Do Competitive Examinations Promote Diversity in Civil Service?

Nathalie Greenan, Joseph Lanfranchi, Yannick L'Horty, Mathieu Narcy, Guillaume Pierne

Nathalie Greenan
Conservatoire National des Arts et Métiers, Paris
Joseph Lanfranchi
Université Paris 2, Panthéon-Assas
Yannick L'Horty
Université Paris-Est
Mathieu Narcy
Conservatoire National des Arts et Métiers, Paris
Guillaume Pierné
Université d'Angers

Research Article

Do Competitive Examinations Promote Diversity in Civil Service?

Nathalie Greenan is professor of economics in the Conservatoire National des Arts et Métiers (CNAM) in Paris and a member of the Interdisciplinary Laboratory of Research in Action Science (LIRSA). She leads the research programs on "Organisational Changes, Work and Employment" in CNAM's Centre for Labor and Employment Studies (CEET) and "Policies of Organisations" in the research federation on Work, Employment and Public Policies (TEPP, National Centre for Scientific Research). Her main interests are the study of changes within public and private organizations, their economic performance, and employee outcomes.
E-mail: nathalie.greenan@lecnam.net

Joseph Lanfranchi is associate professor of economics at Université Paris 2, Panthéon-Assas and a member of the Laboratory of mathematical economics and applied microeconomics (LEMMA). His research focuses on economics of personnel and human resource management, with a special interest in the comparison of workplace practices and motivation and performance of employees in the public, nonprofit, and for-profit sectors.
E-mail: joseph.lanfranchi@u-paris2.fr

Yannick L'Horty is professor at Université Paris-Est and director of the research federation on Work, Employment, and Public Policies (National Centre for Scientific Research), which consists of 10 research units and 200 researchers. His main research areas are labor economics, urban economics, and the evaluation of public policies. He is author of 250 publications.
E-mail: yannick.lhorty@u-pem.fr

Mathieu Narcy is associate professor of economics in the Conservatoire National des Arts et Métiers in Paris. He is also member of the Interdisciplinary Laboratory of Research in Action Science (LIRSA) and the Centre for Labor and Employment Studies (CEET). His main research area is labor economics, with a special interest in the study of differences in gender inequalities, prosocial motivation, and incentives between nonprofit, public, and for-profit organizations.
E-mail: mathieu.narcy@lecnam.net

Guillaume Pierné holds a PhD in economics from the University of Paris-Saclay. He is currently a fellow member of the Angevin Research Group in Economics and Management (GRANEM) at the Université d'Angers, Centre for Labor and Employment Studies (CEET), and the research federation on Work, Employment and Public Policies (TEPP, National Centre for Scientific Research). His research focuses on inequalities and discrimination in the labor market, especially in access to employment.
E-mail: guillaume.pierne@univ.angers.fr

Public Administration Review, Vol. 79, Iss. 3, pp. 370–382. © 2019 by The American Society for Public Administration.
DOI: 10.1111/puar.13053.

Abstract: *The representative bureaucracy literature provides a growing body of empirical evidence that a representative public workforce enhances the efficacy and legitimacy of public services. However, little attention has been paid to the capacity of civil service competitive examinations to give equal opportunity of access to public jobs to equally competent citizens. To fill this gap, the authors use French databases to analyze whether competitive examinations comprising both written and oral tests ensure equality of treatment for all candidates regardless of gender, place of residence, or place of birth. The results challenge the capacity of these examinations to treat candidates equally, identifying inequalities in the written tests as well as evaluation biases in the oral tests. However, oral evaluation biases tend to offset inequalities in the written tests. Therefore, selection boards take a sort of affirmative action toward the sole successful members of groups suffering such inequalities.*

Evidence for Practice

- Little is known about the capacity of civil service competitive examinations to provide equal opportunity of access to public jobs for equally competent citizens.
- Although written tests preserve candidates' anonymity, they are not free of adverse impacts against women and candidates coming from deprived areas or born abroad.
- Oral tests (which follow the written ones) offer selection boards the opportunity to partially compensate for inequalities stemming from written tests, with the objective to better integrate all components of society into the civil service.
- The challenge for public managers wanting to promote diversity alongside proficiency is to reduce the scale of inequalities in the written tests while preserving the ability to select the best candidates.

The theory of representative bureaucracy suggests that a public workforce should reflect the composition of the general population in terms of gender, race, ethnicity, and family social class in order to improve the efficacy and legitimacy of public services. Mosher (1982) distinguishes two types of representation: Passive representation refers to the degree to which the members of the bureaucracy mirror the society as a whole. Active representation assumes that public bureaucrats act, consciously or unconsciously, for the interests and desires of those who share their demographic and social background. Hence, active representation helps policy outcomes reflect the interests of all groups represented in society, including minorities. In general, the theory of representative bureaucracy holds that passive representation tends to lead to active representation. Nevertheless, recent research has shown that passive representation, through its symbolic importance, can enhance trust in public organizations and produce a greater willingness on the part of citizens to coproduce, regardless of bureaucratic actions (Riccucci, Van Ryzin, and Li 2016; Theobald and Haider-Markel 2009).

The aim of this article is to analyze the capacity of civil service competitive examinations to provide equal opportunity of access to public jobs for equally competent citizens and to protect against discrimination. In fact, very little attention has been paid to the best system of recruitment to achieve bureaucratic representation, even though a wide range of empirical evidence supports the positive effects of a representative public workforce.

Numerous empirical studies have demonstrated a positive relationship between the proportion of minorities in decision-making roles and policy outcomes compatible with the interests of those minorities, without distinguishing the effects of passive and active representations (Bradbury and Kellough 2011; Nicholson-Crotty et al. 2016). For example, Dee (2005) finds that public school students from minority groups perform better when they are in classrooms

with teachers sharing the same origin. Meier and Nicholson-Crotty (2006) show a positive relationship between the presence of women among police officers and sexual assault reports and arrests. However, as these studies typically use organization-level data, it is not possible to separate the effects of minority bureaucrats' behavior (active representation) from the effects resulting from the willingness of minorities to cooperate and coproduce with bureaucrats coming from the same minority group (passive representation).

On one hand, the active representation of bureaucrats needs to be assessed at the individual level, which only few studies have done so far. One example is the study carried out by Close and Mason (2006), who find that both Latino and African American drivers are better treated by police officers from those minority groups than by white police officers.

On the other hand, a growing empirical literature gives evidence of the positive effects of passive representation, especially for street-level bureaucrats such as police officers. A study carried out by Theobald and Haider-Markel (2009) shows that white and black citizens are more likely to perceive police actions as legitimate if the citizens share the same demographic characteristics as police officers. In the same way, according to Hong (2017), the more ethnically representative a police force is, the lower the number of black citizens' complaints against it. More generally, the greater willingness of citizens to contribute to the production of public services when public organizations are more representative is found in many other areas of public action (see, e.g., Guul 2018; Riccucci, Van Ryzin, and Li 2016).

Moreover, some studies suggest that positive outcomes not directly linked to the primary objective of the public organization can still be generated by bureaucratic representation. For example, Atkins and Wilkins (2013) find that the presence of minority and female teachers helps reduce teen pregnancy rates. Finally, representativeness can also improve public organization performance by providing more inclusive workplaces (Andrews and Ashworth 2015) and less discriminatory recruitment procedures (Meier, Wrinkle, and Polinard 1999).

To reap the benefits of a representative public workforce while recruiting candidates with the skills best suited to public jobs, public managers must ensure that recruitment strategies select the best candidates regardless of their gender, race, ethnicity, or socioeconomic background. In developed countries, the recruitment of public servants is most often centralized and based on performance in open competitive examinations. Historically, competitive examinations have been instituted to prevent patronage and nepotism. For example, in the United States, the adoption of the Pendleton Act of 1883 set up a system of civil service competitive examinations so that government jobs are awarded on basis of merit rather than political connections. In France, although equal access of citizens to public employment on merit alone was affirmed in the Declaration of the Rights of Man and of the Citizen,[1] the general use of open competitive examinations was not put into practice until after World War II.

