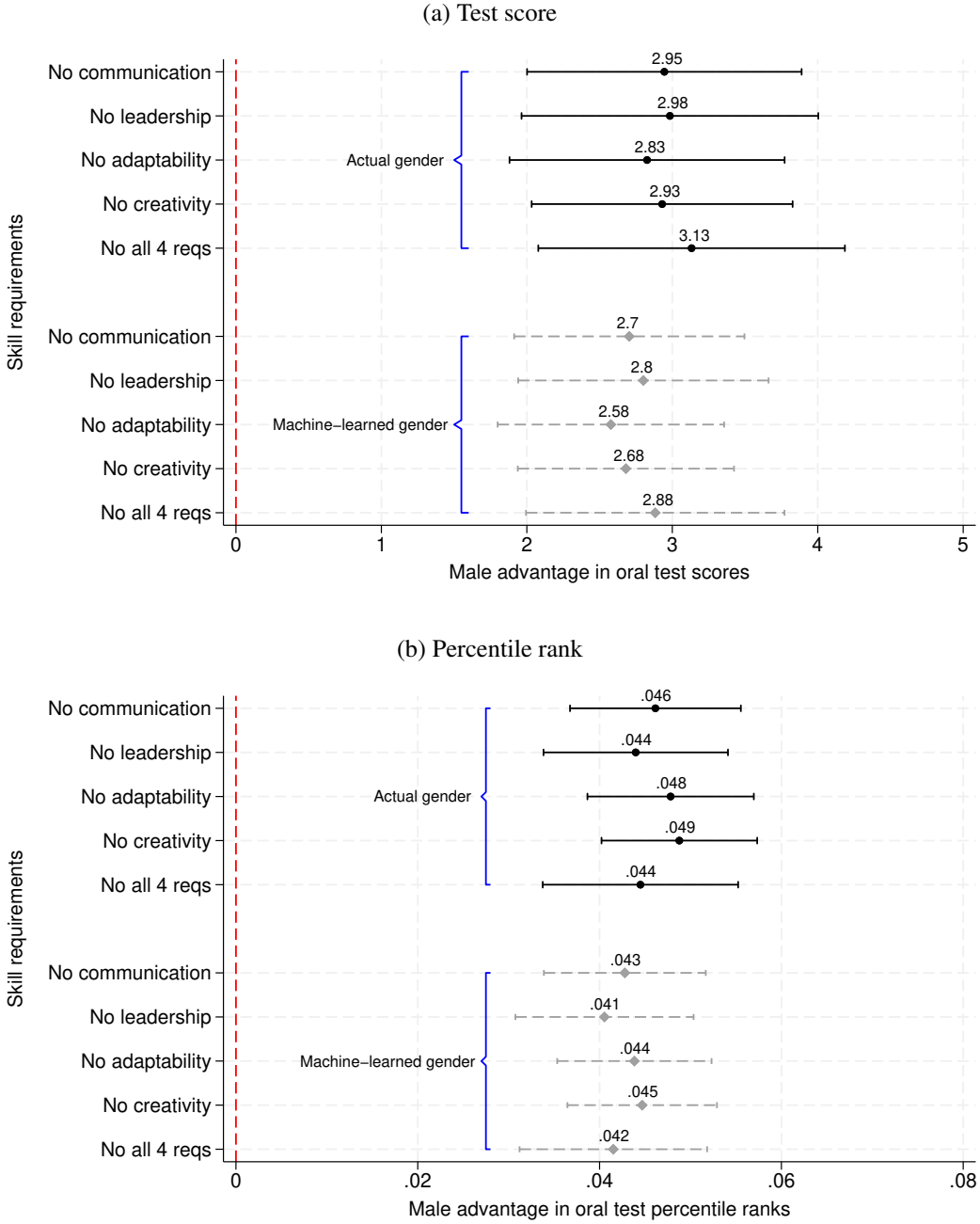


Figure 3 notes: this figure presents estimates of α_3 from Equation 2, along with 95 percent confidence intervals. The dependent variable is one's test score in Panel (a) or percentile rank in Panel (b). We re-estimate men's relative advantages in the oral test after dropping jobs that specify the need for gender-related skills.

Table 7. Exploring explanations: cognitive or non-cognitive skills associated with education

	Test score (1)	Test score (2)	Percentile rank (3)	Percentile rank (4)
Male	-2.876***	-2.551***	-0.123***	-0.108***
	(0.261)	(0.222)	(0.004)	(0.005)
Oral test	20.220***	19.618***	0.003	-0.017
	(1.087)	(0.999)	(0.024)	(0.023)
Male * Oral test	3.352***	3.006***	0.056***	0.051***
	(0.518)	(0.424)	(0.005)	(0.005)
N	133,912	172,858	133,860	172,802
R^2	0.767	0.771	0.026	0.024
Machine-learned gender	No	Yes	No	Yes

Table 7 notes: this table presents estimates of α_1, α_2, and α_3 from Equation 2 by further controlling for individual education level and its interaction with the oral test dummy. The sample is limited to those who have both essay and oral test scores. The dependent variable is test score or percentile rank of an individual who took the written essay or oral test in a given year. We control for job position fixed effects, which absorb all job-related variables. We use actual gender information in columns (1) and (3) and machined-learned gender information based on applicants' names in columns (2) and (4). Standard errors are clustered at the city-year level. *** < 0.01, ** < 0.05, * < 0.1.

includes whether to graduate from an elite university, from an overseas university, or from a junior college. As Table 7 shows, the estimates are similar to our main estimates when limiting the main analysis to the sample without missing values on the education variable. This suggests that men's relative advantages in the oral test may not reflect unobservable gender differences in candidates' attributes that affect differentially the blind and non-blind test performance.[37]

5.3.2 Gender bias in the oral test

In this subsection, we explore whether gender bias is the likely explanation for men's relative advantages in the oral test. We limit the analysis to candidates whose oral test scores are near the threshold on either side for each job position. In our analysis sample, we have only two observations for each job position: a candidate who narrowly passes the threshold score (NP) and a candidate whose score is narrowly below the threshold score (NB). We assume small differences

[37]The main estimates with the restricted sample are shown in Table A4 in the Appendix.

in both cognitive and non-cognitive skills between them.

We characterize the difference in the oral and essay test scores of the candidate who narrowly passes the threshold score (NP) and the whose score is narrowly below the threshold score (NB) in four distinct cases: when the NP is male and the NB is female; when the NP is male and the NB is also male; when the NP is female and the NB is male; and when the NP is female and the NB is also female. We present these results in Figure 4.

This analysis provides suggestive evidence of gender bias in candidate assessment. The comparison group is jobs where a female candidate is selected over another female candidate. In Panel (a), we show the NP-NB difference in essay test scores, and in Panel (b) we show it for oral test scores. In Panel (a), we find that when a female is selected over a male, she needs to have a much better essay score than a male, while a male can have a much lower essay score when he is chosen over a female. In contrast, the essay score difference for a male chosen over another male is nearly identical to that of a female chosen over another female.

In Panel (b) we show analog results for the oral assessment. These document a similar pattern: the largest oral test score gap occurs when a female is selected over a male. In contrast, we find that when a male is selected over a male, this gap is closest to that when a female is selected over a female. We believe that it would be very difficult to explain this pattern only on the basis of gender differences in unobserved characteristics that differentially affect the essay and oral assessments.

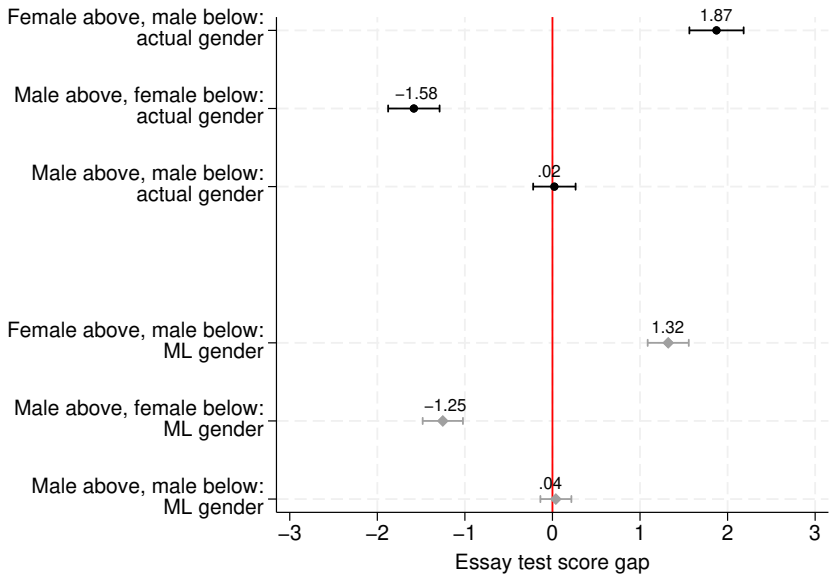
5.3.3 Other explanations

In this subsection, we explore other explanations for men's relative advantages in the face interview: gender differences in self-selection into the interview, social networks, interview preparation, and the likelihood of handwriting being detected.

First, the candidates who attend the face interview may be self-selected in terms of ability among eligible candidates who pass the first-round written exam. If less able men are more likely than women to drop out of the oral exam, then the male advantage in the oral exam may be interpreted as men's better quality, rather than as gender bias in favor of men. To address this concern,

Figure 4. Comparing assessment scores of candidates vs. runners up under different gender pairings

(a) Essay score



(b) Oral test score

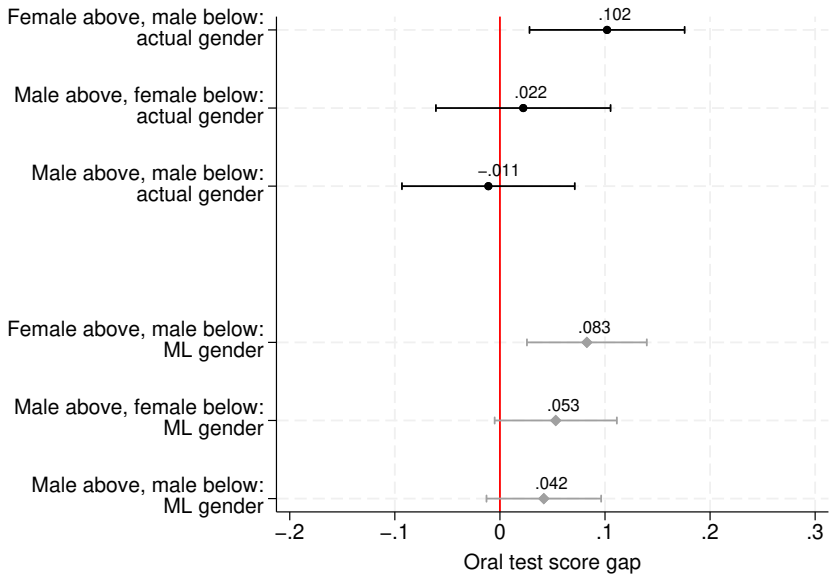


Figure 4 notes: this figure compares gender score gaps between narrowly successful and failed candidates in terms of gender composition. The dependent variable is one's essay score gap in Panel (a) or oral test score gap in Panel (b). We limit the analysis to the candidates whose oral test scores are at the threshold or narrowly below the threshold score for each position.

we drop all candidates of the same positions that have missing values on oral test scores. As Table A5 shows, the results are similar to our main estimates after excluding those who belong to a position where at least one candidate has missing value on the oral test score, suggesting the compositional change concern is minor.

Second, men's relative advantages in the face interview may be due to their better social networks, which help them obtain higher oral test scores. This explanation, however, is unlikely due to the following reasons. On one hand, most evaluators are randomly selected from a provincial-level database of examiners and are randomly assigned to candidates in the oral test. This makes social networks have little impact on candidates' oral test scores. On the other hand, even if social networks may play a role in the civil servant exam (Qin et al., 2024), the social networks impact is unlikely to differ across genders: parents of sons and parents of daughters are unlikely to have different social networks in the civil service sector because parents were permitted to have only one child in the 1990s in China.

Third, men's relative advantages in the face interview may come from male candidates studying harder after the written exam. If men are better prepared for the oral test, we cannot interpret men's relative advantages in the oral test as gender bias in favor of men. This explanation, however, is unlikely in our setting. On one hand, women usually care more about the opportunities of being hired as civil servants and thus are unlikely to invest less time in preparing the oral test than men. On the other hand, the lag between the written and oral tests is only two months, alleviating such possible confounding time trends (Goldin and Rouse, 2000; Breda and Ly, 2015).

Finally, the essay grading may be less "blind" and the gender gap in essay test scores may not reflect the "true" performance gap if examiners can guess candidates' gender based on their handwriting (Lavy, 2008; Ferman and Fontes, 2022). But such an explanation is unlikely in our setting due to the following arguments. First, examiners are required to grade the essay strictly according to detailed scoring criteria. Given the heavy workload in a short period, examiners are unlikely to have time to guess candidates' gender and then grade the essay in favor of women. Second, the success rate of guessing gender based on handwriting would not be high even if examiners

were given time to do so.[38] Examiners, therefore, are uncertain about candidates' gender based on their guesses from handwriting (Breda and Ly, 2015; Breda and Hillion, 2016). Third, because the civil service exam is very important for candidates, both men and women would be careful in their handwriting to avoid bad impressions on examiners and thus there may be no significant gender differences in handwriting. Finally, even if examiners can detect candidates' gender based on handwriting, the pattern in Figure 4 and the men's relative advantage in almost all subgroups are unlikely to be explained by handwriting detection.

5.4 Gender preference in employment decisions

In this subsection, we explore how evaluators' gender preference in the oral test translates into final employment outcomes. According to China's Civil Service Law (Zhonghua Renmin Gongheguo Gongwuyuan Fa), the final employment outcome is based solely on the civil service exam score, which is a weighted sum of both written test and oral test scores.

