DISCUSSION PAPER SERIES

Culture as a Hiring Criterion: Systemic Discrimination in a Procedurally Fair Hiring Process

Dominique Meurs
Patrick A. Puhani

DISCUSSION PAPER SERIES

Culture as a Hiring Criterion: Systemic Discrimination in a Procedurally Fair Hiring Process

Dominique Meurs
Université Paris-Ouest and INED

Patrick A. Puhani
Leibniz Universität Hannover, University College London, University of St. Gallen and IZA

FEBRUARY 2021

ABSTRACT

Culture as a Hiring Criterion: Systemic Discrimination in a Procedurally Fair Hiring Process*

Criteria used in hiring workers often do not reflect the skills required on the job. By comparing trainee performance for newly hired workers conditional on competitive civil service examination scores for hiring French public sector workers, we test whether women and men with the same civil service examination score exhibit similar performance in a job-related trainee programme. Both the civil service examination and trainee scores contain anonymous and non-anonymous components that we observe separately. We find that by the end of the trainee programme (first year of employment), women are outperforming men on both anonymous written and non-anonymous oral evaluations, a finding that holds both conditionally and unconditionally for the civil service examination results. According to further analysis, however, it is the anonymously graded "essay on common culture" civil service examination that, unlike the other CSE components, disadvantages women in this particular context.

Corresponding author:
Patrick A. Puhani
Leibniz Universität Hannover
Institut für Arbeitsökonomik
Königsworther Platz 1
D-30167 Hannover
Germany
E-Mail: puhani@aoek.uni-hannover.d

Tous les Citoyens … sont également admissibles à toutes dignités, places et employs publics, selon leur capacité, et sans autre distinction que celle de leurs vertus et de leurs talents

— Déclaration des droits de l'homme et du citoyen, 1789[1]

1   Introduction

Despite serious efforts to increase procedural fairness in hiring, discrimination against socio-demographic or ethnic groups remains a potential source of labour market inequalities, in particular because the skills systematically tested during hiring may vary among such subpopulations. As a result, even apparently fair hiring practices may discriminate (whether knowingly or unknowingly) by either testing for skills that are not relevant to job performance or failing to test for those that are. This systematic failure to acknowledge differences in applicants' potential productivity may result in disparate hiring outcomes, hereafter labelled "systemic discrimination" or "disparate impact". The concept of systemic discrimination (disparate impact) is rooted in a 1971 U.S. Supreme Court ruling (*Griggs v. Duke Power Co.*) that because having a high school diploma appeared uncorrelated with work productivity in certain jobs, its requirement discriminated against Black employees, who were about half as likely as White North Carolinians to have one. The court ruling thus required that any hiring criteria resulting in very different success rates for groups protected under Title VII of the Civil Rights Act must be shown to be necessary for job performance (Bornholz and Heckman, 2005; Peresie 2009).

This aspect, although widely discussed in the legal literature (Selmi 2006, Peresie 2009), is little addressed in the business and economic literature, which focuses instead on gender-based differences or hiring and promotional aspects like willingness to compete (Sutter and Glätzle-Rützler 2015), jury composition (Bagues and Esteve-Volart 2010), and joint or separate candidate evaluation (Bohnet et al. 2016). One of a handful of exceptions is a study by McCrary (2007), which illustrates disparate impact by showing that American Black under-representation among successful police force applicants is partly explainable by their greater likelihood (relative to Whites) of failing the police

---

[1] "All citizens … are equally admissible to all honours, places, and public employments, according to their abilities, and without distinctions other than their virtues and their talents" (author translation).

academy entrance examination. Minorities also perform worse on standardized tests, which, although they may facilitate fair recruitment with respect to abilities tested, are subject to the issue of efficiency versus equity in the employment process (Autor and Scarborough 2008) Even more relevant for our analysis, whereas introducing blind auditions into professional American symphony orchestra hiring increased the share of female musicians (Goldin and Rouse 2000), blind review processes for scientific grant proposals may disadvantage women because of "gender differences in writing and communication" (Kolev, Fuentes-Medel, and Murray 2019) that are unrelated to scientific productivity. When only *successful* grant proposals are considered, however, women tend to outperform men on scientific output measures (Kolev, Fuentes-Medel, and Murray 2019).