While some studies stress the heavily bureaucratic system of competitive examinations that would select candidates based on their knowledge or their ability to conform to an academic formalism rather than identify the skills suited to a specific position (Sundell 2014), this article questions the capacity of competitive examinations to treat candidates equally whatever their sociodemographic characteristics. More precisely, we seek to determine whether the system of competitive examinations avoids penalizing some protected groups of candidates on the basis of factors that are irrelevant to the jobs for which they apply. In other words, we analyze the extent to which this recruitment process contributes to the achievement of a representative bureaucracy while also allowing the selection of the best available candidates. However, it is important to specify that competitive examinations free of adverse impacts are a necessary but not sufficient condition for obtaining a representative bureaucracy. The skills required for success on the tests and the desire to become a civil servant also need to be evenly spread out in the population.

In this empirical work, we focus on competitive examinations that comprise a first elimination stage consisting of anonymous written tests and a second stage consisting of oral tests. First, this allows us to assess whether the tests are able to treat candidates equally, depending on whether anonymity is guaranteed. Indeed, unlike the written tests, the oral tests can give rise to evaluation biases toward some groups of candidates. Nevertheless, even though the written tests preserve candidates' anonymity, their object and their content may lead to substantial adverse impacts against some protected groups of candidates. Second, as the oral tests follow the written ones, oral evaluation bias can be evaluated by taking into account candidates' differences in unobservable abilities, as long as the latter are constant across the tests. Finally, this two-stage recruitment procedure allows us to assess the relationship between evaluation biases and inequalities in the written tests. Indeed, selection boards may attempt to compensate for inequalities in the written tests to balance the objective to recruit the best available candidates with the objective to integrate all components of society into the civil service.

The French context is useful for analyzing the capacity of competitive examinations to maintain equal opportunity for four reasons. First, almost all French civil service competitive examinations adopt this two-stage recruitment procedure. Second, we were able to get full access to individual, annual, exhaustive data for 72 competitive examinations over the period from 2008 to 2015. We conducted a separate study for each competitive examination to estimate inequalities in the written tests and oral evaluation biases. The different effect sizes were then combined into an overall effect by carrying out a meta-analysis. Moreover, this large data set allows us to estimate the relationship between oral evaluation biases and inequalities in the written tests through a meta-regression.

Third, the competitive examinations concern seven ministries and thus the recruitment of different types of bureaucrats.[2] More particularly, we can consider the demographic representativeness of street-level bureaucrats such as teachers and police officers. Their representativeness is of particular interest because street-level bureaucrats typically interact with citizens and therefore serve as representatives of the government and its policies.

Fourth, the studies focusing on inequalities of access to public jobs are scarce in the French context. To our knowledge, previous

studies have focused only on gender inequalities in the competitive examinations of two ministries: the Ministry of National Education and the Ministry of Higher Education, Research, and Innovation (Bréda and Hillion 2016; Combes, Linnemer, and Visser 2008). In this study, we consider not only a much larger sample of competitive examinations but also two new group criteria in addition to gender: place of birth and place of residence.

Research Hypotheses

Inequalities in Written Tests

At first glance, anonymous written examinations are an effective way of combining equality and meritocracy. However, these tests are not free of adverse impacts. Bourdieu and Passeron (1964) theorize the risk of favoring the selection of individuals who have specific cultural and social skills. Hence, the formal equality associated with the written tests would not ensure true equality between candidates from different sociocultural backgrounds. The possession of cultural capital and the internalization of an academic culture would facilitate success in the written tests, as upper-class candidates may approach academic exercises calmly and self-confidently and have a better understanding of the standards on which judgments are made.

A vast literature in social psychology has identified the existence of adverse impacts against minority groups members associated with the implementation of cognitive tests in schools or in the workplace (Campbell 1996; Hartigan and Wigdor 1989). For example, the General Aptitude Test Battery, which has been widely used as a measure of cognitive ability in employment selection, provides strong adverse impacts against African Americans (Hartigan and Wigdor 1989). Consequently, when cognitive ability tests are used in recruitment decisions, differences in test scores between minority and majority groups invariably lead to lower job selection rates for minority groups, feeding debates about affirmative action. The causes of these adverse impacts are numerous and partly related to the design of the tests, including their measurement methods, cultural standards of reference, and choice of selection criteria (Hough, Oswald, and Ployhart 2001). Written tests in French competitive examinations rely more on cognitive skills than those in the United States, where the tests are more work related—for instance, using job simulated exercises that minimize adverse impacts against protected groups.

Finally, limitations associated with examination logistics (examinations held in central cities for people living in remote locations or overseas) and more generally with their costs for candidates add up with those associated with the content of the written tests. Candidates who can rely on their family or social networks to organize the time they need to prepare for competitive examinations will benefit. The cost of the competitive examination is also psychological. Indeed, candidates may experience competition differently. Experimental tests have shown that women tend to refuse competitive mechanisms more often than men and are less effective in such contexts (Datta Gupta, Poulsen, and Villeval 2013; Ors, Palomino, and Peyrache 2013). These observations lead us to the first hypothesis in our study:

> **Hypothesis 1:** Written tests are not free of adverse impacts against some groups of candidates.

Sources of Evaluation Biases in Oral Tests

In addition to the possibility of inequalities in the written tests, the non-anonymous oral test stage may give rise to differences in evaluation between candidates once their characteristics become observable. Indeed, lifting anonymity may generate discriminatory evaluation biases in the form of an advantage given to a specific group of candidates independently of their true ability. An evaluation bias can be interpreted as positive discrimination when it favors an allegedly unfairly treated group.

Economic theory gives two main explanations for discriminatory behaviors. The taste-based discrimination theory (Becker 1957) justifies discriminatory behavior by the aversion of decision makers to members of particular groups. As competitive examinations constitute an imperfect information framework, statistical discrimination theory (Arrow 1973) suggests that members of selection boards may also have developed beliefs about the distribution of unobservable determinants of the performance of members of certain groups.

Theories from psychology discuss alternative sources for discriminatory evaluation biases. The theory of in-group favoritism states that evaluators may be tempted to favor members of their own group (Anderson, Fryer, and Holt 2006). Other forms of favoritism might also occur, directly helping a candidate belonging to a professional or family network or favoring those belonging to a group that suffered from previous penalization in order to compensate for those inequalities.

Several examples from the literature on academic evaluations and on recruitment illustrate these forms of discriminatory evaluation biases. For instance, a change in the audition procedures for American symphony orchestras' entrance examinations demonstrated the extent of the discrimination against female candidates (Goldin and Rouse 2000). The introduction of blind auditions explains a substantial part of the increase in the number of female instrumentalists recruited in the 1970s and 1980s. By comparing classroom evaluations against an anonymous national examination, Lavy (2008) shows that in Israeli high schools, boys suffer from the bias of their teachers. A controlled experiment carried out in Germany shows that teachers grade the papers signed by people with surnames of Turkish origin lower than papers signed by people with surnames of German origin (Sprietsma 2013).

The literature on academic evaluations also shows that teachers sometimes give lower grades to pupils when they belong to the same group. In some subjects, Norwegian female high school students achieve their best results under the supervision of male teachers, but not in mathematics (Falch and Naper 2013). Hanna and Linden (2012) show that the lower the caste they belong to, the less well Indian students are rated, with such evaluation bias coming more likely from lower-caste teachers.

Studies of civil service competitive examinations have demonstrated the existence of evaluation biases. Some studies exhibit evaluation biases against women. For example, in the context of recruitment for the Spanish Ministry of Justice between 1987 and 2007, Bagues and Esteve-Volart (2010) show that selection boards are less likely to recruit female candidates when they are composed of more women.

The same type of result was found by Bagues, Labini, and Zinovyeva (2017), who analyze how the gender composition of scientific committees affects the chances of success of female and male candidates in competitive examinations to become full and associate professors in Italy and Spain. The authors provide two interpretations of this evidence. First, female evaluators are not statistically more likely to vote in favor of female candidates than male evaluators. Second, male evaluators become less favorable toward female candidates when women are present on the committee.

There is also empirical evidence of evaluation biases in favor of candidates according to their social background. Bagues and Esteve-Volart (2008) observe a strong hereditary component in the recruitment of Spanish top civil servants, partly because of "favoritism based on kinship" in oral tests. Combes, Linnemer, and Visser (2008) analyze competitive examinations from 1984 to 2003 for the recruitment of French university professors in economics. They show that network effects between the selection board and the candidates favor the latter, to the extent of compensating in some cases for a substantially lower publication record.