We implement this analysis using the following estimation equation:

$$Y_{ipct} = \alpha_0 + \alpha_1 * Male_i + \alpha_2 * Essay_{ipct} + \alpha_3 * Male_i * Essay_{ipct} + \lambda_p + \varepsilon_{ipct} \tag{3}$$

Our outcome variable, Y_{ipcto}, is measured at the level of candidate i who has applied for position p in city c in year t. It is an indicator capturing whether the candidate is employed by the civil service. The key explanatory variable is the interaction between gender (Male) and essay score or percentile rank (Essay), which we use to detect possible gender preference in final employment. Because only the rank among candidates applying to a given position determines final employment opportunities, we interact the gender variable with the percentile rank in the essay instead of the raw score. We control for position fixed effects to account for unobservable factors and sorting across positions. Standard errors are clustered at the city-by-year level. We present these results in Table 8.

We find that male candidates have a larger chance of being employed given the same rank

[38] According to an experiment conducted by (Breda and Ly, 2015), the percentage of correct gender guesses is only 68.6%.

Table 8. Gender preference in final employment

	(1)	(2)	(3)	(4)
Male	0.050	0.056***	0.009	0.012***
	(0.042)	(0.024)	(0.007)	(0.005)
Essay score	0.032***	0.032***		
	(0.001)	(0.001)		
Male * Essay score	-0.000	-0.000		
	(0.001)	(0.000)		
Essay percentile rank			0.316***	0.317***
			(0.010)	(0.007)
Male * Essay percentile rank			0.034***	0.026***
			(0.011)	(0.008)
N	132,753	267,829	132,753	267,829
R^2	0.125	0.111	0.117	0.101
Machine-learned gender	No	Yes	No	Yes

Table 8 notes: this table presents estimates of α_1, α_2, and α_3 from Equation 3. The dependent variable is a dummy variable indicating whether an individual is employed finally. We control for job position fixed effects, which absorb all job-related variables. We use actual gender information in columns (1) and (3) and machined-learned gender information based on applicants' names in columns (2) and (4). Standard errors are clustered at the city-year level. *** < 0.01, ** < 0.05, * < 0.1.

in the essay test. This, along with our results in the previous section showing evidence of gender preference in oral assessments, suggests that this gender preference in favor of men has a measurable impact on the gender composition of China's civil servants. According to our preferred specification in the second column, being male on average increases a candidate's employment chance by 4.65% (= 0.036*0.5/0.387).

5.5 Discussion: Statistical or taste-based discrimination?

In this subsection, we discuss whether the gender bias against female candidates in the oral test is due to statistical discrimination or taste-based discrimination.

Statistical discrimination against female candidates may occur if evaluators believe that the non-blind essay score is a less valuable measure of skills for females than for males (Lavy, 2008).

For instance, people may think that women are better than men at preparing written tests but men may perform better than women after being hired. We provide suggestive evidence that is more consistent with taste-based rather than statistical discrimination against female candidates.

First, the large female advantage in the essay scores, where gender identity is concealed, makes it hard to rationalize the relative male advantage in non-blind oral test scores with statistical discrimination. Although evaluators of the oral test are not provided with information on individual candidates' blind test scores, as senior officials or experts, they should have learned from their previous experience that female candidates on average perform much better than males in the blind written exam (Breda and Hillion, 2016).

Second, we divide the analysis sample in terms of whether a job position has the following gender-specific attribute requirements: attributes that biased evaluators may assume to be more male-specific, including work dedication, leadership, independent working, and emergency handling, and attributes that biased evaluators may assume to be more female-specific, including writing, communication, and service orientation.

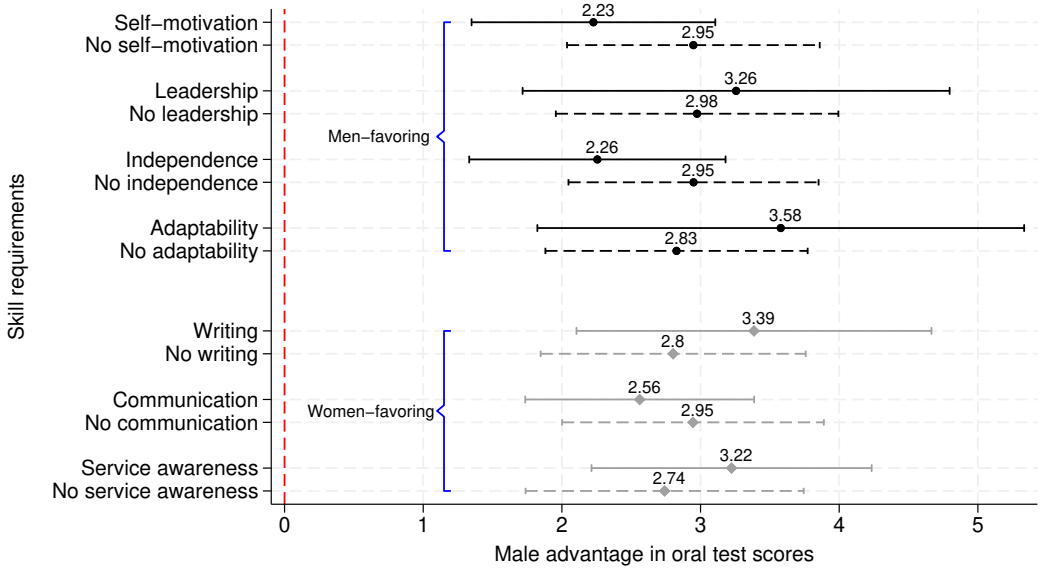
As Figure 5 shows, we find that men have a relative advantage in the oral test in all specifications. Men's relative advantage is even larger when we limit the analysis sample to job positions with women-favoring skill requirements such as writing, communication, and service orientation. The results suggest that the gender bias against women is likely due to evaluators' taste-based discrimination.[39]

Third, we explore the gender gap in test scores among candidates of job positions explicitly specifying gender. We limit the analysis to job positions that have the same names and are from the same employers but explicitly require different genders. Therefore, job positions that explicitly require men are comparable to those that explicitly require women. If statistical discrimination dominates, we expect to find that men perform better than women in the oral test among this group of candidates. Evaluators do not have to discriminate against female candidates at the interview because men and women do not compete for the same job position. As Table 9 shows, however,

[39]The results also suggest that our main estimates are not confounded by possible gender-specific attributes requested by the job position. Figure A3 of the Appendix shows the results with the machine-learned gender information.

Figure 5. Men's advantages in the oral test: gender-specific skill requirements

(a) Test score



(b) Percentile rank

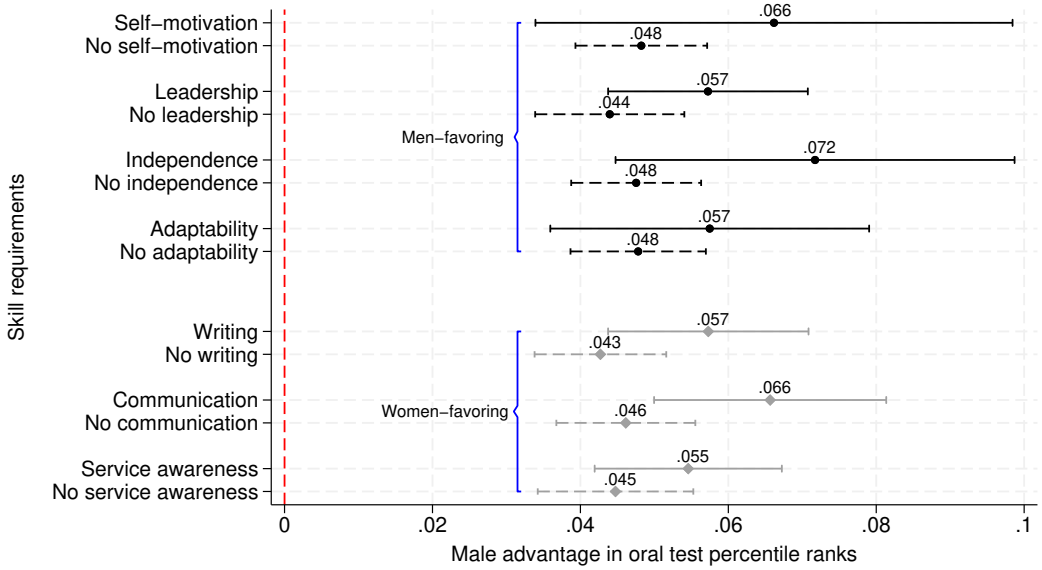


Figure 5 notes: this figure shows results with or without gender-specific skill requirements. The estimates are those of α_3 from Equation 2. The dependent variable is one's test score in Panel (a) or percentile rank in Panel (b).

Table 9. Explicit gender job requirements: test scores and employment

	(1) Admin score	(2) Essay score	(3) Written score	(4) Oral score	(5) Total score	(6) Employed
Male	-2.237***	-3.806***	-3.018***	-1.442***	-2.500***	-0.003
	(0.232)	(0.218)	(0.190)	(0.149)	(0.168)	(0.007)
N	7604	7604	7644	7644	7644	7644
R^2	0.288	0.534	0.482	0.147	0.336	0.053

Table 9 notes: this table shows gender differences in test scores or employment in positions with explicit gender requirements. We regress test scores or a dummy for being employed on a dummy indicating being male. We control for job position fixed effects, which absorb all job-related variables. We use the actual gender information in all columns. The results with machined-learned gender information are A6. Standard errors are clustered at the city-year level. *** < 0.01, ** < 0.05, * < 0.1.

men perform worse than women both in the written exam and in the oral test, suggesting that statistical gender discrimination is unlikely.

Finally, as shown in Subsection 7.2, men's relative advantages in the oral test are smaller when evaluators are less likely to care candidates' gender attribute, suggesting that the gender bias against female candidates in the oral test is due to taste-based discrimination. If men's relative advantages in the oral test stem from evaluators' statistical gender discrimination, we would expect no changes, as the policy should not change evaluators' misperception that men make better future civil servants than women.

6 Promotion

In this section, we study the promotion of candidates who are ultimately selected for the civil service, and how this varies by gender. We use the following two samples that complement each other.

First, we use the civil servants sample from one province located in central China (Hubei Province). We collect promotion information of civil servants from local government websites, which is matched to our civil servant exam dataset. We have successfully obtained promotion information of 586 civil servants recorded in the civil servant exam dataset.

Second, we focus on civil servants in the judicial system on the track to become judges in

China's courts. We match the civil servant exam dataset with an administrative dataset that contains rich information on both work performance and promotion of 0.34 million judges.[40] We have successfully obtained the work performance and promotion history of 5100 court judges, which allows us to explore gender differences in promotion in details.

We will first use the two samples to show gender promotion gap among civil servants. Then we explore why the gender promotion gap exists: we document whether gender differences in work performance and cognitive or non-cognitive skills contribute to the gender promotion gap.

6.1 Gender promotion gap among civil servants

We first study gender promotion gap among civil servants using the Hubei province sample, which contains civil servants of all government departments. As shown by Figure 6, women have higher civil servant exam scores and are more likely to graduate from elite universities, suggesting that women are more qualified than men at hiring. However, women are less likely than men to be promoted to higher ranks (i.e. ranks above the unit level) within ten years after being selected for the civil service.

We next estimate gender differences in promotion within ten years after hiring. We study promotions from newly recruited civil servants to different ranks: vice unit level (fu gu ji), unit level (gu ji), vice section level (fu ke ji), section level (ke ji) or above. These four ranks are low-level positions in China's civil servant system, with vice unit being the lowest level.[41]

We formalize this analysis by estimating the following equation:

$$Y_{ip} = \alpha_0 + \alpha_1 * Male_i + \alpha_2 * X_i + \lambda_p + \varepsilon_{ip} \tag{4}$$

[40] To promote judicial transparency and fairness, China's national-level central court (the Supreme People's Court) has required all local courts to publicize both contemporary and historical verdicts on its official website (China Judgments Online, `https://wenshu.court.gov.cn/`) since 2014. The judgment files contain names and ranks of judges who are responsible for the verdicts, which can be retrieved to comprise our main judge sample. We retrieved 0.34 million judges from the 85 million judgment files between 2013 and 2021 from the official website.

[41] Because our data contain only promotion information within ten years after hiring, newly recruited civil servants are unlikely to be promoted to high ranks such as division level (chu ji). See Subsection 2.1 for details on China's civil servant system.

Figure 6. Gender differences in civil servants' characteristics and promotion

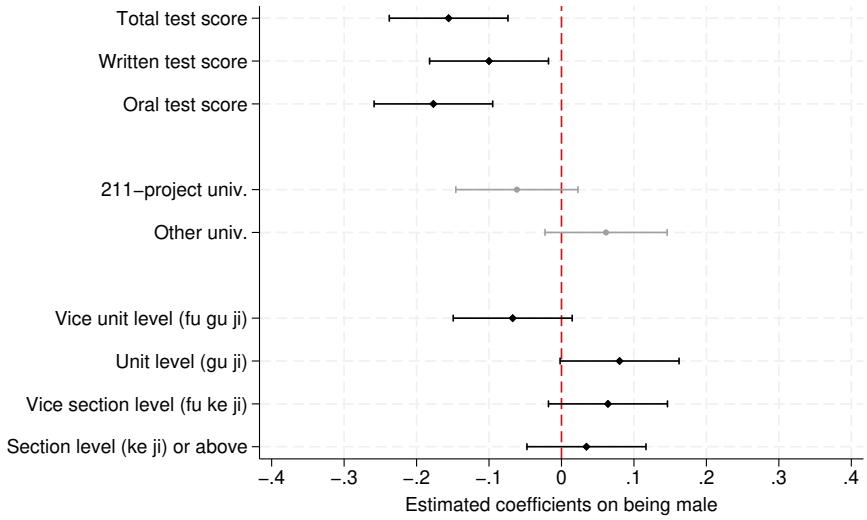


Figure 6 notes: this figure shows gender differences in civil servants' background information and their promotion. The coefficient estimates are scaled by the standard deviation of outcome variables.