To further explore this issue of gender disparity, we perform a regression analysis that tests for the presence of gender differences in trainee evaluations (as a proxy for job performance) in the French civil service once the results of the service entrance examination (CSE, *concours administratifs*) are held constant. The primary basis for our analysis is CSE data for the entire population of those aspiring to higher management positions in French regional administration. These include "external" candidates (those not yet working in the administration, such as university graduates or private sector workers) and "internal" candidates (those having at least 4 years such employment, usually in a lower-level position), most of whom must already have a college degree. Not only do we have access to all CSE scores, but unlike audit studies that stop at the invitation to a job interview, our analysis encompasses the entire recruitment process plus one year of job-related training, together with all test and evaluation scores (both anonymous and non-anonymous) after the two-semester trainee programme.

Because hiring decisions are mechanically determined based on CSE score ranking *only,* any differences other than CSE results (e.g., university degree or major) are irrelevant to the hiring decision and thus to our assessment of systemic discrimination. If these CSE scores were truly gender neutral, we would expect no systemic differences between female and male application success once the CSE results are held constant. Otherwise, we would infer systemic (disparate impact) gender

discrimination, which is distinct from procedural discrimination in the competitive hiring process.[2]

Our most general observation is that even when the competitive examination scores are held constant, women outperform men in the subsequent trainee programme. This finding implies that the criteria reflected in the competitive examination so little resemble job requisites as to generate systemic discrimination against women, which an analysis of the separate CSE components reveals to be caused primarily by the "essay on common culture" (hereafter, the common culture essay). Although the effect itself, while highly statistically significant, is not very large, it is of particular interest given recent proposals to abolish elite French institutions like the *école nationale d'administration* (ENA) − the President's own alma mater − or to scrap private British schools like Eton College, which the Labour Party wishes to strip of its tax privileges. This debate is fuelled by a suspicion that elites distinguish and protect themselves via "a hidden curriculum" of hard-to-acquire codes not taught in the general school system (Bourdieu and Passeron 1964, Bourdieu 1979). Quirky examples include French cultural conventions for cutting different types of cheeses or a proscription against wearing brown shoes with business suits in the City of London's investment banker culture.[3] By making it costly to acquire the cultural knowledge that provides access to privileged jobs, elites may protect themselves and hamper others' social mobility through systemic discrimination.[4]

In our case, although women are generally over-represented in the civil service applicant share, they are slightly under-represented in the share of those who pass the entrance examination. Yet once on the job, females who pass show higher performance (as proxied by trainee evaluations) than their male counterparts, indicating that the CSE, by failing to adequately test female candidates'

---

[2] Although Meurs (2019) is able to provide immigrant statistics based on 2007 and 2010 survey data, the administrative data we use do not record migration status, so we are unable to investigate how immigrants fare in this hiring process.

[3] See "Emmanuel Macron's plan to close ENA shakes French elite", *Financial Times*, April 18, 2019; "Labour's John McDonnell Backs Motion to Scrap Private Schools", *Financial Times*, September 18, 2019; "'Brown Shoes with a Suit? A high court ruling is urgently required'" *Financial Times*, April 18, 2019 "What to Wear to Get Ahead in Finance—According to HBO's 'Industry'"; *Wall Street Journal*, December 19, 2020; and online information from several French cheese vendors on how to cut different cheeses (retrievable using the search term *découper des fromages*).

[4] For recent newspaper features on "systemic" discrimination, see for example "How to Make Society Less Racist", *Financial Times*, June 19, 2020; "Sustainability Investors Shift Their Focus to Social Issues", *Wall Street Journal*, October 10, 2020; "How to Make Artificial Intelligence Less Biased", *Wall Street Journal*, November 3, 2020; "Black and Minority Ethnic Workers Hit Harder by Covid Job Losses", *Financial Times*, January 20, 2021; "Joe Biden Embraces His Inner Radical to Confront Winter of Peril", *Financial Times*, January 21, 2021.

job competencies, inherently lowers the share of women hired and thus promotes systemic discrimination. We further show that even though comparing the written and oral CSE scores for males versus females suggests a hiring process that is procedurally fair or even slightly favouring of women in the non-anonymous oral tests, an analysis of each score component separately identifies the common culture essay as a main reason that the women in our sample perform worse than the men.