Finally, considering that the diversity of the population in the civil service is of major importance in exercising democratic rule, some selection boards may advantage groups of candidates when they are underrepresented in public employment. In the French context, two studies suggest that such compensation strategies occur. Bréda and Ly (2015) analyze the entrance examinations at the École Normale Supérieure,[3] and Bréda and Hillion (2016) the competitive examinations for the recruitment of primary and secondary school teachers. They show that once aware of candidates' identities, selection boards favor female candidates in oral tests in fields in which women are in a minority. Selection boards thus seem to contribute to rebalancing gender representation in these sectors. This literature leads us to formulate the following hypothesis:

> **Hypothesis 2:** Competitive examinations are not free of evaluation biases because of the inclusion of non-anonymous oral tests.

Sources of the Relationship between Inequalities in Written Tests and Evaluation Biases

Evaluation biases may express selection boards' initial preferences or beliefs with respect to some groups of candidates. They may also relate to observed inequalities in the written tests. Indeed, the two-stage design of competitive examinations allows for such a relationship. While in the oral test stage of the examination, the selection board members ignore the grades obtained individually by candidates in the written stage, they may have access to official information about average group performance.[4] Further, information about performance in the written tests of some groups of candidates can be inferred from their relative participation in the oral tests. Finally, during its final deliberations, when the list of successful candidates is established and all of their written and oral grades are available, the selection board can still discuss and revise at the margin the grades of the oral stage.

The theory of representative bureaucracy gives efficiency grounds to the behavior of selection boards that pursue the objective to better represent society in the civil service by compensating in the oral tests for the adverse impacts observed in the written tests. Hence, selection boards, aware of the existence of inequalities in the written tests, may choose to rebalance the outcomes of groups of candidates by biasing the grades given in the oral tests in the opposite direction. As individuals who are unsuccessful in the written tests do not proceed to the oral tests, this bias offsets inequalities at the group level only. In this case, the relationship between inequalities in the written tests and evaluation biases will be negative.

While positive discrimination practices are contrary to the principle of equality, Levade (2004) shows that French law does include differentiations that can be linked to positive discrimination. For example, the French Defender of Rights indicates that it is "possible for an employer, when selecting candidates, to give a compensatory advantage to people belonging to an objectively disadvantaged group because of a prohibited criterion." In this case, the employer must use such advantage to "decide between candidates whose skills have been judged as equal" (Défenseur des droits and CNIL 2012, 19). This new vision of the principle of equality under French law could provide the basis for compensatory practices by selection boards in the oral tests. Moreover, Holzer and Neumark (2000) show that affirmative action programs encourage the integration of minorities with relatively little efficiency loss. These explanations justify the following hypothesis:

> **Hypothesis 3:** Selection boards' evaluation biases may compensate for the inequalities in the written tests and thereby contribute to passive representation in the bureaucracy.

Competitive Examination Data and Econometric Methodology

To test these hypotheses, we need a data set on competitive examinations that fulfills three conditions. First, it must relate to competitive examinations that take place in two stages, with candidate anonymity lifted between the two stages. This setting offers a natural experiment that allows us to approach evaluation biases through the comparison of the anonymous and non-anonymous stages. Second, it must give information about not only the success or failure of candidates at each stage of the competitive examination but also the grades obtained. Third, the data should describe relevant sociodemographic characteristics of the candidates, permitting the analysis of inequality of access, all else being equal.

We gathered data from the French context because we were able to get full access to individual information for all candidates, covering a large set of competitive examinations.[5] Indeed, we processed data for 72 external state civil service competitive examinations for all qualification categories (A, B, and C)[6] and for seven ministries: Agriculture; Economy and Finance; Foreign Affairs; Higher Education, Research, and Innovation; Interior; Labor; and National Education. In all, 58 percent of the candidates applied for positions as secondary school teachers or police officers. Since the seminal work of Lipsky (1980), it is well known that these street-levels bureaucrats act as quasi policy makers in their area of duty. Therefore, the ability of recruitment strategies to achieve representativeness in these professions is essential.

For most examinations, we have yearly data relating to variable periods of observation, generally starting in the late 2000s and ending in 2014 or 2015. Columns 1 and 2 in table 1 list the

Table 1 Description of the Meta-Analysis Data Set

Ministry	Number of Competitive Examinations	Period of Observation	Number of Competitive Sessions	Number of Estimations			Number of Candidates
				Gender	Place of Residence	Place of Birth	
Economy							
G1	6	2007–2015	2	12	12	12	19,695
G2	4	2007–2015	2	8	8	—	7,468
G3	1	2007–2015	13	13	13	13	21,083
G4	1	2007–2016	15	15	15	15	30,149
Total	12	—	32	48	48	40	78,395
Agriculture	3	2000–2016	2	6	6	—	25,381
Foreign Affairs	3	2010–2015	2	6	6	6	13,390
Interior							
G1	1	2011–2014	2	2	2	2	1,636
G2	2	2011–2014	5	10	10	10	50,921
Total	3	—	7	12	12	12	52,557
Labor	2	2009–2015	2	4	4	4	9,123
National Education	45	2008–2015	8	360	360	—	355,745
Higher Education, Research, and Innovation	4	2009–2015	2	8	8	—	29,215
Total	**72**	**—**	**55**	**444**	**444**	**62**	**563,806**

Notes: The number of competitive sessions corresponds to the years or groups of years for which we estimated the effect of each criterion. In the Ministries of Interior and Economy, we divide the competitive examinations into two and four groups, respectively, because they differ in the number of competitive sessions and/or the availability of some membership criteria. The number of competitive sessions exceeds the total number of years when several sessions are organized within the same year (G2, police officers) or, in the case of competition merger, in the period preceding the merger (G3, public finance controller, and G4, public finance inspector). The number of candidates refers to those who completed all the written tests.

number of competitive examinations and periods of observation covered in each ministry. Within each competitive examination, a competitive session is a set of written and oral tests designed to select a given number of candidates. Most of the time, sessions are organized on a yearly basis, but the periodicity also varies with the number of vacant positions. Column 3 in table 1 reports for each examination in each ministry the number of competitive sessions for which we estimate the inequalities and evaluation biases associated with each criterion.[7] The last column presents the number of candidates who took part in all the written tests. In total, more than half a million candidates took part in all the written tests, representing almost 50 percent of the recruitment of civil servants during this period.

We observe three applicant characteristics—gender, place of birth, and place of residence—among the list of 25 criteria of discrimination prohibited under French law. We consider three places of residence: Paris, municipalities with more than 25 percent of the population living in sensitive urban zones (*zone urbaine sensible* or ZUS),[8] and other municipalities. For place of birth, we distinguish candidates according to whether they were or were not born in metropolitan France.[9] Most of the empirical studies devoted to the evaluation of passive representation in Anglo-Saxon countries are concerned with gender, race, and ethnicity criteria. However, under French law, the use of racial and ethnic statistics is forbidden. Nevertheless, places of residence and place of birth reflect the unequal situations of applicants in terms of their economic conditions, social background, and ethnic origin.

The median household disposable income is more than 25 percent higher in Paris than in the country as a whole. In addition, the proportion of executives is two and a half times larger. On the other hand, ZUS residents have experienced degraded economic conditions. In these areas, the poverty rate is 38.4 percent, compared with 13.9 percent for the whole of France.[10] The ZUS areas are also characterized by high concentrations of certain minorities. Among young people recently out of the education system, 29 percent of those of North African origin and 40 percent of those of sub-Saharan African origin live in ZUS areas, compared with 4 percent of young people of French origin.[11] Finally, several studies suggest that people living in these areas suffer from discrimination in the labor market, resulting in less access to employment (Beffy, Coudin, and Rathelot 2014; Bunel, L'Horty, and Petit 2016) and lower wage levels (Couppié, Giret, and Moullet 2010).

The effect of place of birth reflects first of all the influence of immigration status, since more than three-quarters of individuals born outside France are immigrants. Some immigrants face economic and social difficulties (mastery of the French language, schooling, unemployment, etc.). Then, among the populations born outside metropolitan France are those from the overseas departments and territories, who may suffer from discrimination based on their skin color.