Our outcome variables Y_{ip} is a dummy variable for being promoted to vice unit level, unit level, vice section level, or section level or above at the end of year 2023. The outcome variables are measured at the level of the employee i who applied for position p at the time of their civil service exam. We control for position fixed effects (λ_p) to account for any gender difference in position-related unobservable factors. To further alleviate possible concerns about individual-level confounding factors, we control for individual-level civil service test scores and educational background (i.e. a dummy variable for graduating from an elite college) in X_i in later analyses. The standard deviation is clustered at the county level.

We use an event study analysis to explore gender differences in the likelihood of being promoted to different ranks over time. We estimate the following equation:

$$Y_{ipt} = \alpha_o + \sum_{d=1}^{10} \alpha_{1\#d} * Male_i * d_{it} + \rho_p + \lambda_t + \varepsilon_{ipt} \tag{5}$$

Y_{ipt} is a dummy variable indicating whether a person i is a vice-unit-level official, unit-level official, vice-section-level official, or section-level-or-above official in year t. This person applied for

position p at the time of their civil service exam. Period variables p_{it} are the distances between the year in which data was observed and the entry year, with $p = 1$ set to the reference year. The terms ρ_p and λ_t capture position- and year fixed effects, respectively. The standard deviation is clustered at the county level.

We present event-study results in Figure 7, plotting our estimates of $\alpha_{1\#d}$ and their confidence intervals from Equation 5. The figure contains four series: estimates with a dummy variable indicating promotion to vice unit level (black dots), estimates with a dummy variable indicating promotion to unit level (grey triangles), estimates with a dummy variable indicating promotion to vice section level (blue dots), and estimates with a dummy variable indicating promotion to section level or above (light blue triangles).

We find almost no gender differences in promotion to the lowest vice-unit-level rank. We, however, find the gender gap in promotion to the unit-level rank four years after being hired. We further find an increase in the gender gap in promotion to the higher vice-section-level rank six years after being hired, though we find no gender differences in promotion to the section-level rank.[42]

To complement the event study results, we present the overall gender promotion gap estimates in Table 10. As a whole, the estimates are consistent with the event study results: a large gender promotion gap exists among civil servants for relatively higher ranks such as unit level or vice section level and no gender promotion gap exists for the lowest vice-unit-level rank. For instance, women are 45% (=0.028/0.062) less likely than men to be promoted to the unit-level rank.

6.2 Gender promotion gap among judges

In this subsection, we focus exclusively on civil servants in the judicial system by using a larger sample with detailed work performance and promotion history. The analysis with the court judges sample will complement our analysis with the sample of civil servants of all departments in the previous subsection.

[42]This is due to much fewer chances of promotion to the high section-level rank within ten years after being hired. As shown in Figure A4, we have similar results when further controlling for civil servant exam scores and education level.

Figure 7. Gender gap in promotion among civil servants over time

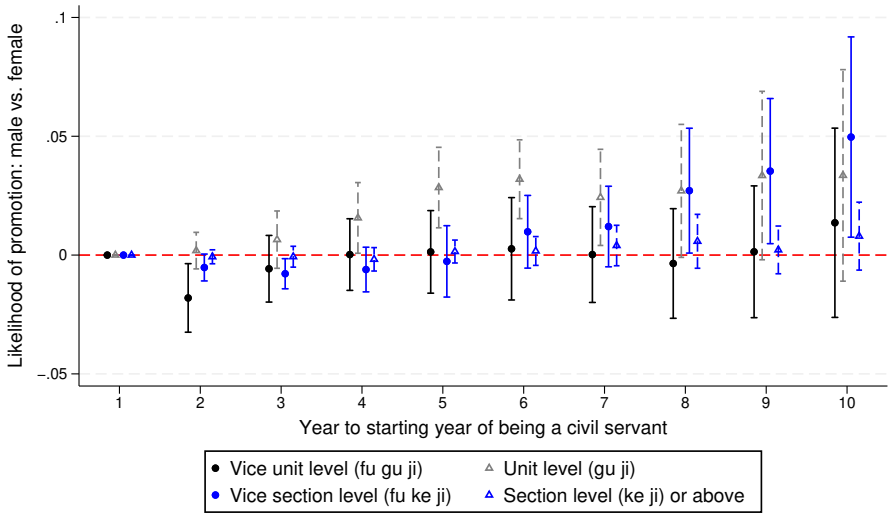


Figure 7 notes: this figure presents estimates of $\alpha_{1\#d}$ from Equation 5, along with 95 percent confidence intervals. The dependent variables are dummy variables indicating whether a person is promoted to a vice-unit-level official, unit-level official, vice-section-level official, or section-level-or-above official. Standard errors are clustered at the county level. We show the tabular analogue to these results in Table 10.

Table 10. Gender gap in promotion among civil servants

	(1) Vice unit	(2) Unit	(3) Vice section	(4) Section or above
Male	-0.014	0.028***	0.016	0.005
	(0.008)	(0.009)	(0.014)	(0.005)
N	2260	2260	2260	2260
R^2	0.231	0.243	0.182	0.209
Outcome mean	.094	.062	.076	.011

Table 10 notes: this table presents estimates of α_1 from Equation 4, the gender gap in promotion to various ranks. The outcome variables are dummy variables for being promoted to vice unit level, unit level, vice section level, or section level or above at the end of year 2023. Standard errors are clustered at the county level. *** < 0.01, ** < 0.05, * < 0.1.

Figure 8. Gender differences in court judges' characteristics, performance, and promotion

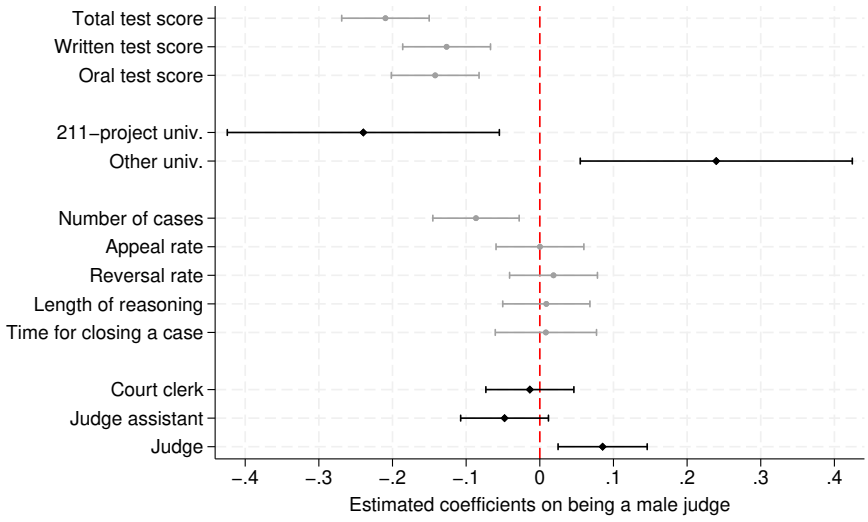


Figure 8 notes: this figure shows gender differences in court judges' background information, work performance, and their promotion. The coefficient estimates are scaled by the standard deviation of outcome variables.

We use our data on candidate assessment to compare the evaluations of male and female employees at time of entry and find a similar gender-specific pattern with the sample of civil servants of all departments. As shown in Figure 8, at time of hiring, women have higher civil service test scores and are more likely to have graduated from an elite university (as defined by the 211-project). Women also complete more court cases and the quality of their judicial decisions is at least as good as that of men's, as measured by the length of reasoning in the judgments, time for closing a case, the appeal and reversal rate of first-instance cases. However, women are less likely than men to be promoted to higher-rank positions and are more likely to work as judge assistants, a position providing support to judges.

Next we explore gender promotion differences within ten years after hiring. We study promotion from the lowest court clerk to judge assistant, and from judge assistant to judge. We use the empirical strategy similar to Equation 5, with the outcome variable replaced with dummy variables indicating whether to be promoted to the rank of judge assistant or whether to be promoted to the

judge rank.[43] The standard deviation is clustered at the city level.

We present event-study results in Figure 9, plotting our estimates of $\alpha_{1\#d}$ and their confidence intervals from Equation 5. The figure contains two series: estimates with a dummy variable indicating promotion to the judge rank (black dots) and estimates with a dummy variable indicating promotion to the judge assistant rank (grey triangles).

We find that men are more likely than women to be promoted from court clerks to judge assistants four years after being hired. We also find a large gender gap in promotion to the judge rank in favor of men starting from the sixth year of the civil service. In contrast, women are more likely than men to be judge assistants starting from the eighth year of the civil service, suggesting a large promotion disadvantage relative to men.[44] Overall, our estimates are consistent with the results with the civil servants of all departments shown in Figure 7.

To complement the event study results, we present the overall gender promotion gap estimates in Table 11. In the first two columns, we present the estimates of the gender gap in promotion from court clerk to the rank of judge assistant and to the judge rank separately. In the third and fourth columns, we present the estimates of the gender gap in number of years taken to be promoted. As a whole, the estimates are consistent with the event study results shown in Figure 9: women are 35% (=0.039/0.112) less likely than men to be promoted to the rank of judge. In contrast, women are more likely than men to be promoted from the rank of court clerk to judge assistant, the second lowest rank in the hierarchy we study. We find no gender differences in time to promotion. The results with court civil servants are also consistent with the estimates with civil servants of all departments shown in Table 10.

[43] A person is defined as a judge if the person is a judge (Shenpan yuan) or a chief judge (Shenpan zhang).

[44] We also redefine ''judges" by including "assistant judges" (Zhuli Shenpanyuan, in Chinese), which can act as the judge role before 2017, the end year of China's Judge Quota Reform. The assistant judge rank was gradually abolished afterwards: assistant judges can apply for the judge position or become judge assistants. (Source: `http://www.npc.gov.cn/zgrdw/npc/xinwen/2019-04/23/content_2086166.htm`.) As Figure A5 shows, the gender promotion gap still exists after including assistant judges as judges.

Figure 9. Gender gap in promotion among court judges over time

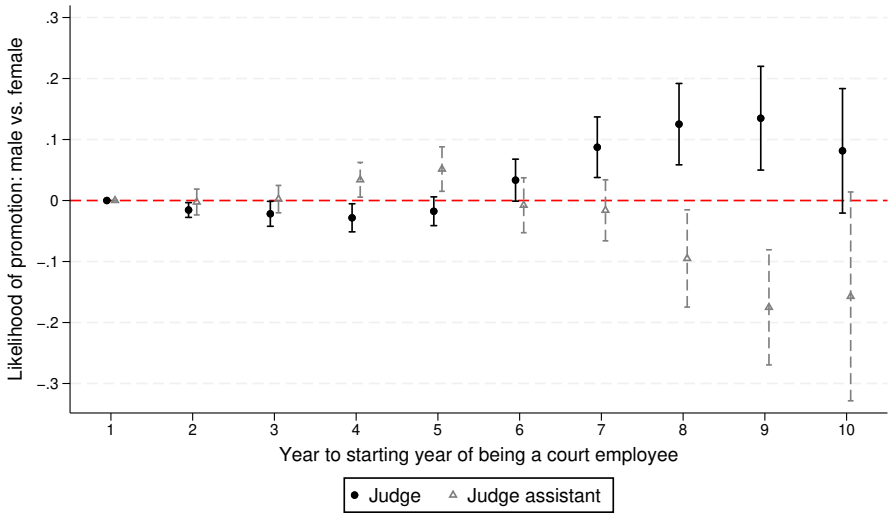


Figure 9 notes: this figure presents estimates of $\alpha_{1\#d}$ from the Equation 5, along with 95 percent confidence intervals. The dependent variables are dummy variables indicating whether a person is a judge assistant or a judge. Standard errors are clustered at the city level. We show the tabular analogue to these results in Table 11.

Table 11. Gender gap in promotion among civil servants

	(1) Promotion to judge asst	(2) Promotion to judge	(3) Time to promotion to judge asst	(4) Time to promotion to judge
Male	-0.022	0.039***	-0.045	0.005
	(0.018)	(0.014)	(0.060)	(0.005)
N	3271	3271	1417	2260
R^2	0.611	0.553	0.875	0.209
Outcome mean	.453	.112	3.536	.011

Table 11 notes: this table presents estimates of α_1 from Equation 4, the gender gap in promotion to various ranks. The outcome variables are dummy variables for being promoted to judge assistant or judge and time to promotion. Standard errors are clustered at the city level. *** < 0.01, ** < 0.05, * < 0.1.

6.3 Why is there a gender promotion gap among civil servants?

The estimates of the previous two subsections both show a large gender promotion gap in favor of men. Why does the gender promotion gap exist? In this subsection, we further explore possible explanations for the gender promotion gap among civil servants. By exploiting detailed work performance information of court judges, we first show whether gender differences in work performance may contribute to the gender promotion gap among civil servants (Azmat and Ferrer, 2017). Second, we explore whether gender differences in cognitive or non-cognitive skills may explain the gender promotion gap by using detailed civil servant exam scores and educational background information. Finally, we discuss other possible explanations such as gender bias in subjective performance evaluations of candidates in promotion decisions (Correll et al., 2020).

6.3.1 Gender differences in work performance

First, we dive deeper into our work performance data, examining gender differences in quantity and quality of output measured at the individual-by-year level. Our performance measures include the number of cases that a court employee completes in a year, the average time for closing a case, the average length of reasoning in the judgments they issue, the proportion of their first-instance cases which are subsequently appealed, and the proportion of their case verdicts which are finally reversed, with the latter four indicators measuring the quality of completed cases. For the four work quality measures, we limit the analysis to the sample of judges, who are mainly responsible for judicial decisions.