Although France has a comparatively low gender wage gap relative to other OECD countries (about 14% vs. 19% in the US according to the OECD database[5]), its two-century old tradition of hiring workers based purely on merit (see opening epigram) makes it a particularly interesting case study. That is, the nation has adhered to a merit-based system for virtually all public sector hiring – as well as for entry into (the mostly public) elite educational institutions – since long before anti-discrimination policies became widespread elsewhere (Greenan et al. 2019). The research context is also highly representative in that the public sector is France's largest employer, with about a 22% share of total employment compared to only about 15% in the US (OECD 2019, p. 85).

The remainder of the paper is structured as follows: in Section 2 we explain the competitive French CSE process − the prevailing system of recruitment into public sector positions and elite higher education − and outline the trainee programme and trainee evaluations used to proxy job performance. After briefly describing our administrative data sets in Section 3, in Section 4, we report the results of our main analysis − the test for systemic discrimination − which probes for male-female differences in job-related trainee test scores conditional on CSE outcomes. In Section 5, to effect a more in-depth examination of these CSE results, we adopt the audit study method of comparing male and female non-anonymous oral scores conditional on anonymous written scores. We then report the results for separate regressions for each year in Section 6. Lastly, in Section 7, we conclude the paper by summarising and discussing the implications of our findings.

---

[5] See https://data.oecd.org/earnwage/gender-wage-gap.htm for an international comparison based on 2019 or latest available data.

2   Recruitment into the French Public Sector via the CSE and Subsequent Trainee Evaluation

Both entry into France's most prestigious schools and colleges and securing a middle or higher management position in its regional public administration[6] (*instituts régionaux d'administration*, IRA) are contingent upon passing competitive examinations (*concours*); in the latter case, *les concours administratifs* or civil service examination. Whereas the *école nationale d'administration* (ENA), attended by many French presidents and other leading figures, accepts only about 80 applicants per year into *top* management levels, the IRA institutions one hierarchical level below recruit 600–700 annually.[7] These competitive examinations are the only path to higher management positions, and the CSE score cut-off points are determined for each training region based on the number of available jobs. During the trainee programme, which includes one semester of general training and a second of specialization, candidates also obtain on-the-job work experience in the form of two mandatory internships in public administration. In addition to a stipend of around €1,600 per month, trainees are *de facto* guaranteed a position. Once the two-semester trainee evaluations determine their ranking, they may pick their preferred regular civil service employment from a list of vacancies based on that ranking.[8]

Before sitting their CSE, candidates must choose one of five trainee centres in Bastia (Corse), Lille (Nord), Lyon (Rhône-Alpes), Metz (Lorraine), or Nantes (Pay de la Loire),[9] all of which serve their surrounding region, as well as some positions in Paris. They are allowed to see each centre's numbers to gauge competition and can change locale up to the last day of registration.[10] As Table 1

---

[6] This contingency holds true not only for recruitment into higher management (considered here) but also for the entire public sector, even to support staff such as secretaries, although these latter must pass a different entry test tailored to their prospective jobs.

[7] A survey question answered by about 60% of the 2007 job candidates on educational attainment (not recorded in the administrative data) shows that about 90% had at least a bachelor's degree and 75% at least a master's degree (Eberhard et al. 2008).

[8] An overview of the hiring and training process (in French), is available at http://www.fonction-publique.gouv.fr/score/ecoles-de-formation/ira-et-ena/ira

[9] Because most candidates are supposed to work in French regions, there is no such training centre in Paris. Here, the regions listed after the cities are from before the 2015 regional reform.