Information about gender and place of residence is available in the seven ministries, but information about place of birth is absent in all competitive examinations from three ministries (Agriculture; National Education; and Higher Education, Research, and Innovation) and in some from the Ministry of Economy.[12]

Tables A1 and A2 in the Supporting Information online show that the average success rates in both the written and oral tests differ significantly by gender, place of residence, and place of birth of the candidates. When considering all ministries together, two categories of candidates are less successful in both the written and oral tests: candidates born outside metropolitan France and ZUS residents. The lesser success of candidates born outside metropolitan France in the written and oral tests is observed in all ministries for which this criterion is available. The differences in success rates in the written tests according to place of residence differ little from one ministry to another. In fact, Parisian candidates perform significantly better in all the ministries except Economy, and ZUS candidates are less successful in written tests than candidates living

in other municipalities in four ministries out of seven (Economy, Agriculture, Foreign Affairs, and National Education) but more successful in the Ministry of Interior. The differences in oral test success rates for these candidates are more pronounced. There is no significant difference in success rates according to place of residence in the Ministries of Agriculture, Foreign Affairs, and Labor. Parisian candidates perform better in oral tests in the Ministries of Economy and National Education, while ZUS candidates underperform in the Ministries of Interior and National Education and fare better in the Ministry of Higher Education, Research, and Innovation.

Results for gender are more complex. If women are doing better overall than men in both the written and oral tests, it is far from being a common finding in all ministries. Female candidates have higher success rates in written tests than males in the Ministries of Economy; Interior; and Higher Education, Research, and Innovation and lower ones in the ministries of Foreign Affairs, Labor, and, to a lesser extent, National Education. Regarding significant differences in oral tests success rates, women are more likely to succeed in the Ministry of National Education and less so in the Ministry of Interior.

Differences in success rates between categories of candidates logically lead to a change in the sociodemographic composition of the pool of candidates at each stage of the competitive examination (see table A3 in the Supporting Information online). The most pronounced distortion concerns the proportion of candidates born outside metropolitan France, since they are only 9.4 percent among the successful candidates in the examination, whereas they account for 17.9 percent among the candidates who completed all the written tests. The same logic applies within each of the four ministries where this criterion is available. Furthermore, for each criterion, the written tests lead to a greater alteration in the composition of the pool of applicants than the oral tests.

These success rates do not necessarily reflect unequal treatment of candidates in the examinations. To assess the ability of competitive examination to treat candidates equally, it is necessary to neutralize the differences in the candidates' characteristics that may influence their performance throughout the competitive examinations. Our data set provides information about three sets of such candidates' traits. First, their level of education and age reflect the general human capital they have accumulated. We also observe occupational status at the date of the examination (inactive, unemployed, private worker, public servant),[13] signaling the possible acquisition of the skills required for the job. Finally, the likely influence of household demands on the level of effort devoted to preparation for the examination is approximated by marital status and number of children.

Whereas age and educational level are known in all competitive examinations, the number of children and marital status are present only for the competitive examinations of the Ministries of Interior and Labor and for some of the examinations of the Ministry of Economy. Finally, occupational status is available only for the competitive examinations of the Ministries of Foreign Affairs, National Education; and Higher Education, Research, and Innovation.

To test hypothesis 1 and hypothesis 2, we need to determine whether gender, place of residence, and place of birth affect the candidates' probability of success in the written test stage and the extent of evaluation biases in the oral test stage while controlling for all observable sources of performance. So, we will bring to light whether written and oral tests are likely to select candidates who are representative of the "competent" subpopulation that is composed of equally able candidates in terms of those observable sources of performance. The heterogeneity of the information available between competitive examinations suggests that it is preferable to run a specific regression within each of the 72 competitive examinations. Performing a single regression pooling all the candidates from all the competitive examinations would restrain the control variables to those available in all examinations, leading to less control over the heterogeneity of performance between the candidates.

In addition, when a given candidate's characteristic is available in several competitive examinations, it may be coded differently because each ministry decided on coding practices independently from one another. This is a particular concern for level of education and occupational status.

To test hypothesis 1, for each of the 72 competitive examinations, we estimate a probit model explaining the probability of passing among the candidates who completed all the written tests, controlling for all of their available characteristics. Thus, candidate i succeeds in the written test if his or her average grade N_i^W exceeds the minimum grade required by the selection board $\tilde{N}^W$, which is unobservable by the candidates. Then we denote Y, the observed binary variable corresponding to whether the candidate is successful or not in the written tests:

$$Y_i = \begin{cases} 1 & if\ N_i^W - \tilde{N}^W \geq 0 \\ 0 & if\ N_i^W - \tilde{N}^W < 0 \end{cases} \quad (1)$$

We model the probability of success in written test as follows:

$$Y_i = P\left(\alpha^W + \beta^W C_i + \gamma^W X_i + \varepsilon_i^W > \tilde{N}^W\right), \quad (2)$$

where C_i is a vector of the three group criteria and X_i is a vector including all control variables available in each competitive examination. Finally, ε_i^W is the error term following standard normal distribution with mean zero and standard deviation σ.To test hypothesis 2, we have to identify the presence of discriminatory evaluation biases in the oral test stage. For that purpose, we perform a double difference estimation (see, e.g., Bréda and Ly 2015). Simply evaluating the influence of each criterion on the average grade obtained in the oral test stage is unsatisfactory. Indeed, the average grade obtained by candidate i in the oral tests N_i^O depends on his or her observed characteristics (X_i) but also on an unobserved ability to perform in oral tests (θ_i^O). We can model this grade as follows, considering for the sake of simplicity a single criterion, C_{1i}:

$$N_i^O = \alpha^O + \beta_1^O C_{1i} + \gamma^O X_i + \theta_i^O + \varepsilon_i^O \quad (3)$$

The vector X_i includes the same control variables as in the probit models. If θ_i^O is correlated with the criterion C_{1i}, then β_1^O does not only reflect a discriminatory evaluation bias by the selection board toward candidates belonging to C_{1i}. To control for this potential

correlation, we need to find a way to take into account this unobserved ability. It may be linked to intrinsic characteristics of the candidate such as the capacity to work hard or the ability to manage emotions, but it may also depend on the logistics of the entrance examination (e.g., the distance between the candidate's home and the entrance examination venue). These unobserved characteristics are likely to influence success in both the written and oral tests. Hence, we assume that the unobserved ability to perform in oral tests θ_i^O is correlated with the unobserved ability to perform in the written tests (θ_i^W) and includes a component that is specific to oral tests (ω_i):

$$\theta_i^O = \rho\,\theta_i^W + \omega_i \tag{4}$$

Moreover, we can model the written grade as follows:

$$N_i^W = \alpha^W + \beta_1^W C_{1i} + \gamma^W X_i + \theta_i^W + \varepsilon_i^W \tag{5}$$

It is then possible to estimate the influence of the membership criterion, not on the average grade obtained in the oral test but on the difference between this grade and the average grade obtained in the written test. We can model it as follows:

$$\begin{aligned} N_i^O - N_i^W = {} & \left(\alpha^O - \alpha^W\right) + \left(\beta_1^O - \beta_1^W\right) C_{1i} + \left(\gamma^O - \gamma^W\right) X_i \\ & + (\rho - 1)\,\theta_i^W + \omega_i + \varepsilon_i^O - \varepsilon_i^W \end{aligned} \tag{6}$$

Estimating this grade differential makes it possible to eliminate the effect of the unobserved ability of the candidates affecting both the average grades obtained in the written and oral tests when it is valued in a similar way in the two types of tests ($\rho = 1$). The coefficient $\left(\beta_1^O - \beta_1^W\right)$ then corresponds to a selection board discriminatory evaluation bias toward candidates belonging to C_{1i}, provided that there are no candidates' aptitudes specifically valued in oral tests ($\omega_i = 0$) or, more realistically, that these aptitudes are not correlated with group criterion C_{1i}. This evaluation bias corresponds to the difference in the points obtained between the written and oral tests between the population belonging to the criterion C_{1i} and the reference population. For example, if the estimated coefficient associated with being a woman is positive, that means that female candidates gain (or lose) relatively more (or fewer) points between the written and oral tests than males do.