We use the empirical strategy similar to the Equation 5, with the outcome variable replaced with work performance measures at the judge-year level. The standard deviation is clustered at the city level. As shown in Figure 10, we find that female employees are more productive than male employees. They complete more court cases, and the quality of their completed cases is at least as good as men's: the length of judicial reasoning is longer for women than for men in judgment files and women are more efficient in handling cases as measured by the time taken to close a case. We find no gender differences in the rate of cases which are appealed or reversed, though these events

are of low frequency.[45]

We further estimate the gender promotion gap conditioning on the above work performance measures. As Figure 11 shows, even after controlling for civil servants' work performance (both work quantity and quality) measures, men are still more likely to be promoted to judges and less likely to be judge assistants six years after being hired. This suggests that gender differences in work performance cannot explain the gender promotion gap among civil servants.

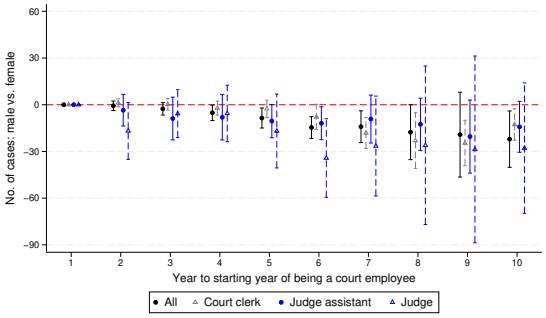
6.3.2 Gender differences in cognitive or non-cognitive skills

In this subsection, we will explore whether the gender promotion gap is due to possible gender differences in cognitive or non-cognitive skills (Kuhn and Weinberger, 2005; Heckman, Stixrud and Urzua, 2006; Azmat and Ferrer, 2017). Our first measure is a dummy variable indicating elite college education, which is associated with both cognitive and non-cognitive skills (Heckman, Stixrud and Urzua, 2006). Our second measure is civil servant exam scores, which are associated with both cognitive and non-cognitive skills because the civil servant exam aims to test candidates' skills such as analytical skill, reasoning, organizational and coordination skills, as shown in Table 1.

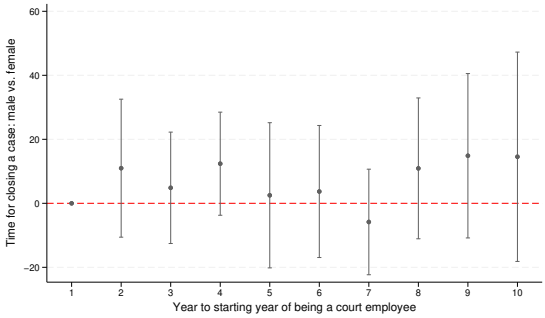
First, we focus on the sample of civil servants of all departments in Hubei Province. Figure 6 has shown that women have higher civil servant exam scores and are more likely to graduate from elite universities. We further control for civil servant exam scores and educational background variables in Equation 5 to explore whether the gender promotion gap is due to gender differences in cognitive and non-cognitive skills measured by civil servant exam scores and elite college education. As Panel (a) of Figure 12 shows, the estimates are similar to our main estimates in Figure 7, suggesting that gender differences in cognitive and non-cognitive skills are unlikely to contribute to the gender promotion gap among civil servants.

[45] One concern is the potential for gender-specific case assignments. If female judges are disproportionately assigned easier cases, the measured performance could be biased against male judges. However, this concern is minimized in our context, as we observe consistent patterns even after 2017, when the Supreme People's Court of China mandated the random assignment of cases to judges. Source: `http://gongbao.court.gov.cn/Details/a4bafde7d2e94a72f12f97204cd1be.html`.

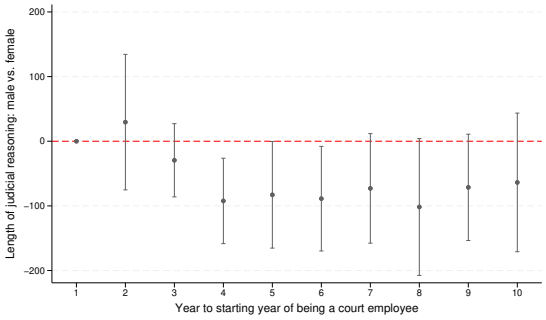
Figure 10. Gender differences in work performance over time



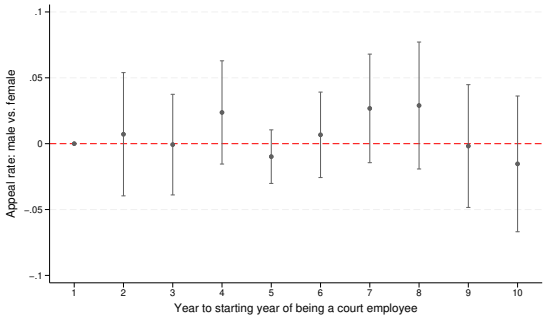
(a) Number of cases



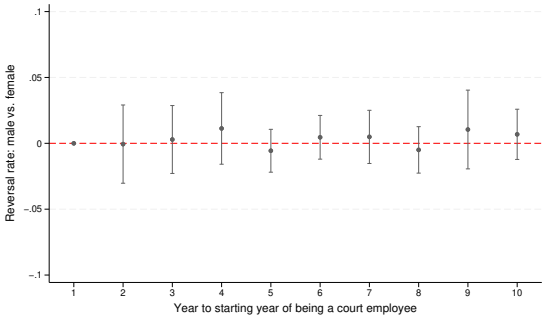
(b) Time for closing a case



(c) Length of judicial reasoning



(d) Appeal rate

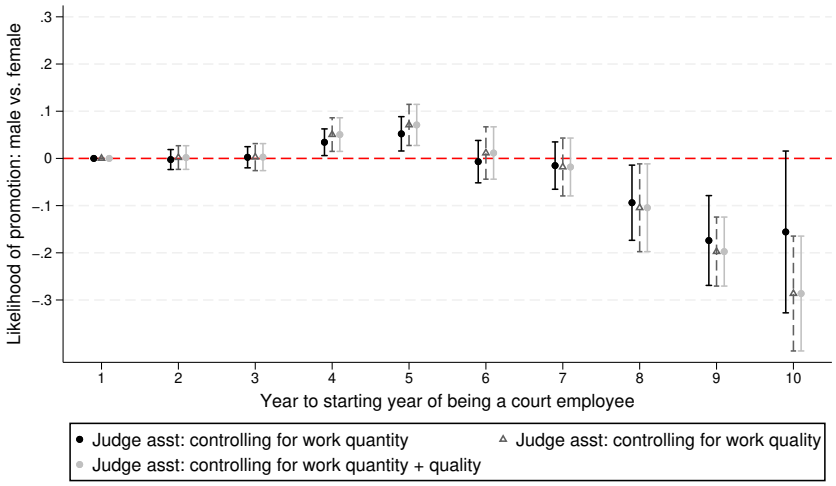


(e) Reversal rate

Figure 10 notes: this figure shows gender differences in work performance over time. The dependent variable is number of cases that a court civil servant attends, time for closing a case, length of judicial reasoning, appeal rate of first-instance cases, or reversal rate of first-instance cases.

Figure 11. Gender promotion gap among civil servants: Accounting for work performance

(a) Promotion to judge assistant



(b) Promotion to judge

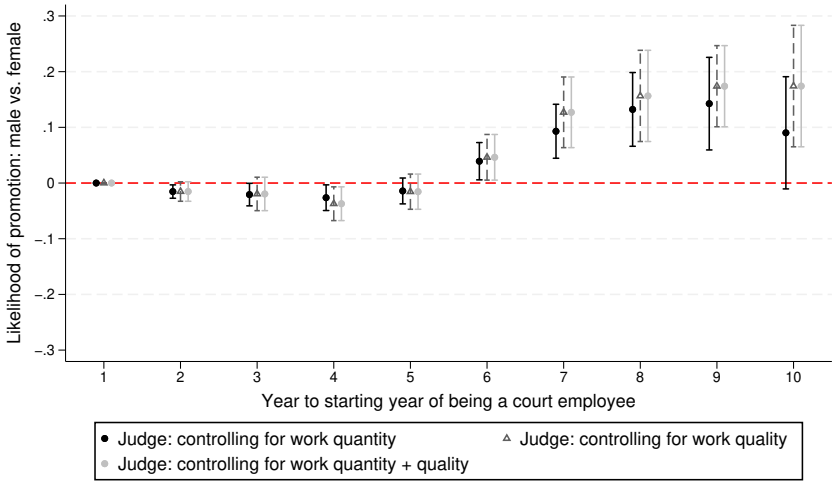


Figure 11 notes: this figure presents estimates of $\alpha_{1\#d}$ from Equation 5, along with 95 percent confidence intervals. The dependent variable is a dummy variable indicating whether a person is promoted to a judge assistant position in a given year in Panel (a) and is a dummy variable indicating whether a person is promoted to a judge position in a given year in Panel (b). Standard errors are clustered at the city level.

Second, we focus on the sample of court judges, who have both civil servant exam scores and detailed work performance history. Similarly, Figure 8 has shown that women have higher civil servant exam scores and better education. As Panel (b) of Figure 12 shows, the estimates are similar to our main estimates in Figure 9 after controlling for both civil servant exam scores and work performance.[46] Overall, the estimates with both samples suggest that gender differences in cognitive and non-cognitive skills are unlikely to contribute to the gender promotion gap among civil servants.

6.3.3 Gender bias

A final explanation for the gender difference in promotion is gender bias in promotion decisions (Régner et al., 2019; Benson, Li and Shue, 2024). The statistical discrimination hypothesis is unlikely in our setting, where supervisors/leaders know candidates' ability and work performance well. Take judge promotions as an example. All application materials including work performance data are publicized within the court system, so evaluators are clear about candidates' ability and performance when making promotion decisions. But according to our main estimates in Figure 9, we still find a large gender promotion gap even after evaluators are informed of candidates' ability and performance, making the statistical discrimination hypothesis unlikely.

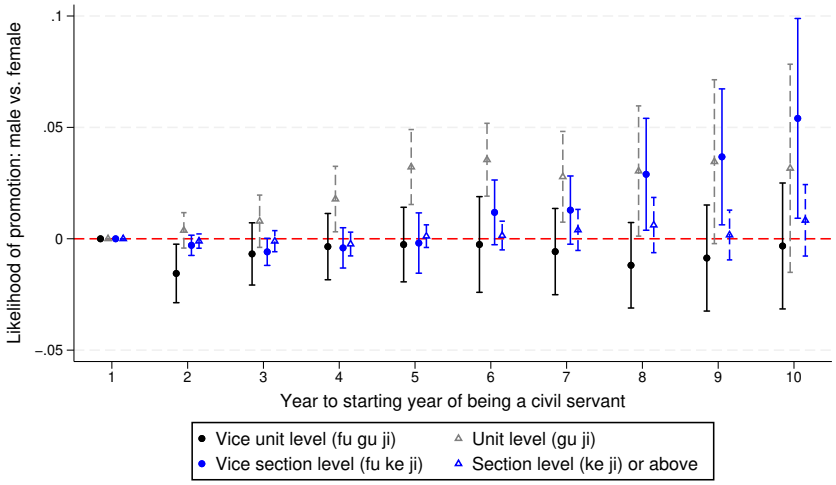
The gender promotion gap among civil servants is likely due to taste-based bias, especially because the promotion of civil servants explicitly requires subjective evaluations (Zhao, Zhao and Yang, 2022). For instance, court judges' promotions depend largely on subjective evaluations of court leaders and colleagues, which usually accounts for more than one quarter of the total assessment score (Zuo, 2020).[47] The minority group such as women are more likely to suffer subjective evaluation bias even conditional on candidates' ability and work performance (Correll

[46] In the court judges sample, only 11% of observations have educational background information. As shown in Figure A6, the gender promotion gap among court judges appears in the seventh years after being hired, which is consistent with our main estimates in Figure 9.

[47] A typical judge promotion decision is based on work performance, the score of a written exam, and subjective evaluations of court leaders and colleagues. The written exam aims to test legal knowledge and ability to apply the knowledge in judicial decisions. Court leaders' evaluation carries a much larger weight in the subjective evaluation. Source: `http://www.zysfy.hbfy.gov.cn/DocManage/ViewDoc?docId=ef8258f0-7f9d-4725-beb4-60f582290258`.

Figure 12. Gender promotion gap among civil servants: Accounting for cognitive and non-cognitive skills

(a) The sample of civil servants of all departments



(b) The sample of court judges

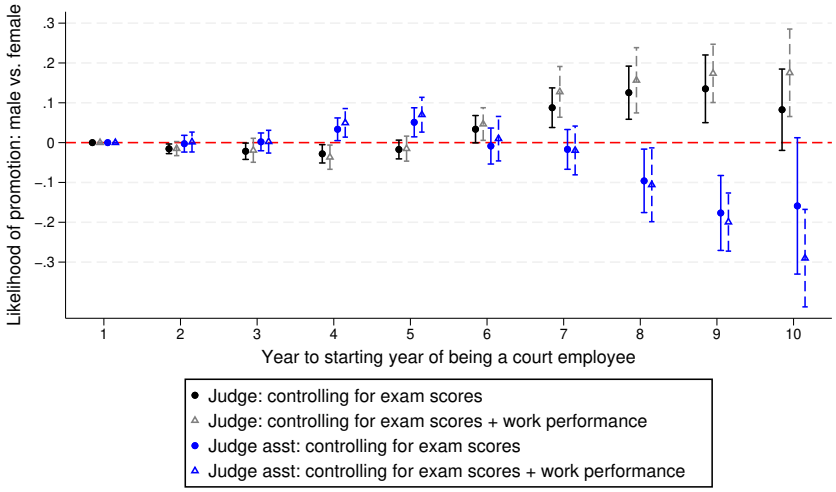


Figure 12 notes: this figure presents estimates of $\alpha_{1\#d}$ from Equation 5, along with 95 percent confidence intervals. The dependent variable is a dummy variable indicating whether a person is promoted to a vice-unit-level official, unit-level official, vice-section-level official, or section-level-or-above official in a given year in Panel (a). The dependent variable is a dummy variable indicating whether a person is promoted to a judge or a judge assistant position in a given year in Panel (b). We control for civil servant exam scores and education in Panel (a) and civil servant exam scores and work performance in Panel (b). Standard errors are clustered at the county level in Panel (a) and at the city level in Panel (b).

et al., 2020).