[10] Although the competitive written tests take place simultaneously in various testing centres across France and in overseas departments and territories like Réunion or Guadeloupe, they are sent for grading to the regional training centres, each of which employs about 20 examiners who correct the test papers anonymously. Of these 20 individuals, two-thirds or more are chosen to build the three-person juries of oral examiners, which should include at least one man and one woman and be equivalent in terms of professional status, work

shows, the CSE structure includes different examinations for external and internal candidates, both sets of which were redesigned in 2008. Before that year, the written test for internal candidates comprised two parts: a dossier (worth 4/9 of the written score) and a short-question test (worth 5/9). The 2007 dossier, for example, contained 10 documents on sustainable development policies (40 pages, including various decrees and government committee reports), based on which the candidate, pretending to be in charge of a prefecture in Provence-Alpes-Côte d'Azur, had to develop a strategy for a programme of sustainable environmental policies in the region. The short-question test (eliminated in 2008) contained 10 items (3 on public law, 2 on European politics, 3 on public finance, and 2 on economics) to be answered within a 3-hour time frame. The pre-2008 examination for external candidates included a similarly structured 3-hour short-question test (5/12), a four-hour essay on a given common culture topic such as "The Internet, A New Floor for Public Debate" (4/12), and a summary of an administrative document (3/12), also eliminated in 2008.

Taking the oral test is contingent on passing the written test, whose exact score is hidden from both the jury and the examinee. These orals are actually a conversational interview with a three-person jury, with scores originally worth 5/14 (5/17) of the total for internal (external) candidates but with different weights since 2008 (see Table 1). The interview comprises a 25-minute discussion of the candidate's motivation for seeking this type of job, as well as previous work experience and training.[11] One final step for successful external candidates is a mandatory modern languages test (an option for internal candidates), which, however, carries little weight (a factor of 1) relative to the orals (a factor of 4). Our test for systemic discrimination compares CSE scores with performance

_____

experience, and career type. In fact, almost all jury members are civil servants at a position hierarchically higher than that to be assumed by successful candidates immediately after training. Juries are also supposed to be from a mix of backgrounds that represents the ministries for which the candidates will work (e.g. half from the national educational administration, with the others from the Departments of the Interior, Defense, Finance, or Transport). Jury members receive one day of training on oral test administration, including sensitization to discrimination and instructions for creating a tabular record of their evaluations of candidate responses. Jury member names and job positions (e.g. principal attaché, director of a prefecture, hospital administrator, or university professor) are published in advance on regional training centre web sites, as well as in official government decrees.

[11] Since 2008, the candidates have also had to submit a dossier that includes a résumé and statement of job application motives. Prior to this time, the interview contained an additional 8- to 10-minute discussion of a text or topic prepared by the candidate during the 30 minutes before the interview (Eberhard et al. 2008, p. 89).

evaluations from the subsequent two-semester trainee programme designed to teach and test professional job requisites.[12]

To explain the nature of the trainee programme, we borrow from detailed documentation on the web pages of the five training centres,[13] which explicitly label the training "professional" and describe two mandatory professional internships in the administration for which candidates are being trained. The web site of the Lyon training centre, for example, explicitly states that the aim of the professional training is to help candidates "acquire a professional culture" that can be "operational immediately". This acquisition includes an understanding of the "technical culture" needed on the job, as well as "various competencies (legal, financial, management, human resources, office communication …)", with "particular stress …. on the functioning of teams, management, team work and communication".

During the first semester, the candidates' training comprises 80 hours of instruction on the public sector framework (*cadre et enjeux de l'action publique*), 160 hours on administration means and methods (*moyens et méthodes de l'administration*), 100 hours on management, and 40 hours of foreign languages (English, German, Italian, Spanish). This training applies various teaching formats, including small group sessions (around 20 individuals), small group preparation of collaborative papers, 2-3 hour larger group sessions on specific topics, whole day seminars, and evening sessions. Once the training centre publishes the available internship positions, the trainees state their preferences, and the training centre head assigns the internships.[14] It is these centre heads (rather than the evaluation juries) who assign the internship scores based on the recommendations of the internship supervisor and an internship report prepared by the trainee. During both the first and second semester internships, which last for 7 weeks and 8 weeks, respectively, the trainees are

---

[12] See the 2016 *concours* description, p. 26, at http://www.fonction-publique.gouv.fr/files/files/score/ecoles-formation/ira/Ira_2016/BrochureIRA_16(2).pdf and Article 4 of the governing law at https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000795736

[13] The quoted terms and passages in this paragraph, translated by the authors, are available at http://www.ira-nantes.gouv.fr/, http://www.ira-metz.fr/, https://extranet.ira-lyon.gouv.fr/fr/Pages/accueil.aspx, https://ira-bastia.gouv.fr/, and https://www.ira-lille.gouv.fr/.