We organize our econometric strategy in three phases.

Phase 1. We perform 72 independent empirical analyses, each corresponding to a given competitive examination. In every model, we interact our three group criteria with dummies for successive competitive sessions. Hence, we obtain the estimated effects of each criterion on the probability of passing the written test stage and on the difference in average grade between the written and oral tests for each competitive session. We interpret the former as a measure of inequality in the written test stage and the latter as a measure of evaluation bias in the oral test stage.

To the extent that we compare individuals who took part in different competitive sessions as well as grades obtained at different stages, it is possible that comparability problems arise. This will be the case, for example, if the evaluators tended to score candidates in a much more severe and/or homogeneous manner in some sessions compared with others. To solve this problem, we have standardized the average grades obtained in the written and oral tests for each competitive session and examination specialty. This standardization involves converting the average scores obtained by the candidates in such a way that each distribution has a zero mean and a unitary standard deviation. It guarantees comparability across time and between examination specialties.

Phase 2. After phase 1, we have 444 estimated effects of gender and place of residence and 62 effects of place of birth (see column 4 of table 1) on the probability of passing the written tests and on the difference in average grade between the written and oral tests. To assess the validity of hypothesis 1 and hypothesis 2, we must aggregate these effects to improve the reliability of the true effect size. For that purpose, we use a meta-analysis in which each competitive session is treated as a specific experiment. We assume that each estimated effect follows a normal distribution with a mean equal to the true effect size and a variance equal to the sum of the variance of this effect size within and between the competitive sessions. Our estimated effects are therefore sampled from a distribution of effects in the different sessions with a nonzero between-sessions variance. This random-effects meta-analysis computes the aggregate effect size as a weighted mean incorporating an estimate of the between-sessions variance into each session weight, equal to the inverse of the sum of the variance of this effect size within and between the sessions. We use the residual maximum likelihood method to estimate the latter (Patterson and Thompson 1971).

We obtain an aggregate effect size for each criterion. To understand how heterogeneous those effect sizes are according to the characteristics of competitive sessions, we also run meta-regression models controlling for the organizing ministry, the qualification level category, and the log number of recruits. We estimate a random-effects linear regression model for each criterion with as many observations as the number of competitive sessions.[14]

Phase 3. To test the validity of hypothesis 3, we perform a meta-regression model for each criterion, explaining the magnitude of the evaluation bias in the oral test stage by the extent of the inequality of success in the written test stage. We run these regressions with and without controls for the characteristics of the competitive sessions.

Results

Inequalities in the Written Test Stage

The top of table 2 presents the aggregate effect size of each criterion on the inequality of success in the written tests, obtained from the meta-analysis. All things being equal, the three criteria are associated with significant inequalities of success in the written test stage. Consequently, anonymity in written tests does not guarantee an equal probability of success between the different groups of candidates. Thus, the results confirm hypothesis 1. The written test design has adverse impacts on candidates born outside metropolitan France, non-Parisian candidates—particularly those living in municipalities with large ZUS—and, to a lesser extent, female candidates.

More precisely, candidates born outside metropolitan France are the most disadvantaged: their probability of succeeding in the written

Table 2 Inequalities of Success in the Written Tests: Overall Effect and Analysis of Heterogeneity

		Place of Residence		
	Woman	Paris Resident	ZUS Resident	Born Outside Metropolitan France
		Aggregate effect size		
	–0.009*** (2.91)	0.062*** (15.19)	–0.012*** (5.92)	–0.086*** (12.83)
Number of observations	442	431	439	62
		Sources of heterogeneity		
Constant	–0.036* (1.88)	–0.028 (1.51)	–0.027** (2.15)	–0.021 (0.65)
Ministry				
Economy	Ref.	Ref.	Ref.	Ref.
Agriculture	0.006 (0.22)	0.054 (1.59)	0.019 (1.59)	—
Foreign Affairs	–0.081*** (3.28)	0.069*** (3.26)	0.004 (0.31)	0.019 (0.84)
Interior	0.018 (0.93)	0.094*** (4.86)	0.015* (1.87)	–0.001 (0.08)
Labor	–0.014 (0.49)	0.116*** (4.42)	0.022* (1.86)	–0.004 (0.16)
National Education	0.005 (0.38)	0.094*** (8.11)	–0.005 (0.72)	—
Higher Education	0.025 (1.14)	0.047** (2.49)	0.022** (2.49)	—
Category				
B/C	Ref.	Ref.	Ref.	Ref.
A	0.007 (0.54)	0.023** (1.82)	0.010 (1.52)	0.003 (0.22)
Log number of recruits	0.003 (1.07)	–0.001 (0.27)	0.001 (0.50)	–0.015** (2.49)
Adjusted R^2	0.051	0.502	0.303	0.194

Note: Absolute *t*-values are in parentheses.
***Significant at the 1% level;
**at the 5% level;
*at the 10% level.

test stage is 8.6 percentage points lower that that of candidates born in metropolitan France. According to table A4 in the Supporting Information online, they are penalized in 88.71 percent of the competitive sessions, and they never benefit from an advantage.

Parisian candidates benefit from a probability of success in the written tests that is 6.2 percentage points higher than candidates living outside Paris in a municipality with no or small ZUS. They have greater chances of success in one competitive session out of three. Nearly half of the Parisian candidates participated in competitive sessions in which they had an advantage in the written tests. Conversely, the candidates living in municipalities with large ZUS suffer from a disadvantage of 1.2 percentage points, and they are penalized in 10 percent of the competitive sessions.

Finally, the chances of success in the written tests are smaller for women (–0.9 percentage points) than for men. Nevertheless, table A4 in the Supporting Information online reveals that although women are penalized in more than one-fifth of the sessions, they still perform better than men in 12.9 percent of those sessions. Thus, as noted in the descriptive statistics, there is more heterogeneity in effects for gender than for other criteria.

The bottom part of table 2 analyzes some likely sources of the heterogeneity in effect size between competitive sessions. The log number of recruits does not explain differences in inequalities of success in the written tests, except for candidates born outside metropolitan France, whose disadvantage in the written tests is larger when the examination offers more positions. The magnitude of the advantage in the written tests for Parisian candidates is larger in competitive examinations for high-qualification positions (A category). Inequality of success in the written tests also varies with the recruiting ministries for place of residence and for gender. Parisian candidates benefit from a lower advantage in the Ministries of Economy and Agriculture, while the penalty experienced by ZUS residents in the written tests is lower in the Ministries of Interior; Labor; and Higher Education, Research, and Innovation. Finally, the advantage of male candidates is greater in the Ministry of Foreign Affairs than in any other ministry.

The systematic presence of adverse impacts for different groups of candidates, as well as their considerable heterogeneity among ministries, suggests that it would be illuminating to analyze to what extent it could be explained by the variability in the design of the written tests.

Evaluation Biases in the Oral Test Stage

The results presented in table 3 and in table A5 in the Supporting Information globally corroborate hypothesis 2. Indeed, table A5 reveals that, except for ZUS residents, candidates experience significant evaluation biases in more than one-fifth of competitive sessions. However, the aggregate effect size of each criterion on the evaluation biases in the oral test stage presented at the top of table 3 is significant only for women and candidates living in municipalities with large ZUS.

More precisely, on average, selection boards tend to favor the former and penalize the latter. Indeed, competitive sessions in which women enjoy a favorable evaluation bias are 1.8 times more frequent than those in which the evaluation bias is unfavorable. Conversely, the sessions in which candidates living in municipalities with large ZUS suffer from a negative evaluation bias are 1.6 times more frequent than those in which they benefit from a positive bias.