Overall, we find that female civil servants are less likely than males to be promoted, despite their higher ability and better educational background at the time of hiring, and their better subsequent work performance. Our results in this section suggest that gender preference in promotion decisions is a likely contributor to the difference in career development we observe between male and female civil servants, which in turn is likely to contribute to gender inequality in the ranks of high-ranking officials and future leaders.

7 Policies to mitigate gender bias

In this section we present quasi-experimental analysis of two policies designed to disrupt the forces of gender bias at two key stages where employers influence the process - the content of job advertisements and the evaluation of candidates when identity is known to the assessor.

7.1 Policy targeting explicit gender preference in job advertisements

In July 2016, the All-China Women's Federation issued the *Interim Measures for Women's Federation Organizations to Promote Fair Employment for Women*.[48] This initiative aims to address gender discrimination in employment during the hiring and recruitment processes by establishing an intervening mechanism. Because this regulation has no legally binding force, local governments can choose whether to enforce it. After an extensive search of local government websites, we identified four provinces (Anhui, Heilongjiang, Hubei, and Jilin) that have implemented this regulation, with two of them (Hubei and Jilin) having civil servant exam records in our main dataset.[49]

In Figure 13 we show the share of civil servant job advertisements with explicit and implicit gender preference. In line with the guidance from the policy, in the two provinces that implemented the anti-discrimination regulation we observe an abrupt decrease in the share of job advertisements with explicit gender preference and an increase in the share of job advertisements with implicit

[48] Source: `https://www.nwccw.gov.cn/2022/06/21/99338527.html`.

[49] Other provinces may also have implemented the regulation but could not be identified through our website search, suggesting that our estimates may represent a lower bound.

gender preference after 2016.[50] In contrast, we find no changes in the share of civil servant job advertisements with explicit or implicit gender preference after 2016 in other provinces.

Next, we use an event study analysis to study the impact of the anti-discrimination policy. Because our application and first-round assessment data do not contain the sample from the two provinces that implemented the policy, we study the policy impact on the following two key stages: one, the likelihood of women progressing past the first round and being invited as a finalist to the second round of assessment; two, the likelihood of final employment for these women. We limit our analysis sample to positions whose advertisements contain either explicit or implicit preference for male applicants.

Our analysis estimates the following equation:

$$Y_{ict} = \alpha_o + \sum_{d=-4}^{1} \alpha_{1\#d} * TreatProv_{ic} * d_{it} + \rho_c + \lambda_t + \varepsilon_{ict} \tag{6}$$

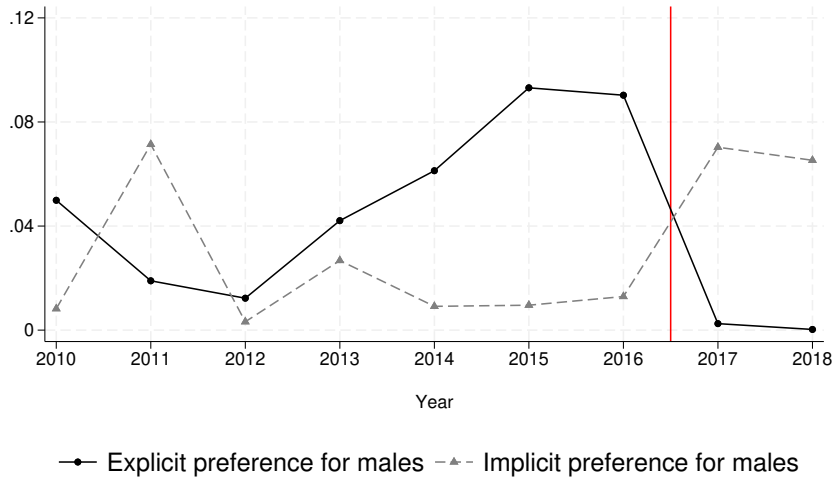
Y_{ict} is male share of candidates associated with a job position i from an employer located in city c and advertised in year t. $TreatProv_{ic}$ is a dummy variable indicating whether the city c of a job position i is from a province that implemented the regulation after 2016. Period variables d_{it} are the distances between the year in which data was observed and the policy year (2017), with $d = 0$ set to the reference year. The terms ρ_c and λ_t capture city- and year-fixed effects, respectively. The standard deviation is clustered at the city level. We present our estimates in Figure 14.

We find that after the policy, women are are roughly twenty percentage points more likely than before to advance to the second-round of assessment and ultimately to fill the role. These results are consistent with Card, Colella and Lalive (2024), which studies the consequence of eliminating explicit gender preferences in job advertisements in Austria. These results suggest that policy prohibiting explicit gender discrimination in job advertisements can be effective in impeding some of the negative forces exerted by employers at this stage of the process, even if its effect is merely to change explicit gender preference in advertisement to implicit preference.

[50] We find no changes in the share of civil servant job advertisements that explicitly request men or women after the policy.

Figure 13. Shares of job advertisements over time: Explicit vs. implicit male preference

(a) Provinces that implemented the regulation



(b) Provinces that did not implement the regulation

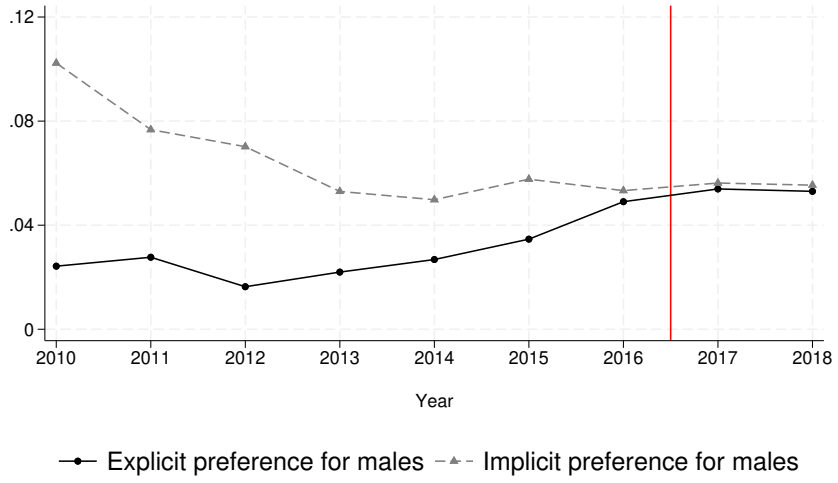


Figure 13 notes: this figure presents the shares of the following two civil servant job advertisements over time: job advertisements that suggest the job is more appropriate for males and explicitly mention "male," and job advertisements that suggest the job is more appropriate for males and do not explicitly mention "male." We show how the two shares of job advertisements change before and after the two provinces implemented an anti-discrimination regulation in Panel (a). We show the trend of the two shares of job advertisements in other provinces in Panel (b).

Figure 14. The anti-discrimination policy effect on women's chances of face interview and final employment

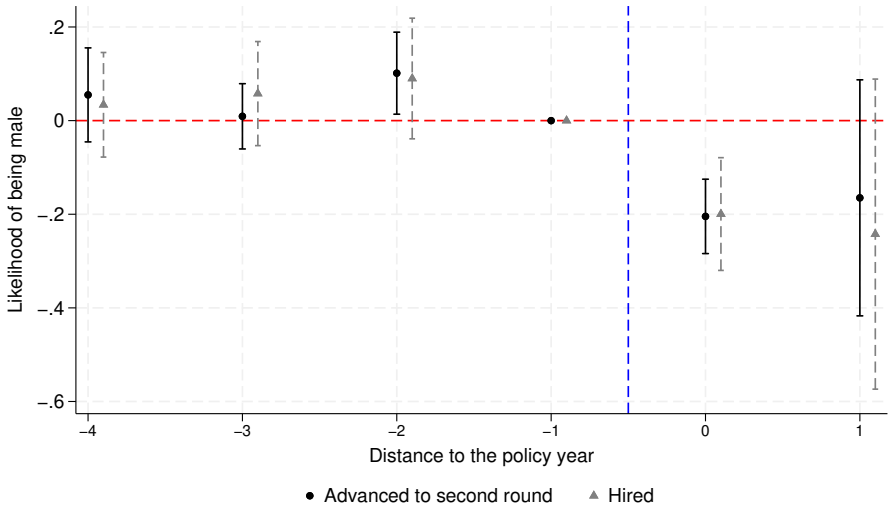


Figure 14 notes: this figure presents estimates of $\alpha_{1\#d}$ from Equation 6, along with 95 percent confidence intervals. The dependent variables are male share of second-round face interview candidates or male share of finally employed people at the job position level. Standard errors are clustered at the city level.

7.2 Policy that severs links between assessors and candidates

How can we mitigate possible gender bias in favor of men in the oral test? In this subsection, we present evidence from a natural experiment which aimed to reduce the connection between evaluators and job candidates.

In China, local city-level civil service administrations are usually responsible for organizing the oral test and selecting evaluators. While the majority of evaluators are randomly selected from a database of outside experts, often one to two examiners are drawn from local senior civil servants related to the job in question (of Human Resources and of Chongqing Municipality), 2009). This linkage allows evaluators from local regions to exert influence on the assessment of candidates. More specifically, if they have preferences over the gender of their future colleagues, this allows for a clear channel for them to reach their objective.[51]

[51] No information on the identity of evaluators is published, which prevents us from exploring how evaluator gender affects assessment of candidate performance by gender.

We evaluate a policy in the Chinese province of Guangdong which aimed to break this linkage. Starting from 2015, Guangdong mandated that all oral tests of the civil service examination be held in a city other than in the city of employers, while other provinces still held oral tests in the city of employers. The goal of the policy was ostensibly to mitigate or prevent problems with nepotism, by reducing the linkage between assessors' choices and their future colleagues, but it also dramatically reduces the incentive for the assessor to exert any gender preference on their assessments, as the assessment no longer determines who they will work with in the future.

We estimate the following model:

$$
\begin{aligned}
(7)\quad Y_{iprto} &= \alpha_0 + \alpha_1 * Male_i + \alpha_2 * Oral_{iprto} + \alpha_3 * Male_i * Oral_{iprto} \\
&+ \alpha_4 * Male_i * (Treat_r * Post_t) + \alpha_5 * Oral_{ipcto} * (Treat_r * Post_t) + \alpha_6 * (Male_i * Oral_{ipcto}) * (Treat_r * Post_t) \\
&+ \alpha_7 * Male_i * Treat_r + \alpha_8 * Oral_{ipcto} * Treat_r + \alpha_9 * (Male_i * Oral_{ipcto}) * Treat_r \\
&+ \alpha_{10} * Male_i * Post_t + \alpha_{11} * Oral_{ipcto} * Post_t + \alpha_{12} * (Male_i * Oral_{ipcto}) * Post_t \\
&+ \lambda_p + \varepsilon_{iprto}
\end{aligned}
$$

Our dependent variable, Y_{iprto} is a candidate's raw test score or percentile rank among applicants to the job. This variable is measured at the level of candidate i who applied for position p in region r in year t, and for exam o (either the oral or essay test). $Oral_{iprto}$ is a dummy variable indicating whether the test is an oral test (as compared to an essay test). $Treat_r$ is a dummy variable indicating whether region r is Guangdong Province which implemented the policy. $Post_t$ is a dummy variable indicating whether year t is after 2015, when Guangdong Province implemented the policy. We include fixed effects at the job level (that is, we include a separate fixed effect for each advertised position, which absorbs both region and year fixed effects) to account for potential sorting across positions. We cluster our standard errors at the region (province) level.

To capture the policy effect on men's relative advantages in the oral test, we interact $Treat_r * Post_t$, $Treat_r$, and $Post_t$ with gender dummy, oral test dummy, and the interaction between gender dummy and oral test dummy, separately. The parameter of our interest is the coefficient α_6, which we in-

terpret as suggestive evidence of how the policy helps mitigate gender bias in the civil servant exam.

Because our main dataset contains only four years (2012/2016/2017/2018) of data on Guangdong Province, we limit our analysis sample to the four years. As shown in Table 12, we control for the interactions between a dummy for Guangdong Province and the three variables (gender dummy, oral test dummy, and their interaction term) in Equation 2 of our main analysis in columns (2) and (5), and the interactions between a dummy for being after the policy and the three variables in columns (3) and (6).

According to the results with our preferred specification in columns (3) and (6), the policy mitigates men's relative advantages in the oral test after reducing the connection between evaluators and job candidates. This analysis suggests a potentially productive means of minimizing the risk of gender preference in oral assessments such as the one we study. By reducing (or, ideally, eliminating) the potential for assessors to directly exert their gender preference on candidates, the male advantage in oral assessments we have measured throughout this analysis thus far either vanishes or reverses.[52]

8 Conclusion

Why are women underrepresented in senior roles across government, industry, and science compared to men? One key explanation is the existence and influence of gender preference by employers. In this paper, we provide evidence from the Chinese civil service, spanning three important stages where gender preference may influence the workforce composition: recruitment, hiring, and promotion. We find evidence of gender preference mainly favoring men at each stage. Nearly a quarter of civil service positions discourage women from applying, and women respond to such gender-specific job requests. During candidate evaluations, women outperform men – until the

[52]Figure A7 show estimates for men's relative advantages in the oral test separately for four subsamples: Guangdong Province before the policy, Guangdong Province after the policy, other provinces before the policy, and other provinces after the policy. The results show a slight increase in the men's relative advantage in both oral test scores and percentile rank in other provinces after the policy. In contrast, we find a decrease in the men's relative advantage in the oral test in Guangdong Province after the policy, which is consistent with the results of Table 12.