[14] The regulations do not, however, clearly specify the mechanism of these internship assignments.

expected to actively participate and immerse themselves in jobs similar to the public administration positions they hope to secure after training.

As it is apparent, this entire paid trainee programme is explicitly targeted towards the candidates' future jobs, which they select from the list of vacancies posted by the administration contingent on their final trainee rankings. This process, in which first-ranked candidates have their choice of open job postings in the region attached to their training centre, typifies a meritocratic principle commonly applied in public sector competitions in France. As a result, trainees already have a job, and almost nobody fails the trainee programme. Nonetheless, having first choice of the most desirable positions provides a strong incentive to do well, especially given that, although the training centre picked at the CSE stage determines the greater geographic region of eventual assignment, jobs still differ by exact location and field. Even the ranking in the first semester impacts a trainee's choice of specialization field – whether central administration (*administration centrale, AC*), territorial administration (*administration territoriale de l'Etat, ATE*), or school and university administration (*administration scolaire et universitaire, ASU*) – which in turn defines the nature of their eventual placement.

The trainee programme evaluation comprises two sets of examinations, one at the end of each semester, which determine where the new employee will work. The first semester examination has four parts: an anonymously graded 5-hour anonymous written test based on an administrative problem that combines legal, organizational, financial, and communication issues (5/20); an internship grade based on supervisory judgement and an internship report (7/20); a 20-minute internship-based oral test of administrative management skills (5/20); and a 40-minute oral defense of an administrative group report, graded at the group (5 to 7 trainees) level (3/20). The second semester examination similarly contains an anonymously graded 5-hour written test solving one or more practical administrative cases (5/16), an internship grade similar to that in the first semester (5/16), a 30-minute oral test in the field of specialization (5/16), and a 20-minute foreign language oral discussion of a professional topic (1/16). The total trainee programme score is derived by dividing the first semester score by 2 and adding it to the second semester score such that these semesters account for 10/26 and 16/26 of the total score, respectively.

These examination sets are evaluated by six-person juries, one for each of the five training centres, appointed by the national ministry and tasked with both setting and evaluating these written and oral examinations. This six-person body is complemented in the second semester by three experts, one in each of the three areas of specialization. All jury members, who are replaced after 3 years or even earlier to guarantee turnover, must be in a position ranked higher than the job type to be filled by the candidate without ever having been a teacher in the trainee programme.

3   Data and Descriptive Statistics

Our data set contains information on all CSE candidates during the 6 years from 2007 to 2012, together with the performance evaluations of successful applicants who subsequently entered the one-year trainee programme. To illustrate the procedural time frame, the 2007 cohort's written tests took place in February 2008, after which successful candidates took the oral CSE in June 2008, entered training in Fall 2008, and completed training in Fall 2009. Hence, our last CSE cohort (2012) completed training in Fall 2014. As Figure 1 shows, the number of candidates taking the written CSE increased greatly over the 2000−2010 decade, rising from around 3,000 to about 7,000 but then declining to nearer 5,800 in 2014. The number of candidates passing the written test (thereby becoming eligible to take the oral) ranged from 1,300 to 1,600, whereas the number of those successfully completing the entire CSE process ranged between 685 (2003) and 995 (2007). These latter would either be offered a job (implying trainee programme entry) or put on a waiting list, compiled because candidate participation in multiple competitive examinations means that not every job offer is automatically accepted. As revealed in both Figure 1 and our analysis of individual-level data in Table 3 (row 1), depending on year, about two thirds to three quarters of the candidates during our study period failed the written test, while 77 to 182 of those who passed (dependent on year but decreasing over time) withdrew from the CSE process (possibly because of other offers).