Although it is not significant, the average evaluation bias faced by candidates born outside metropolitan France is positive, and it is the

Table 3 Evaluation Bias in the Oral Tests: Overall Effect and Analysis of Heterogeneity

	Woman	Place of Residence		Born Outside Metropolitan France
		Paris Resident	ZUS Resident	
		Aggregate size effect		
	0.044*** (4.37)	–0.0035 (0.23)	–0.022* (1.85)	0.025 (0.66)
Number of observations	442	431	439	62
		Sources of heterogeneity		
Constant	0.097 (1.51)	0.003 (0.01)	–0.118* (1.65)	–0.190 (1.00)
Ministry				
Economy	Ref.	Ref.	Ref.	Ref.
Agriculture	0.133 (1.35)	–0.731*** (5.40)	–0.159 (1.43)	—
Foreign Affairs	0.043 (0.46)	0.065 (0.57)	0.092 (0.85)	–0.071 (0.44)
Interior	–0.031 (0.45)	0.114 (1.17)	0.097 (1.33)	–0.057 (0.57)
Labor	–0.216** (2.11)	0.132 (0.91)	0.018 (0.15)	–0.081 (0.53)
National Education	0.015 (0.37)	–0.092 (1.64)	–0.142*** (2.94)	—
Higher Education	–0.091 (1.14)	–0.043 (0.39)	0.083 (0.81)	—
Category				
B/C	Ref.	Ref.	Ref.	Ref.
A	–0.128*** (2.62)	0.106 (1.59)	0.112** (2.05)	–0.081 (0.53)
Log number of recruits	0.010 (0.96)	–0.004 (0.26)	0.020* (1.67)	0.058 (1.64)
Adjusted R^2	0.040	0.142	0.052	0.031

Note: Absolute *t*-values are in parentheses.
***Significant at the 1% level;
**at the 5% level;
*at the 10% level.

criterion with the highest share of sessions with positive evaluation bias (in one session out of five). In addition, being Parisian gives rise to evaluation bias in more than one-fifth of competitive sessions. Nevertheless, these biases are sometimes negative and sometimes positive, which explains why the aggregate effect size is not significant.

The bottom of table 3 displays the sources of heterogeneity in evaluation biases. The magnitude of evaluation biases differs slightly from one ministry to another for gender and for the place of residence, whereas it does not depend on the recruiting ministry for place of birth. It is within the Ministry of Labor that women are the least likely to benefit from a favorable evaluation bias. If we refer to table A5 in the Supporting Information, half of the sessions in this ministry exhibit a disadvantage for female candidates. The only ministry in which Parisians suffer from a significant negative evaluation bias is the Ministry of Agriculture. Indeed, according to table A5, it is the only ministry characterized by no evaluation bias in their favor, whereas in half of the sessions, these biases are to their disadvantage. ZUS residents are more likely to suffer from negative evaluation biases in the Ministry of National Education compared with other ministries.

Evaluation biases are less favorable for women in high-qualification competitive exams (A category) than in lower-qualification ones (B and C categories). In contrast, evaluation biases are more favorable for candidates living in municipalities with large ZUS in high-qualification examinations and in examinations for which the number of offered positions is high.

At this stage of the analysis, two results stand out. On the one hand, there are very high inequalities of success in the written tests. On the other hand, the oral tests are far from being free of evaluation biases. A question now arises: do evaluation biases exist to partially offset inequalities of success in the written tests? In other words, do selection boards tend to overevaluate (resp. underevaluate) groups of candidates who have been less (resp. more) successful in the written tests to make the French civil service more representative? It is relevant to emphasize that the fact of not observing a significant evaluation bias in favor of a given group of candidates is not enough to discard the compensating behavior of selection boards toward this group. Indeed, it is sufficient that the evaluation biases giving advantage (resp. disadvantage) to a group of candidates are significantly more frequent when this group is less (resp. more) successful in the written tests. To investigate the likelihood of a compensation mechanism in oral test grades, we need to estimate the relationship between inequalities in the written tests and evaluation biases in the oral tests.

Relationship between Inequalities in Written Tests and Evaluation Biases

Table 4 gives, for each criterion, the results of the meta-regressions that relate the extent of inequalities of success in the written tests with the magnitude of evaluation biases, with and without controls for the characteristics of the competitive sessions.

The results validate hypothesis 3 for gender and place of birth. The more male candidates are successful in the written tests compared with women, the more women candidates benefit from evaluation biases to their advantage in the oral tests, and vice versa. The same type of compensation mechanism seems to prevail for candidates born outside metropolitan France. However, the coefficient associated with our measure of inequality of success in the written tests becomes nonsignificant when the characteristics of the competitive sessions are taken into account. In fact, supplementary

Table 4 Results of Meta-Regression: Effect of Inequalities of Success in the Written Tests on Evaluation Biases in the Oral Tests

	Woman	Place of Residence		Born Outside Metropolitan France
		Paris Resident	ZUS Resident	
		Without independent variables		
Constant	0.037*** (3.82)	–0.002 (0.08)	–0.019 (1.60)	–0.130* (1.72)
Inequality of success	–0.834*** (5.71)	–0.026 (0.15)	0.213 (1.02)	–1.636** (2.35)
Adjusted R^2	0.14	–0.003	–0.003	0.09
		With independent variables		
Constant	0.067 (1.09)	0.001 (0.01)	–0.112 (1.57)	–0.259 (1.35)
Inequality of success	–0.858*** (5.72)	–0.079 (0.42)	0.221 (1.06)	–1.256 (1.66)
Ministry				
Economy	Ref.	Ref.	Ref.	Ref.
Agriculture	0.122 (1.30)	–0.731*** (5.40)	–0.166 (1.49)	—
Foreign Affairs	–0.077 (0.83)	0.069 (0.61)	0.088 (0.81)	–0.026 (0.16)
Interior	–0.015 (0.24)	0.122 (1.23)	0.094 (1.29)	–0.077 (0.79)
Labor	–0.235** (2.42)	0.140 (0.96)	0.012 (0.11)	–0.079 (0.53)
National Education	0.015 (0.39)	–0.084 (1.40)	–0.143*** (2.91)	—
Higher Education	–0.073 (0.96)	–0.040 (0.036)	0.077 (0.076)	—
Category				
B/C	Ref.	Ref.	Ref.	Ref.
A	–0.114** (2.47)	0.107 (1.60)	0.110** (2.00)	–0.073 (0.91)
Log number of recruits	0.013 (1.24)	–0.005 (0.32)	0.021* (1.68)	0.047 (1.32)
Adjusted R^2	0.17	0.14	0.05	0.07

Note: Absolute *t*-values are in parentheses.
***Significant at the 1% level;
**at the 5% level;
*at the 10% level.

investigation reveals that the compensation phenomenon favoring candidates born outside metropolitan France takes place more often in the Ministry of Labor and in the largest competitive examinations.

Hence, selection boards seem to take into account the gender of the candidates and, to a lesser extent, their place of birth in order to offset the inequalities observed in the written tests during the oral tests. This behavior is aimed at improving the representativeness of the civil service in terms of gender and place of birth. On the other hand, there is no relationship between inequalities in the written tests and oral evaluation biases concerning place of residence. Therefore, the competitive examinations in which ZUS candidates are disadvantaged in the oral tests are not the same as those in which they suffer from the lowest probability of success in the written tests.

Discussion

According to the "four-fifths rule," which is the most common measure of adverse impact used in the United States, differences in average success rates in the written tests of French civil service competitive examinations show evidence of adverse impacts. This rule, codified in the 1978 Federal Uniform Guidelines on Employee Selection Procedures, states that a selection rate for a given group that is less than 80 percent of that of the highest-scoring group may be regarded as adverse impact. According to this rule, table A1 in the Supporting Information reveals that the written tests provide adverse impacts against candidates born outside metropolitan France in all ministries and against non-Parisian candidates with the exception of the Ministries of Economy and National Education. These tests also generate adverse impacts against women, but only in the Ministries of Agriculture and Foreign Affairs.