Table 12. The non-employer city testing policy effect on men's relative advantages in the oral test

	Test score			Percentile rank		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat * Post * Male	1.343***	-0.145***	0.893	0.039***	-0.044***	-0.017
	(0.456)	(0.000)	(0.749)	(0.008)	(0.000)	(0.014)
Treat * Post * Oral test	-0.173	4.683***	10.919***	0.019***	-0.004***	0.007
	(1.687)	(0.000)	(2.852)	(0.002)	(0.000)	(0.007)
Treat * Post * Male * Oral test	-2.705***	-1.209***	-2.584*	-0.043***	-0.011***	-0.030***
	(0.787)	(0.000)	(1.313)	(0.005)	(0.000)	(0.010)
Male	-1.672***	-1.704***	-0.703	-0.096***	-0.097***	-0.073***
	(0.456)	(0.455)	(0.790)	(0.008)	(0.008)	(0.014)
Oral test	13.359***	13.405***	19.060***	-0.014***	-0.014***	-0.004
	(1.687)	(1.702)	(2.923)	(0.002)	(0.002)	(0.008)
Male * Oral test	1.548*	1.591*	0.204	0.049***	0.049***	0.032***
	(0.787)	(0.790)	(1.496)	(0.005)	(0.005)	(0.011)
N	299458	299458	299458	297258	297258	297258
R^2	0.638	0.639	0.645	0.017	0.017	0.017
Interactions with 'Treat'	No	Yes	Yes	No	Yes	Yes
Interactions with 'Post'						

Table 12 notes: this table presents estimates of α_1, α_2, α_3, α_4, α_5, and α_6 from Equation 7. The dependent variable is a candidate's raw test score or percentile rank. The policy is implemented in 2015 by Guangdong Province to reduce the connection between evaluators and job candidates. We control for job position fixed effects, which absorb all job-related variables. Standard errors are clustered at the province level. *** < 0.01, ** < 0.05, * < 0.1.

evaluator knows the candidate's gender, at which point the bias becomes apparent, leading to a lower likelihood of women being hired. Even after being hired, women continue to outperform men, but are 35-45% less likely to be promoted.

Our study documents the existence and consequences of gender preference within one of the world's largest employers, highlighting the need to counter this bias for fair and equitable employment and career advancement. We demonstrate that countering this bias is not only necessary but feasible. We assess the impact of two policies that minimize the scope for gender preference at different stages of the hiring process. Both policies significantly influence candidate evaluations and ultimately increase the proportion of women hired. While other efforts are also necessary for gender equity, our findings highlight the importance and potentially high returns of these initiatives in promoting equality and ensuring that talent is allocated efficiently.

References

Amer, Abdelrahman, Ashley C Craig, and Clémentine Van Effenterre. 2023. "Does Better Information Reduce Gender Discrimination in the Technology Industry?"

Aneja, Abhay, and Guo Xu. 2022. "The costs of employment segregation: Evidence from the federal government under Woodrow Wilson." Quarterly Journal of Economics, 137(2): 911–958.

Armstrong, Brenna, Tiffany D Barnes, Daina Chiba, and DIANA Z O'BRIEN. 2023. "Financial Crises and the Selection and Survival of Women Finance Ministers." American Political Science Review, 1–19.

Azmat, Ghazala, and Rosa Ferrer. 2017. "Gender gaps in performance: Evidence from young lawyers." Journal of Political Economy, 125(5): 1306–1355.

Bai, Ying, and Ruixue Jia. 2016. "Elite recruitment and political stability: the impact of the abolition of China's civil service exam." Econometrica, 84(2): 677–733.

Benson, Alan, Danielle Li, and Kelly Shue. 2024. ""Potential" and the Gender Promotion Gap." American Economic Review, in press.

Bertrand, Marianne, and Esther Duflo. 2017. "Field experiments on discrimination." Handbook of economic field experiments, 1: 309–393.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination." American Economic Review, 94(4): 991–1013.

Blau, Francine D, and Lawrence M Kahn. 2017. "The gender wage gap: Extent, trends, and explanations." Journal of economic literature, 55(3): 789–865.

Born, Andreas, Eva Ranehill, and Anna Sandberg. 2022. "Gender and willingness to lead: Does the gender composition of teams matter?" Review of Economics and Statistics, 104(2): 259–275.

Botelho, Fernando, Ricardo A Madeira, and Marcos A Rangel. 2015. "Racial discrimination in grading: Evidence from Brazil." American Economic Journal: Applied Economics, 7(4): 37–52.

Breda, Thomas, and Clotilde Napp. 2019. "Girls' comparative advantage in reading can largely explain the gender gap in math-related fields." Proceedings of the National Academy of Sciences, 116(31): 15435–15440.

Breda, Thomas, and Mélina Hillion. 2016. "Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in France." Science, 353(6298): 474–478.

Breda, Thomas, and Son Thierry Ly. 2015. "Professors in core science fields are not always biased against women: Evidence from France." American Economic Journal: Applied Economics, 7(4): 53–75.

Burns, John P. 2007. "Civil service reform in China." OECD Journal on Budgeting, 7(1): 1–25.

Burns, John P, and Wang Xiaoqi. 2010. "Civil service reform in China: Impacts on civil servants' behaviour." China Quarterly, 201: 58–78.

Cahuc, Pierre, Stephane L Carcillo, Andreea Minea, and Marie-Anne Valfort. 2019. "When correspondence studies fail to detect hiring discrimination."

Cao, Guangyu, Chenran Liu, and Li-An Zhou. 2023. "Suing the government under weak rule of law: Evidence from administrative litigation reform in China." Journal of Public Economics, 222: 104895.

Card, David, Fabrizio Colella, and Rafael Lalive. 2024. "Gender preferences in job vacancies and workplace gender diversity." Review of Economic Studies.

Ceci, Stephen J., and Wendy M. Williams. 2011. "Understanding current causes of women's underrepresentation in science." Proceedings of the National Academy of Sciences, 108(8): 3157–3162.

Chen, Minglu. 2022. ""Innocent Young Girls": The Search for Female Provincial Leaders in China." The China Quarterly, 251: 751–775.

Colonnelli, Emanuele, Mounu Prem, and Edoardo Teso. 2020. "Patronage and selection in public sector organizations." American Economic Review, 110(10): 3071–3099.

Corbett, Christianne, Jan G Voelkel, Marianne Cooper, and Robb Willer. 2022. "Pragmatic bias impedes women's access to political leadership." Proceedings of the National Academy of Sciences, 119(6): e2112616119.

Correll, Shelley J, Katherine R Weisshaar, Alison T Wynn, and JoAnne Delfino Wehner. 2020. "Inside the black box of organizational life: The gendered language of performance assessment." American Sociological Review, 85(6): 1022–1050.

Dahis, Ricardo, Laura Schiavon, and Thiago Scot. 2023. "Selecting top bureaucrats: Admission exams and performance in brazil." Review of Economics and Statistics, 1–47.

Estrada, Ricardo. 2019. "Rules versus discretion in public service: Teacher hiring in Mexico." Journal of Labor Economics, 37(2): 545–579.

Exley, Christine L, and Kirby Nielsen. 2024. "The gender gap in confidence: Expected but not accounted for." American Economic Review, 114(3): 851–885.

Ferman, Bruno, and Luiz Felipe Fontes. 2022. "Assessing knowledge or classroom behavior? Evidence of teachers' grading bias." Journal of Public Economics, 216: 104773.

Flory, Jeffrey A, Andreas Leibbrandt, Christina Rott, and Olga Stoddard. 2021. "Increasing workplace diversity: Evidence from a recruiting experiment at a Fortune 500 company." Journal of Human Resources, 56(1): 73–92.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in competitive environments: Gender differences." Quarterly Journal of Economics, 118(3): 1049–1074.

Goldin, Claudia. 2014. "A grand gender convergence: Its last chapter." American Economic Review, 104(4): 1091–1119.

Goldin, Claudia, and Cecilia Rouse. 2000. "Orchestrating impartiality: The impact of "blind" auditions on female musicians." American Economic Review, 90(4): 715–741.

Greenan, Nathalie, Joseph Lanfranchi, Yannick l'Horty, Mathieu Narcy, and Guillaume Pierné. 2019. "Do competitive examinations promote diversity in civil service?" Public Administration Review, 79(3): 370–382.

Heckman, James J, and Yona Rubinstein. 2001. "The importance of noncognitive skills: Lessons from the GED testing program." American Economic Review, 91(2): 145–149.

Heckman, James J, Jora Stixrud, and Sergio Urzua. 2006. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." Journal of Labor economics, 24(3): 411–482.

Hinnerich, Björn Tyrefors, Erik Höglin, and Magnus Johannesson. 2011. "Are boys discriminated in Swedish high schools?" Economics of Education review, 30(4): 682–690.

Hospido, Laura, Luc Laeven, and Ana Lamo. 2022. "The gender promotion gap: evidence from central banking." Review of Economics and Statistics, 104(5): 981–996.

Huang, Junming, Alexander J. Gates, Roberta Sinatra, and Albert-László Barabási. 2020. "Historical comparison of gender inequality in scientific careers across countries and disciplines." Proceedings of the National Academy of Sciences, 117(9): 4609–4616.

Jiang, Xinhui, Sarah Eaton, and Genia Kostka. 2023. "Women's Work: The Gendered Nature of Appointment Politics in Subnational China." Government and Opposition, 1–23.

Kuhn, Peter, and Catherine Weinberger. 2005. "Leadership skills and wages." Journal of Labor Economics, 23(3): 395–436.

Kuhn, Peter, and Kailing Shen. 2013. "Gender discrimination in job ads: Evidence from china." Quarterly Journal of Economics, 128(1): 287–336.

Kuhn, Peter, Kailing Shen, and Shuo Zhang. 2020. "Gender-targeted job ads in the recruitment process: Facts from a Chinese job board." Journal of Development Economics, 147: 102531.

Lavy, Victor. 2008. "Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment." Journal of Public Economics, 92(10-11): 2083–2105.

Li, Feng. 2008. "Annual report readability, current earnings, and earnings persistence." Journal of Accounting and Economics, 45(2-3): 221–247.

Liu, Ernest, Yi Lu, Wenwei Peng, and Shaoda Wang. 2022. "Judicial independence, local protectionism, and economic integration: Evidence from China." National Bureau of Economic Research.

Liu, Hanzhang. 2023. "Meritocracy as Authoritarian Co-Optation: Political Selection and Upward Mobility in China." American Political Science Review, 1–17.

Mangal, Kunal. 2024. "How much are government jobs in developing countries worth?" World Bank Economic Review.

Matsa, David A, and Amalia R Miller. 2011. "Chipping away at the glass ceiling: Gender spillovers in corporate leadership." American Economic Review, 101(3): 635–639.

Mocanu, Tatiana. 2023. "Designing gender equity: Evidence from hiring practices and committees."

Neumark, David. 2018. "Experimental research on labor market discrimination." Journal of Economic Literature, 56(3): 799–866.

of Human Resources, DHRSSCQ (Department, and Social Security of Chongqing Municipality). 2009. "Implementation Rules for Civil Service Oral Exam in Chongqing Municipality." Available at https://gongwuyuan.eol.cn/difang/201103/t20110322_590581.shtml. Accessed on 2014/1/21.

of Human Resources, DHRSSHN (Department, and Social Security of Hainan Province). 2013. "Implementation Rules for Civil Service Oral Exam in Hainan Province." Available at http://hrss.hainan.gov.cn/hrss/0600/201310/71353185ad1741f4953a0da42ece4a5b.shtml. Accessed on 2014/1/21.

of Human Resources, MOHRSS (Ministry, and Social Security). 2015. "Provisional Measures for the Organization and Management of Civil Service Oral Exam." Available at https://www.gd.gov.cn/zwgk/wjk/zcfgk/content/post_2723797.html. Accessed on 2014/1/21.

Qin, Yu, Ya Tan, Geer Ang, and Fan Zhang. 2024. "From connections to merit: Anti-corruption reform and occupational mobility in China's civil service." Economic Journal, Revise and Resubmit.

Régner, Isabelle, Catherine Thinus-Blanc, Agnès Netter, Toni Schmader, and Pascal Huguet. 2019. "Committees with implicit biases promote fewer women when they do not believe gender bias exists." Nature human behaviour, 3(11): 1171–1179.

Wolfers, Justin. 2006. "Diagnosing discrimination: Stock returns and CEO gender." Journal of the European Economic Association, 4(2-3): 531–541.

Xue, Melanie Meng, and Boxiao Zhang. 2023. "The Short-and Long-Run Effects of Affirmative Action: Evidence from Imperial China." Available at SSRN 3997918.

Zhao, Jujun, Zhiyuan Zhao, and Guang Yang. 2022. "Evaluating the Promotion of China's Local Middle-level Cadres: The Role of Professional Résumés." China: An International Journal, 20(4): 23–40.

Zimmerman, Seth D. 2019. "Elite colleges and upward mobility to top jobs and top incomes." American Economic Review, 109(1): 1–47.

Zuo, Weimin. 2020. "Empirical research on the reform of mechanisms for the selection of quota judges: The case of Province A." China Legal Science, 37(4): 261–282.

Appendices

A Appendix

A.1 Appendix Figures and Tables

Figure A1. Word Cloud: Gender Preference in civil service Job Ads



Figure A1 notes: this word cloud is based on the text of job requirements of civil service job ads.