Because of France's post-revolutionary tradition of anonymity, the only information provided by our data set other than gender (implicitly recorded in the salutation field) is pass or fail indicator, CSE scores and trainee evaluations, internal versus external candidate indicator, distribution across training centres, and year. To illustrate sample means at different stages of the application process, we

first report the means for candidates attending the trainee programme after having passed both the written and oral tests (Table A1), then those for the larger sample of all candidates who had passed the written and were taking the oral tests (Table A2), and lastly, those for the widest set of candidates, all those taking the written tests. In all three tables, we show the sample means separately by external versus internal candidate and by male versus female.

As shown by the raw means of the standardized trainee evaluations (Online Appendix Table A1 for the trainee sample means), both external and internal female candidates outperform men in almost all assessments, although more so among the former than the latter (for external candidates, all except one of these differences are significantly different from zero). Among the external candidates, although women significantly outperform men on the orals (Online Appendix Table A2 for sample means for candidates having passed the written and taking the oral CSE), men significantly outperform women in the two written CSE, that is the common culture essay and the short-question test (Online Appendix Table A3 for sample means for candidates taking the written CSE; the third written CSE was abolished in 2008). Among the internal candidates for whom the job-related dossier replaces the essay, however, the reverse is true in that women significantly outperform men on this exam (Online Appendix Table A3). Men do outperform women unconditionally on the short-question test among both external and internal candidates (Online Appendix Table A3), but this result turns out to be of no significance in our subsequent conditional analysis.

4 Systemic Discrimination (Disparate Impact)

To assess the male-female performance differences noted above, as well as their implications for potential gender discrimination, we regress the trainee evaluation scores on a gender dummy; the CSE scores; control dummies for year, internal versus external candidate, and training centre; and interaction terms between year and internal examination and year and training centre. Our regression analysis is based on the following equation:

$$traineescore_i = \tau\, female_i + \rho\, CSE\, scores_i + \beta x_i + \varepsilon_i \qquad (1)$$

In the specifications outlined in Table 4, we first include only the female gender dummy (column (1)); then add in controls for year, internal/external candidate, training centre, and interactions between

year and internal/external candidate as well as between year and training centre (column (2)); and lastly incorporate the CSE scores, which are also interacted with the year dummies (column (3)). Because adding in these regressors leads to only small changes in the female gender coefficients, we focus on the final column, which indicates the female performance relative to male when the CSE scores are *held constant*.

If the CSE were designed to measure all job-relevant characteristics correlated with gender, then the coefficient of female gender should be zero. In fact, as Table 4 shows, even holding the CSE scores constant, the women's trainee performance is significantly better than the men's, with overall trainee test scores 0.19 standard deviations higher, a statistically significant difference given a mere 0.03 standard error. Likewise, when we analyse overall trainee scores by semester, women significantly outperform men in both the first and second semesters, although somewhat more markedly in the latter, which requires field specialization. All these estimates are highly statistically significant.

For the separate trainee evaluation components, although the female gender coefficients are highly statistically significant for both semesters, ranging between 0.16 and 0.20 standard deviations for both the anonymous written trainee test and the internship, the results are less clear for the oral, defense, and language evaluations.[15] Yet the fact that these female gender coefficients are zero for the defense score is not surprising since it is the only one assigned on a group level (5 to 7 trainees). Hence, just as we expect gender mixing within each group (even though group composition data are unavailable), we also expect the female coefficient for a group defense score to be close to zero, which is empirically true (for the performance effects of gender mixing in groups, see Apesteguia et al. 2012). The female coefficients in the regression with end-of-first-semester orals and language test scores as the dependent variable are also close to zero and statistically insignificant, although they refer to individually assigned not group scores. However, for the second semester orals, the coefficient of female gender is again statistically significant, albeit small in size, at 0.06 standard deviations.

---

[15] The distributions of the standardized training evaluations overall, as well as the written and internship scores in both semesters, are graphed in Figure 3.

Given that candidates may sit various entrance examinations for different job types while also applying to private sector jobs, in theory our findings may be related to selection in that men who perform higher in the training programme may be more likely to leave the application process after successfully passing the CSE. Although the absence of trainee scores for candidates who do not enter the training programme prevents us from testing this hypothesis, we can still assess whether gender explains post-CSE exit versus programme entrance. The last row of Table 4 reports the coefficient of female gender for linear probability models with post-CSE exit as the dependent variable. Not only are no coefficients statistically significant, but all point estimates are small with estimated gender effects of 1 or 2 percentage points. It is thus very unlikely that gender differences in post-CSE exit explain our finding that women outperform men in the trainee programme even after we control for CSE scores.