As ruled by the U.S. Supreme Court, the important question that arises is, are the inequalities recorded in the written tests "reasonably related" to the duties performed in the jobs? Many reasons lead us to think that the answer is negative. First, adverse impacts do not disappear when the differences in the candidates' characteristics that may influence their performance in written tests are neutralized. Second, the differences in success rates according to place of birth and place of residence are of such magnitude that one may legitimately believe that some of the differences are based on selection criteria irrelevant to the job. Third, a lower mastery of the French language (relevant criterion for public employment) cannot fully explain the differences in success rates according to place of birth. Indeed, 63.4 percent of the candidates born outside metropolitan France come from the French overseas departments and territories where French is the official language. Moreover, these candidates have a significant lower success rate on the written tests than immigrant candidates: 13.2 percent versus 17.1 percent (26 percent for candidates born in metropolitan France). Fourth, although inequalities of success in the written tests between women and men are smaller, we carried out a complementary analysis suggesting that they are essentially based on irrelevant criteria for public employment. To do this, we considered competitive examinations to become a teacher, and more particularly the Agrégation examination. Some of the candidates for the Agrégation examination are already teachers who have passed successfully the CAPES competitive examination. Thus, we know that they possess the required skills to be a teacher as well as some professional experience. Agrégation provides higher status with higher pay and shorter working hours. It also offers the opportunity to teach in higher education institutions. When we compare the effect of gender on the probability of success in the written tests, depending

on whether or not the candidates are already teachers, we find that in both cases, women have a significantly lower probability of success than men. More precisely, this probability is 11.3 percentage points lower when all candidates are considered and 6.7 percentage points lower when we only consider those who are already teachers. Therefore, female teachers who succeeded in passing the CAPES competitive examination are still less likely to pass written tests of Agrégation examination than their male counterparts. To sum up, it appears that the significant inequalities of success in the written tests are partly based on irrelevant selection criteria.

Furthermore, our results show that in two-stage competitive examinations, the disadvantaged (resp. advantaged) candidates in the written tests can find themselves at an advantage (resp. at a disadvantage) in the oral ones. This is particularly the case for women, for whom the evaluation biases in the oral tests tend to offset, at least in part, the inequalities observed in the written tests. We interpret this compensation as a result of the lifting of the candidates' anonymity, allowing selection boards to make their preferences prevail.

Two main arguments favor such interpretation. First, this compensation is observed particularly for gender, the most visible characteristic in the oral tests, whereas it is not observed for place of residence, which cannot be directly identified by selection boards. Second, our econometric estimation strategy controls for unobservable skills that influence results in both types of tests in the same way. However, any difference in design between the written and oral tests is likely to explain, at least in part, these evaluation biases provided that they lead to valuing differently unobservable skills that are unequally distributed between women and men. For example, an evaluation bias favoring women may be due to the fact that the oral tests require less preparation time than the written ones and that women have less average time available than men because of their social roles. In this case, the evaluation bias would not reflect the fact that the selection board wishes to promote women but would come from a female disadvantage that would be less prejudicial to success in the oral tests than in the written tests.

To confirm whether gender evaluation biases are really related to deliberate selection board behavior, we carried out an additional statistical analysis by comparing gender evaluation biases within pairs of competitive examinations whose tests designs are very similar.[15] The results show that the magnitude of gender evaluation biases is significantly different for each pair of similar examinations analyzed, suggesting that these biases reflect different preferences of selection boards as the valuation of skills in the written and oral tests is by definition similar.[16]

Finally, the preferences of selection boards toward gender could stem from stereotypes as well as from the desire to promote gender diversity within civil service. Additional statistical analyses reveal that evaluation biases in favor of women are negatively influenced by the share of women among the candidates who took part in all the written tests. Furthermore, the fact of including this share as an explanatory variable in the meta-regression presented in table 4 does not modify the effect of inequalities of success in the written tests on evaluation biases. In other words, during the oral tests, selection boards seek to improve gender diversity among the hired candidates not only by offsetting the inequalities observed in the written tests, but also by taking into account the gender disequilibrium among the initial pool of candidates.

Such an offsetting behavior could prevail for candidates born outside metropolitan France. However, the small number of observations we have in our meta-analysis does not allow us to dig further into the understanding of the compensation mechanism.

Conclusion

While much of the public administration research advances convincing normative arguments for a representative civil service workforce, very little attention has been given to investigating the capacity of competitive examinations to treat candidates equally whatever their sociodemographic characteristics while selecting the best candidates. Our research fills this gap by considering competitive examinations that have a first stage consisting of anonymous written tests and a second one consisting of non-anonymous oral tests. This specific setting acts as a natural experiment for assessing evaluation biases by selection board through the comparison of the anonymous and non-anonymous stages of the competition.

We use a unique French database that allows us to estimate the effects of candidates' gender, place of birth, and place of residence on inequalities of success in written tests and on selection boards' evaluation biases for a large number of competitive sessions. While the written tests are the most selective step of competitive examinations, they are significantly less successful, all else being equal, for candidates born outside metropolitan France, non-Parisian candidates, and, to a lesser extent, female candidates. These results suggest that adverse impacts can occur during the written tests. Consequently, if the French civil service wishes to build a diverse, inclusive, and efficient workforce, revising the design of the written tests while preserving their ability to select the best candidates should be considered.

However, the oral tests lead to evaluation biases aimed at compensating for the inequalities related to gender and place of birth encountered in written tests. This strategy of selection boards can help improve the diversity among the civil servants, but it may also be considered positive discrimination, which can stigmatize those who benefit from it (Coate and Loury 1993; Foley and Williamson 2019). Furthermore, biases in the oral tests operate to offset inequalities in the written tests at the group level only. It does not correct the fact that individuals unfairly excluded because of written tests adverse impacts have been denied opportunity to prove their true ability.

The overall negative effect size for evaluation biases concerning ZUS resident candidates, which is independent of inequalities of success in written tests, should be investigated further, as it could be related to a discriminating behavior of selection boards toward candidates from more deprived areas.

Finally, to better guide the state in its revision of competitive examinations, further research is necessary to determine to what extent the nature of written tests is responsible for the inequalities of success and how it penalizes some groups of candidates. In this context, the use of the grades obtained in the various written tests

could highlight which specific subjects or types of test explain the observed inequalities.

Acknowledgments

This article is part of a project supported by the French Ministry of Public Action and Accounts (Directorate General of Administration and Public Service).

Notes

1. Article 6 lays down the principle according to which "all citizens . . . are also eligible for all dignities, places and public positions, according to their ability, and without distinction other than that of their virtues and talents."
2. The ministries considered are Agriculture; Economy; Foreign Affairs; Higher Education, Research, and Innovation; Interior; Labor; and National Education.
3. The École Normale Supérieure prepares students for high-level teaching and academic careers.
4. For example, in many French competitive examinations, a report summarizing the profile of the average successful candidate in the written tests scores is available for oral examiners.
5. In June 2015, the prime minister entrusted Yannick L'Horty with a mission to assess the risk of discrimination in access to public employment (L'Horty 2016). We were able to access these data as part of that mission.
6. The external civil service competitive examinations are grouped into three hierarchical categories (A, B, and C) based on minimum qualification requirements.
7. When only a small number of candidates engaged in a competitive session, we group successive sessions as if they formed a single session to obtain estimates that are more statistically reliable.
8. A sensitive urban zone (ZUS) is a deprived area defined by the authorities to be a high-priority target for city policy, taking into consideration the housing, employment, and education problems of its residents.
9. Metropolitan France is the European territory of France. Hence, candidates born outside metropolitan France were born either in French overseas departments and territories or in another country.
10. The poverty rate is the share of households whose income falls below 60 percent of the median household income of the total population.
11. A young person is from a given geographic area if both parents were born in that area.
12. In the Ministries of National Education and Higher Education, Research, and Innovation, we used first names to identify the gender of candidates, which implies that it remains unknown for candidates with gender-neutral first names. We control for this lack of information about gender using a dummy variable in our regressions.
13. Our study deals with external competitive examinations, but civil servants are allowed to take part in these.
14. The numbers of observations in tables 2 and 3 for gender and place of residence are somewhat smaller than in table 1. This is because in some competitive sessions, the number of candidates for a given criterion was too small to estimate the effect.
15. We considered three pairs of competitive examinations. The first pair corresponds to the two police officer competitive examinations of the Ministry of Interior, which are distinguished only by their place of assignment. The second pair consists of competitive examinations for the recruitment of secondary school teachers in physics and chemistry. Finally, within the Ministry of Economy, we distinguished a third pair of examinations for the recruitment of public finance inspectors (general and operating system programmers). For these last two pairs, the design of the tests is not exactly identical but very similar.
16. The results are available from the authors upon request.

References

Anderson, Lisa R., Roland G. Fryer, and Charles A. Holt. 2006. Discrimination: Experimental Evidence from Psychology and Economics. In *Handbook on the Economics of Discrimination*, edited by William M. Rodgers, III, 97–115. Northampton, MA: Edward Elgar.