Figure A2. Major Requirements and Gender Preference in Job Ads

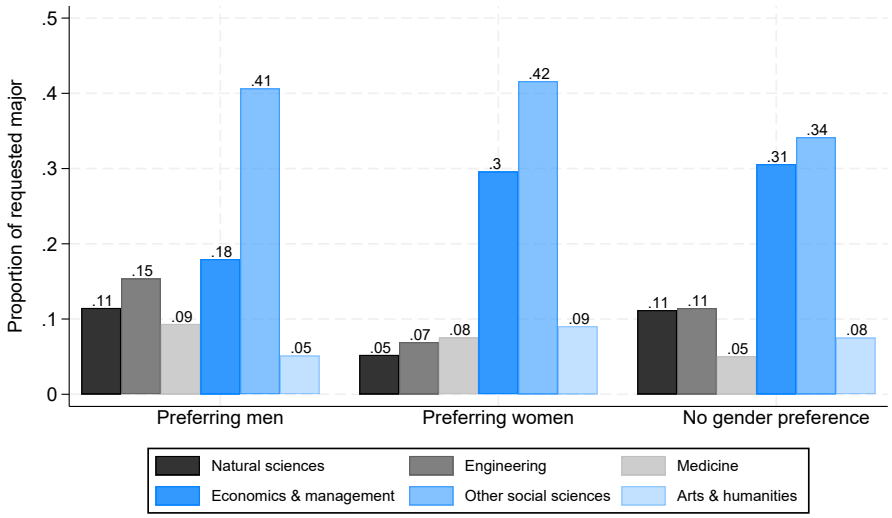
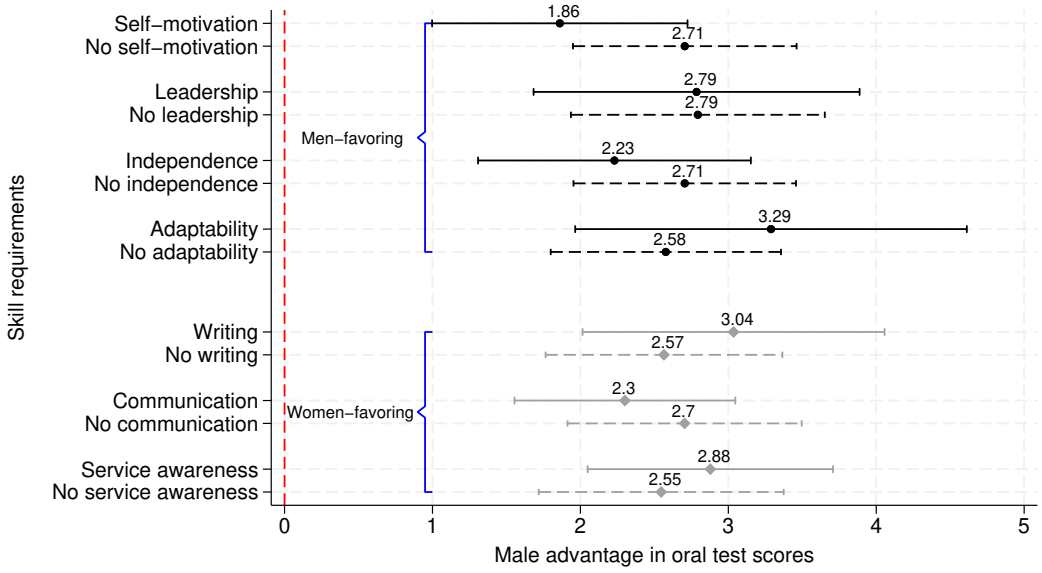


Figure A2 notes: this figure shows proportions of requiring various majors for each gender preference category in civil service job ads.

Figure A3. Men's advantages in the oral test: gender-specific skill requirements

(a) Test score



(b) Percentile rank

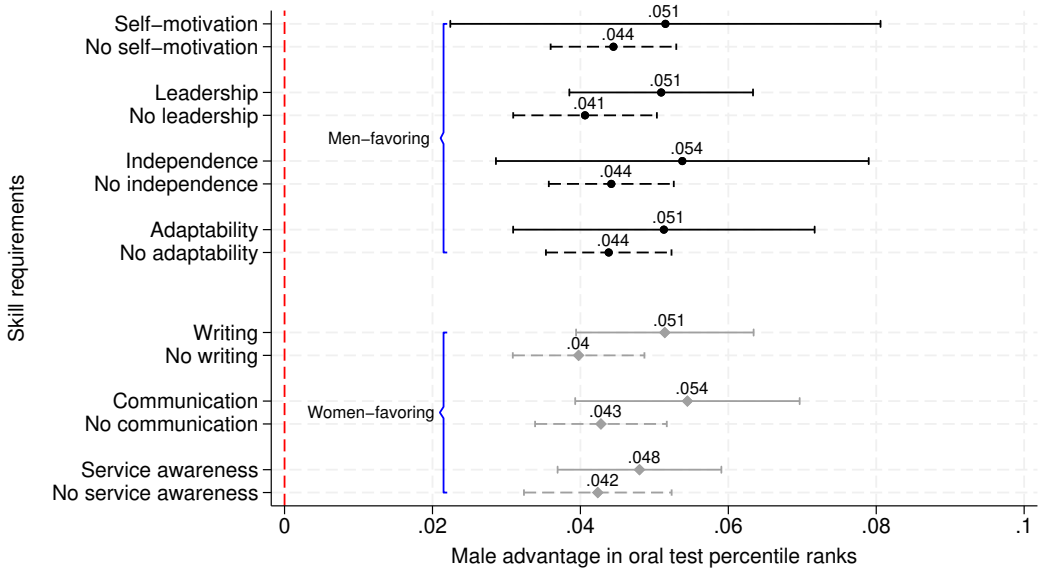


Figure A3 notes: this figure shows results with or without gender-specific skill requirements using the machine-learned gender information. The estimates are those of α_3 from Equation 2. The dependent variable is one's test score in Panel (a) or percentile rank in Panel (b).

Figure A4. Gender gap in promotion among civil servants over time: Controlling for civil servant exam scores and education

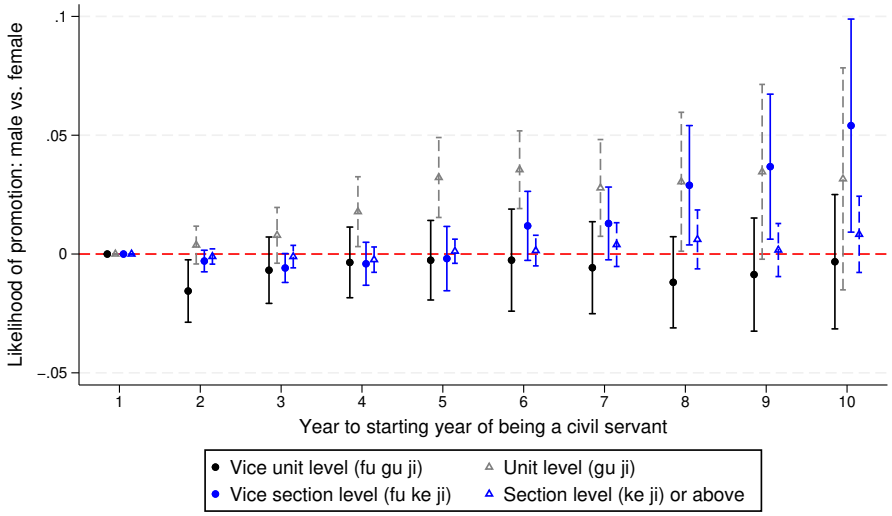


Figure A4 notes: this figure presents estimates of $\alpha_{1\#d}$ from Equation 5, along with 95 percent confidence intervals by further controlling for civil servant exam scores and education level. The dependent variables are dummy variables indicating whether a person is a vice-unit-level official, unit-level official, vice-section-level official, or section-level-or-above official. Standard errors are clustered at the county level.

Figure A5. Gender promotion gap among court judges: redefine "judge"

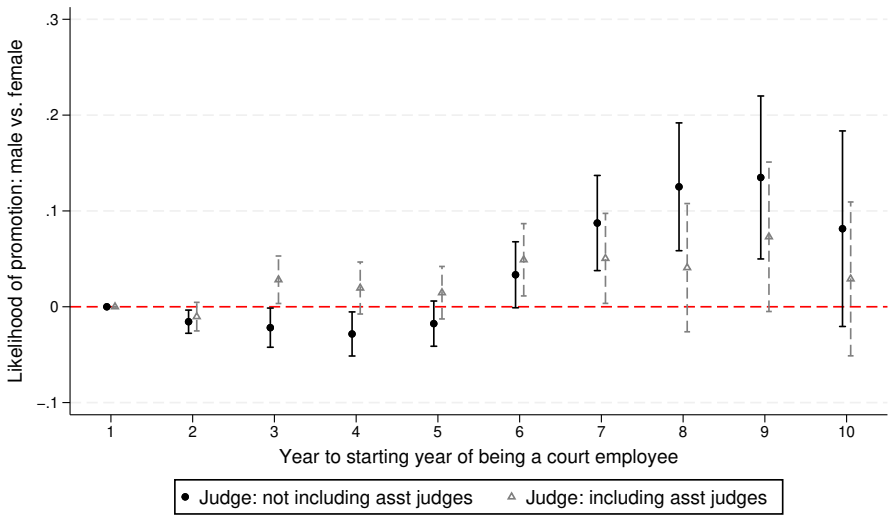


Figure A5 notes: this figure presents estimates of $\alpha_{1\#d}$ from Equation 5, along with 95 percent confidence intervals. The dependent variable is a dummy variable indicating whether a person is a judge. We show estimates with redefining "judges" by also including "assistant judges." Standard errors are clustered at the city level.

Figure A6. Gender promotion gap among court judges: further controlling for education

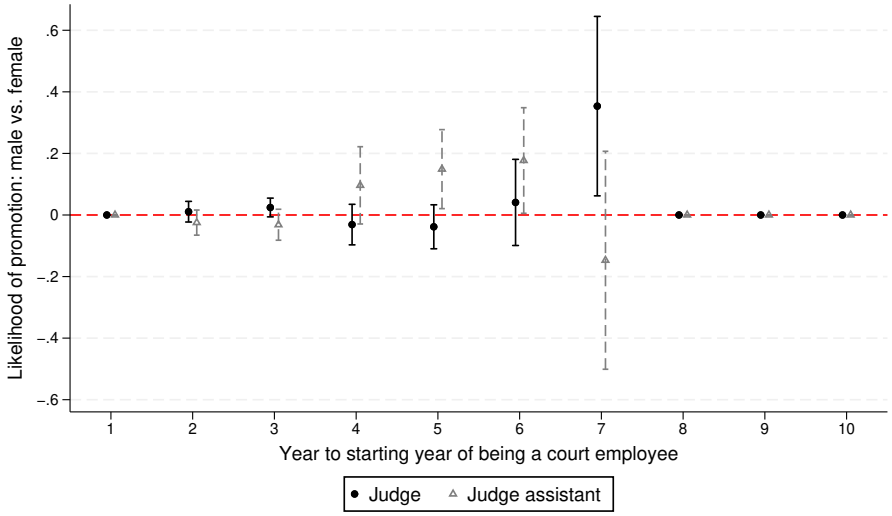
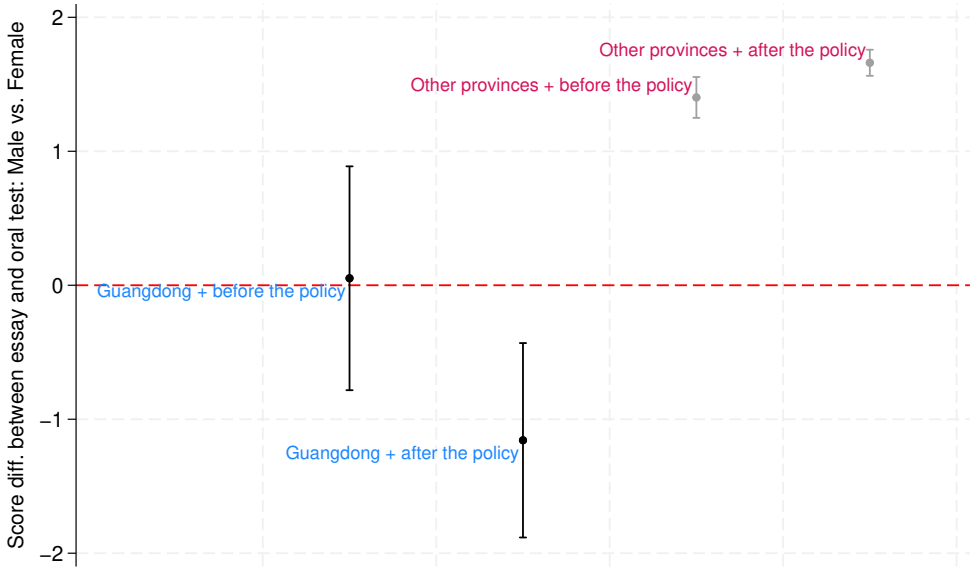


Figure A6 notes: this figure presents estimates of $\alpha_{1\#d}$ from Equation 5, along with 95 percent confidence intervals. The dependent variable is a dummy variable indicating whether a person is a judge assistant or a judge. We show estimates by further controlling for civil servant exam scores, work performance measures, and elite college education. Standard errors are clustered at the city level. Because the judges hired in 2013-2015 do not have education information, we have missing estimates for eight years after hiring.