To check the robustness of these results to using an alternative control for CSE score when regressing trainee scores on gender (see Table 5), we replace the separate variable controls for each CSE score (essay/dossier, short-question test, administrative summary, oral test, language test) with the total CSE score (column (2)). Whereas the multi-variable specification reflects the two-step CSE process of written (oral) test as the first (second) hurdle, the overall score defines successful candidates. Here, the results are similar but somewhat larger in size, with women now outperforming men by 0.22 instead of 0.19 standard deviations conditional on the total CSE score and other controls. Both numbers are highly statistically significant. When we allow for a more flexible systemic form − namely, a third-order polynomial in the total CSE score as a control variable − the gender gaps in the different trainee evaluations are almost identical to those when the total CSE score is included as a linear control variable (Table 5, column (3)).

In general, then, women do better than men in job-related trainee evaluations even conditional on CSE scores. In fact, as Table 4 demonstrates, although controlling for this latter only marginally changes the gender gap in trainee evaluation scores, the larger conditional than unconditional gender gap implies that not all CSE scores assess the job-related skills that women possess to a greater extent than men. Hence, the competitive examination process, while seemingly procedurally fair based on

our comparison of non-anonymous oral and anonymous written CSE scores (see Section 5 below), may in fact subject women to systemic discrimination.

In Table 6 and Table 7, we estimate the same models as in Table 4 but separately for external and internal candidates, respectively. As the tables show, although women outperform men in the trainee evaluations in both groups, the gender differences are much more pronounced for external candidates in terms of both size and statistical significance. Among external candidates, women outperform men by about a quarter of a standard deviation in the total trainee score, that is by 0.29 and 0.24 standard deviations in the unconditional and conditional comparisons, respectively. These estimated differences are statistically significant at the 1% level. Among internal candidates, the numbers are still positive but smaller, at 0.08 and 0.13 standard deviations in the unconditional and conditional comparisons, respectively, with only the latter number statistically significant at the 1% level. The last column of Table 6 exhibits the differences in the coefficients estimated in the coefficient for females between external (Table 6) and internal candidates (Table 7) in the full specification (that is also controlling for job entry/CSE exam scores). The results, which are obtained through a joint estimation for both external and internal candidates with interaction terms, indicate women's outperformance of men is statistically significantly higher (by 0.12 standard deviations) for external than internal candidates when considering the total trainees score.

When we examine the anonymous written trainee evaluations, women consistently outperform men among external candidates in both semesters, with estimated conditional differences in the former of between 0.22 and 0.24 standard deviations (all statistically significant at the 1% level, see Table 6, column (3)). Among internal candidates, the estimated conditional differences for the written trainee evaluations are smaller, at an insignificant 0.05 and a highly significant 0.12 standard deviations in the first and second semester, respectively (see Table 7, column (3)). In both written trainee evaluations, women's outperformance of men is statistically significantly higher for external than internal candidates (by 0.17 and 0.12 in the first and second semester, respectively).

The gender gap differences among external and internal candidates are more similar for the internship scores, with conditional gaps of 0.19 and 0.17 among external and 0.17 and 0.14 standard deviations among internal candidates, respectively (all statistically significant at the 1% level,

although the differences between internal and external candidates is not statistically significant). For the oral, defense, and language scores, we observe statistically significant and positive gender gaps only in the second semester and only among external candidates (with point estimates among internal candidates close to zero).

In sum, the phenomenon that women outperform men in the trainee evaluations is more pronounced for external than for internal candidates, suggesting that systemic discrimination (disparate impact) in the civil service examination process applies primarily to external candidates. One probable explanation is that the CSE dossier provides internal candidates with the opportunity to show how they would solve a job-related administrative problem (see Section 2), while the common culture essay required of external candidates is only loosely related to future job requirements, if at all. Admittedly, the 10 short questions included with the latter may be job relevant; however, the internal candidates' dossier is obviously far better tailored to assessing future job potential.