Andrews, Rhys, and Rachel Ashworth. 2015. Representation and Inclusion in Public Organizations: Evidence from the U.K. Civil Service. *Public Administration Review* 75(2): 279–88.

Arrow, Kenneth J. 1973. The Theory of Discrimination. In *Discrimination in Labor Markets*, edited by Orley Ashenfelter and Albert Rees, 3–33. Princeton, NJ: Princeton University Press.

Atkins, Danielle N., and Vicky M. Wilkins. 2013. Going beyond Reading, Writing, and Arithmetic: The Effects of Teacher Representation on Teen Pregnancy Rates. *Journal of Public Administration Research and Theory* 23(4): 771–90.

Bagues, Manuel F., and Berta Esteve-Volart. 2008. *Top Civil Service: Meritocracy or Nepotism? Unpublished manuscript*.

———. 2010. Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment. *The Review of Economic Studies* 77(4): 1301–28.

Bagues, Manuel F., Mauro S. Labini, and Natalia Zinovyeva. 2017. Does the Gender Composition of Scientific Committees Matter? *American Economic Review* 107(4): 1207–38.

Becker, Gary. 1957. *The Economics of Discrimination*. Chicago: University of Chicago Press.

Beffy, Magali, Elise Coudin, and Roland Rathelot. 2014. For Whom Are Permanent Jobs Off Limits? A Markov-Chain-Based Analysis of Individual Labor Market Dynamics. *Annals of Economics and Statistics* 115–116: 315–42.

Bourdieu, Pierre, and Jean-Claude Passeron. 1964. *Les Héritiers: Les étudiants et la culture*. Paris: Les Editions de Minuit.

Bradbury, Mark, and J. Edward Kellough. 2011. Representative Bureaucracy: Assessing the Evidence on Active Representation. *American Review of Public Administration* 41(2): 157–67.

Bréda, Thomas, and Son Thierry Ly. 2015. Professors in Core Science Are Not Always Biased against Women: Evidence from France. *American Economic Journal: Applied Economics* 7(4): 53–75.

Bréda, Thomas, and Mélina Hillion. 2016. Teaching Accreditation Examinations Reveal Grading Biases Favor Women in Male-Dominated Disciplines in France. *Science* 353(6298): 474–8.

Bunel, Mathieu, Yannick L'Horty, and Pascale Petit. 2016. Discrimination Based on Place of Residence and Access to Employment. *Urban Studies* 53(2): 267–86.

Campbell, John P. 1996. Group Differences and Personnel Decisions: Validity, Fairness, and Affirmative Action. *Journal of Vocational Behavior* 49(2): 122–58.

Close, Billy R., and Patrick L. Mason. 2006. After the Traffic Stops: Officer Characteristics and Enforcement Actions. *Topics in Economic Analysis & Policy* 6(1): 1–41.

Coate, Stephen, and Glenn C. Loury. 1993. Will Affirmative-Action Policies Eliminate Negative Stereotypes? *American Economic Review* 83(5): 1220–40.

Combes, Pierre-Philippe, Laurent Linnemer, and Michael Visser. 2008. Publish or Peer-Rich: Skills and Networks in Hiring Economics Professors. *Labour Economics* 15(3): 456–74.

Couppié, Thomas, Jean-François Giret, and Stéphanie Moullet. 2010. Lieu de résidence et discrimination salariale: Le cas des jeunes habitant dans une zone urbaine sensible. *Économie et statistique* 433–434: 47–70.

Datta, Gupta, Nabanita Anders Poulsen, and Marie-Claire Villeval. 2013. Gender Matching and Competitiveness: Experimental Evidence. *Economic Inquiry* 51(1): 816–35.

Dee, Thomas S. 2005. A Teacher Like Me: Does Race, Ethnicity, or Gender Matter? *American Economic Review* 95(2): 158–65.

Défenseur des droits and CNIL [Defender of Rights and National Commission on Informatics and Liberty]. 2012. Mesurer pour progresser vers l'égalité des chances: Guide méthodologique à l'usage des acteurs de l'emploi. https://www.cnil.fr/sites/default/files/atoms/files/ddd_gui_20120501_egalite_chances_0.pdf [accessed April 3, 2019].

Falch, Torberg, and Linn R. Naper. 2013. Educational Evaluation Schemes and Gender Gaps in Student Achievement. *Economics of Education Review* 36: 12–25.

Foley, Meraiah, and Sue Williamson. 2019. Managerial Perspectives on Implicit Bias, Affirmative Action, and Merit. *Public Administration Review* 79(1): 35–45.

Goldin, Claudia, and Cecilia Rouse. 2000. Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians. *American Economic Review* 90(4): 715–41.

Guul, Thorbjørn S. 2018. The Individual-Level Effect of Gender Matching in Representative Bureaucracy. *Public Administration Review* 78(3): 398–408.

Hanna, Rema N., and Leigh L. Linden. 2012. Discrimination in Grading. *American Economic Journal: Economic Policy* 4(4): 146–68.

Hartigan, John, and Alexandra Wigdor. 1989. Fairness in Employment Testing. *Science* 245(4913): 14.

Holzer, Harry J., and David Neumark. 2000. What Does Affirmative Action Do? *ILR Review* 53(2): 240–71.

Hong, Sounman. 2017. Does Increasing Ethnic Representativeness Reduce Police Misconduct? *Public Administration Review* 77(2): 195–205.

Hough, Leatta M., Frederick L. Oswald, and Robert E. Ployhart. 2001. Determinants, Detection and Amelioration of Adverse Impact in Personnel Selection Procedures: Issues, Evidence and Lessons Learned. *International Journal of Selection and Assessment* 9(1–2): 152–94.

L'Horty, Yannick. 2016. Les discriminations dans l'accès à l'emploi public. Report to the Prime Minister, June. https://www.fonction-publique.gouv.fr/files/files/Espace_Presse/girardin/Rapport_LHorty_final.pdf [accessed April 3, 2019].

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): 2083–105.

Levade, Anne. 2004. Discrimination positive et principe d'égalité en droit français. *Pouvoirs* 111: 55–71.

Lipsky, Michael. 1980. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage Foundation.

Meier, Kenneth J., and Jill Nicholson-Crotty. 2006. Gender, Representative Bureaucracy, and Law Enforcement: The Case of Sexual Assault. *Public Administration Review* 66(6): 850–60.

Meier, Kenneth J., Robert D. Wrinkle, and J.L. Polinard. 1999. Representative Bureaucracy and Distributional Equity: Addressing the Hard Question. *Journal of Politics* 61(4): 1025–39.

Mosher, Frederick C. 1982. *Democracy and the Public Service*. Oxford: Oxford University Press.

Nicholson-Crotty, Sean, Jason A. Grissom, Jill Nicholson-Crotty, and Christopher Redding. 2016. Disentangling the Causal Mechanisms of Representative Bureaucracy: Evidence from Assignment of Students to Gifted Programs. *Journal of Public Administration Research and Theory* 26(4): 745–57.

Ors, Everen, Frédéric Palomino, and Eloic Peyrache. 2013. Performance Gender Gap: Does Competition Matter? *Journal of Labor Economics* 31(3): 443–99.

Patterson, H. Desmond, and Robin Thompson. 1971. Recovery of Inter-block Information When Block Sizes Are Unequal. *Biometrika* 58(3): 545–54.

Riccucci, Norma M., Gregg G. Van Ryzin, and Huafang Li. 2016. Representative Bureaucracy and the Willingness to Coproduce: An Experimental Study. *Public Administration Review* 76(1): 121–30.

Sprietsma, Maresa. 2013. Discrimination in Grading: Experimental Evidence from Primary School Teachers. *Empirical Economics* 45(1): 523–38.

Sundell, Anders. 2014. Are Formal Civil Service Examinations the Most Meritocratic Way to Recruit Civil Servants? Not in All Countries. *Public Administration* 92(2): 440–57.

Theobald, Nick A., and Donald P. Haider-Markel. 2009. Race, Bureaucracy, and Symbolic Representation: Interactions between Citizens and Police. *Journal of Public Administration Research and Theory* 19(2): 409–26.

Supporting Information

A supplemental appendix can be found in the online version of this article at http://onlinelibrary.wiley.com/doi/10.1111/puar.13053/full.