Figure A7. Reducing evaluators' ties to employers

(a) Test score



(b) Percentile rank

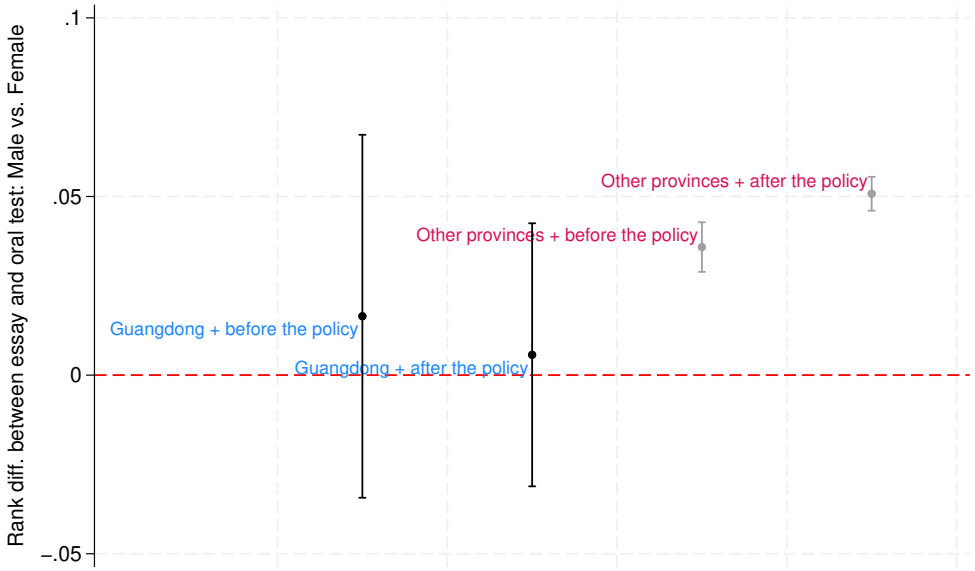


Figure A7 notes: this figure compares gender score gaps across provinces (we divide provinces to reform and non-reform regions). The estimates are those of α_3 from Equation 2. The dependent variable is one's test score in Panel (a) or percentile rank in Panel (b).

Table A1. Applicants' responses and first-round performance: job ads of the same position

	No. of applicants	Prop. of male applicants		Admin score	Essay score	Written score
	(1)	(2)	(3)	(4)	(5)	(6)
Requesting men	-20.952***	1.000***	0.804***	-1.991**	-2.969***	-2.480***
	(3.924)	(0.000)	(0.041)	(0.760)	(0.831)	(0.765)
N	176	35	176	176	176	176
R^2	0.617	1.000	0.952	0.501	0.655	0.588
Machine-learned gender	–	No	Yes	–	–	–

Table A1 notes: this table presents gender differences in applicants' responses and first-round written exam performance for job advertisements of the same position. The dependent variable is number of applicants, proportion of male applicants, applicants' administrative test scores, essay scores, or total written exam scores for a job position in a given year. We control for job-related variables in all columns, including number of slots of the job position, education requirements, major requirements, a dummy variable for CPC (Communist Party of China) membership, job experience, local hukou (household registration status), and cognitive or non-cognitive skill requirements. We use actual gender information in columns (2) and machined-learned gender information based on applicants' names in columns (3). Standard errors are clustered at the city-year level. *** < 0.01, ** < 0.05, * < 0.1.

Table A2. Applicants' second-round performance relating to gender preference in job ads of the same position

	Prop. of male applicants		Written scores	Oral scores	Total scores
	(1)	(2)	(3)	(4)	(5)
Requesting men	0.988***	0.919***	-2.630***	-1.579***	-2.618***
	(0.010)	(0.033)	(0.321)	(0.306)	(0.287)
N	325	382	382	382	382
R^2	0.985	0.946	0.679	0.438	0.613
Machine-learned gender	No	Yes	–	–	–

Table A2 notes: this table presents gender differences in civil service exam performance for job advertisements of the same position. The sample is limited to those who attend the second-round oral test. The dependent variable is proportion of male applicants, written test scores, oral test scores, or total exam scores for a job position in a given year. We control for job-related variables in all columns, including number of slots of the job position, education requirements, major requirements, a dummy variable for CPC (Communist Party of China) membership, job experience, local hukou (household registration status), and cognitive or non-cognitive skill requirements. We use actual gender information in columns (1) and machined-learned gender information based on applicants' names in columns (2). Standard errors are clustered at the city-year level. *** < 0.01, ** < 0.05, * < 0.1.

Table A3. Final employment relating to gender preference in job advertisements of the same position

	Prop. of male candidates		Written scores	Oral scores	Total scores
	(1)	(2)	(3)	(4)	(5)
Requesting men	0.991***	0.923***	-2.495***	-0.891***	-2.154***
	(0.007)	(0.037)	(0.372)	(0.312)	(0.447)
N	283	340	340	340	340
R^2	0.984	0.917	0.584	0.438	0.611
Machine-learned gender					

Table A3 notes: this table presents gender differences in civil service exam performance for job advertisements of the same position. The sample is limited to those who are finally employed through the civil service exam. The dependent variable is proportion of male applicants, written test scores, oral test scores, or total exam scores for a job position in a given year. We control for job-related variables in all columns, including number of slots of the job position, education requirements, major requirements, a dummy variable for CPC (Communist Party of China) membership, job experience, local hukou (household registration status), and cognitive or non-cognitive skill requirements. We use actual gender information in column (1) and machined-learned gender information based on applicants' names in column (2). Standard errors are clustered at the city-year level. *** < 0.01, ** < 0.05, * < 0.1.

Table A4. Gender preference in oral test: limiting to the sample with education information

	Test score		Percentile rank	
	(1)	(2)	(3)	(4)
Male	-2.907***		-0.122***	
	(0.273)		(0.004)	
Oral test	15.080***	15.476***	-0.009***	-0.008***
	(0.970)	(0.851)	(0.003)	(0.003)
Male * Oral test	3.410***		0.054***	
	(0.546)		(0.005)	
Male (ML)		-2.574***		-0.107***
		(0.232)		(0.005)
Male (ML) * Oral test		3.050***		0.050***
		(0.446)		(0.005)
N	133912	172858	133860	172802
R^2	0.767	0.771	0.026	0.024
Machine-learned gender	No	Yes	No	Yes

Table A4 notes: this table presents estimates of α_1, α_2, and α_3 from Equation 2. The sample is limited to those who have essay and oral test scores and education information. The dependent variable is test score or percentile rank of an individual who took the written essay or oral test in a given year. We control for job position fixed effects, which absorb all job-related variables. We use actual gender information in columns (1) and (3) and machined-learned gender information based on applicants' names in columns (2) and (4). Standard errors are clustered at the city-year level. *** < 0.01, ** < 0.05, * < 0.1.

Table A5. Gender preference in the oral test: Dropping candidates of positions with missing values

	Test score		Percentile rank	
	(1)	(2)	(3)	(4)
Male	-2.497***		-0.124***	
	(0.166)		(0.004)	
Oral test	15.414***	15.050***	-0.023***	-0.022***
	(0.691)	(0.552)	(0.002)	(0.001)
Male * Oral test	2.703***		0.050***	
	(0.322)		(0.004)	
Male (ML)		-1.512***		-0.093***
		(0.156)		(0.003)
Male (ML) * Oral test		1.374***		0.045***
		(0.306)		(0.003)
N	204272	421718	202508	418176
R^2	0.743	0.658	0.012	0.008
Machine-learned gender	No	Yes	No	Yes

Table A5 notes: this table presents estimates of α_1, α_2, and α_3 from Equation 2. The sample is limited to those who have essay and oral test scores. We also drop all candidates of the same positions that have missing values on oral test scores. The dependent variable is test score or percentile rank of an individual who took the written essay or oral test in a given year. We control for job position fixed effects, which absorb all job-related variables. We use actual gender information in columns (1) and (3) and machined-learned gender information based on applicants' names in columns (2) and (4). Standard errors are clustered at the city-year level. *** < 0.01, ** < 0.05, * < 0.1.

Table A6. Explicit gender job requirements and gender differences in test scores and employment: using ML gender

	Test score					Employed
	(1)	(2)	(3)	(4)	(5)	(6)
Male	-2.246***	-3.779***	-2.958***	-1.436***	-2.433***	-0.005
	(0.223)	(0.215)	(0.186)	(0.142)	(0.163)	(0.007)
N	7964	7964	8087	8087	8087	8087
R^2	0.285	0.523	0.475	0.161	0.350	0.051

Table A6 notes: this table shows gender differences in test scores or employment in positions with explicit gender requirements by using the machine-learned gender information instead of the actual gender information. We regress test scores or a dummy for being employed on a dummy indicating being male. We control for job position fixed effects, which absorb all job-related variables. Standard errors are clustered at the city-year level. *** < 0.01, ** < 0.05, * < 0.1.

A.2 Factors associated with gender preference in civil servant job advertisements

In this Appendix, we explore what factors are associated with gender preference in civil servant job advertisements using the following regression specification:

$$GenderPref_{ict} = \alpha_0 + \alpha_1 * Edu_{ict} + \alpha_2 * Major_{ict} + \alpha_3 * OthReq_{ict} + \lambda_{ct} + \varepsilon_{ict} \quad \text{(A1)}$$

where $GenderPref_{ict}$ is a dummy variable indicating whether job i from employers located in city c and advertised in year t prefers men (or women).[53] We explore how gender preference in job advertisements is associated with job requirements on education (Edu_{ict}), major ($Major_{ict}$), and a number of other job requirements ($OthReq_{ict}$), including job experience requirement, local hukou (household registration location), CPC (Communist Party of China) membership, number of available places, and skill requirements.[54] We also control for city-year fixed effects (λ_{ct}) in the equation to account for possible factors (such as local gender norms) at the city-year level. The standard deviation is clustered at the city-year level.

As shown in Table A7, we find that job advertisements requesting men demand lower education level and are more likely to demand STEM majors, compared with other job ads. In contrast, job advertisements requesting women are more likely to demand social sciences majors rather than STEM majors, reflecting gender stereotypes in job advertisements.

We further limit the analysis to the job advertisements that separately request men and women but are from the same position of the same employer. As shown in Table A8, we find that job advertisements requesting women are more likely to demand social sciences or humanities majors rather than STEM majors, compared with job advertisements requesting men. We also find that job advertisements requesting women have higher educational requirements and are more likely to require a local hukou, suggesting higher job requirements for women (Kuhn and Shen, 2013).

[53] The dummy variable of jobs preferring men includes jobs explicitly requesting men (not allowing women) and implicitly preferring men (still allowing women) while the indicator for jobs "preferring" women includes jobs explicitly requesting women (not allowing men) and implicitly preferring women (still allowing men).

[54] Our skill requirement measures include a number of dummy variables indicating whether a job requires cognitive skills, computer skills, writing skills, communication skills, self-motivation, self-study ability, hard-working attitudes, good service attitudes, leadership, teamwork, creativity, independence, and adaption.

Table A7. Factors associated with gender preference in job ads

	Preferring men		Preferring women	
	(1)	(2)	(3)	(4)
Education requirements				
Bachelor	-0.035***	-0.028***	-0.002	0.000
	(0.005)	(0.005)	(0.002)	(0.002)
Master or above	-0.068***	-0.056***	-0.015***	-0.010***
	(0.010)	(0.010)	(0.004)	(0.004)
Major requirements				
Enigneering	0.025***	0.024***	-0.003**	-0.004***
	(0.006)	(0.006)	(0.001)	(0.001)
Medicine	0.076***	0.059***	0.012***	0.010***
	(0.009)	(0.008)	(0.002)	(0.002)
Economics and management	-0.086***	-0.082***	0.014***	0.014***
	(0.005)	(0.005)	(0.001)	(0.001)
Other social sciences	0.003	0.001	0.016***	0.014***
	(0.004)	(0.004)	(0.001)	(0.001)
Arts and humanities	-0.070***	-0.051***	0.021***	0.022***
	(0.006)	(0.006)	(0.002)	(0.002)
Others	0.046***	0.038***	0.019***	0.017***
	(0.006)	(0.005)	(0.002)	(0.001)
Observations	862796	861367	862796	861367
R^2	0.158	0.181	0.070	0.074
Other job requirements	No	Yes	No	Yes

Table A7 notes: this table presents estimates of factors associated with gender preference in job ads. The reference group for "Education requirement" is "Junior college or below." The reference group for "Major requirement" is "Natural sciences." We control for other job requirements in columns (2) and (4), including number of slots of the job position, a dummy variable for CPC (Communist Party of China) membership, job experience, local hukou (household registration status), and cognitive or non-cognitive skill requirements. Standard errors are clustered at the city-year level. *** < 0.01, ** < 0.05, * < 0.1.

Table A8. Factors associated with gender preference in job advertisements of the same position

	Preferring men			
	(1)	(2)	(3)	(4)
Education requirements				
Bachelor	-0.025***	-0.015**	-0.057***	-0.040***
	(0.008)	(0.007)	(0.012)	(0.012)
Master or above	-0.075***	-0.048***	-0.080**	-0.055
	(0.015)	(0.014)	(0.038)	(0.039)
Major requirements				
Enigneering	0.029	0.024	0.032	0.017
	(0.022)	(0.022)	(0.029)	(0.028)
Medicine	-0.024	-0.022	0.016	0.024
	(0.023)	(0.023)	(0.031)	(0.030)
Economics and management	-0.167***	-0.166***	-0.215***	-0.222***
	(0.019)	(0.019)	(0.026)	(0.024)
Other social sciences	-0.112***	-0.116***	-0.123***	-0.134***
	(0.020)	(0.020)	(0.026)	(0.027)
Arts and humanities	-0.178***	-0.173***	-0.224***	-0.215***
	(0.022)	(0.021)	(0.029)	(0.027)
Others	-0.090***	-0.114***	-0.096***	-0.139***
	(0.020)	(0.021)	(0.028)	(0.027)
Observations	27583	27563	27573	27545
R^2	0.064	0.082	0.121	0.152
Other job requirements	No	Yes	No	Yes
City-year FEs	Yes	Yes	No	No
Position FEs	No	No	Yes	Yes

Table A8 notes: this table presents estimates of factors associated with gender preference in job advertisements of the same position. The reference group for "Education requirement" is "Junior college or below." The reference group for "Major requirement" is "Natural sciences." We control for other job requirements in columns (2) and (4), including number of slots of the job position, a dummy variable for CPC (Communist Party of China) membership, job experience, local hukou (household registration status), and cognitive or non-cognitive skill requirements. Standard errors are clustered at the city-year level. *** < 0.01, ** < 0.05, * < 0.1.