To throw more light on exactly which parts of the CSE are responsible for this systemic discrimination, Table 8 reports the results of separately regressing each CSE component on a dummy variable for female gender and all the future trainee evaluation scores and training centre indicators for the population of candidates who completed the trainee programme. The typical estimating equation looks as follows:

$$CSE\ score_i = \tau\ female_i + \phi trainee scores_i + \beta x_i + \varepsilon_i \tag{2}$$

Although not exactly the reverse of the regressions reported in Table 4 given that the whole set of trainee evaluations substitute for the whole set of CSE scores in the set of control variables, these new calculations do highlight how each CSE component advantages or disadvantages women when this proxy for candidate (future) job performance is held constant. Because the examination structure differs slightly for external versus internal candidates, we also report separate regressions for these two groups. Comparing the coefficients for females across the various regressions reveals that, with trainee evaluations (future job performance) held constant, women significantly underperform on the common culture essay required for external candidates only, with a coefficient of -0.11 standard deviations, significant at the 5% level. Admittedly, among these same external candidates, once trainee performance is held constant, women outperform men on the language test; however, this

latter carries very little weight in the overall CSE score and only candidates who pass the written tests are admitted to the oral and language tests.

Among internal candidates, women not only underperform men by 0.11 standard deviations on the oral test (significant at the 5% level) conditional on future trainee performance (see Table 8, column (3)), but also on the language test, which, being optional for this group, carries very little weight. Because the two-stage CSE process mandates that candidates must pass the written test before taking the oral, systemic discrimination in this first stage is potentially more harmful than in the second. Indeed, comparing the results in Tables 6 and 7 confirms that external candidates are more affected by systemic discrimination than internal candidates, although the lack of any published set of pre-set questions for the orals makes it far more difficult to interpret oral test outcomes for internal candidates than the common culture essay results for external candidates.

Our identification of the common culture essay as the primary mechanism of systemic discrimination against female candidates in the CSE hiring process raises the question of whether this bias is a pure gender effect or whether female candidates in fact have a different cultural background to males from an assumedly similar society. In separate surveys administered to a subset of candidates for the 2007 and 2010 CSE, when asked about field of study, women were more likely than men to have a law degree (48 versus 36%, respectively), whereas men were more likely than women to hold a degree in social sciences such as history or political science (27 versus 15%, respectively). This higher prevalence of social sciences background over law among male versus female candidates may well explain why men outperform women on the common culture essay. It is through just such mechanisms that systemic (disparate impact) works, even when unintended. That is, if hiring depends on nonjob-related characteristics acquired in a particular context and one group is more privy than others to such acquisition, this privileged group has a greater chance of securing employment because of selection based on irrelevant characteristics. In the case of CSE scores as a hiring criterion, the ability to write the (job irrelevant) common culture essay appears more likely to have been acquired by men, perhaps via field of study preferences. Hence, just as the high school diploma in *Griggs v. Duke Power Co.* (1971) privileged Whites over Blacks, this essay improves men's chances of being hired while effectively discriminating against women, albeit unintentionally.

In fact, when we list the pairwise correlations between each CSE score and the trainee evaluations for external versus internal candidates by gender (Table 9), most correlations, although positive, are quite small, usually under 0.20. The most notable exception is the correlation between the CSE and trainee foreign language scores, which is around 0.60 in all four panels. However, as previously noted, the scores for this test carry little weight in either the overall CSE score or trainee evaluations. The next highest set of correlations, which are considerably lower at around 0.20 or below, are between the CSE oral scores and oral trainee evaluations, followed by those between the various CSE written scores and written trainee evaluations, which are usually well below 0.20 and sometimes statistically insignificant or even close to zero. The one exception is the correlations between the short-question CSE scores and certain trainee evaluations for internal candidates, which in two cases are above 0.20.[16]

Based on the above observations, we then wonder whether the systemic discrimination experienced by women through the common culture essay leads to a disparate impact in the sense that the share of women finally receiving a job in the regional administration is lower than the share of women among the initial candidate set. Unfortunately, comparing the share of female candidates who

---

[16] Given that among external candidates, the correlations between the CSE common culture essay scores and both the written and internship trainee evaluations are all smaller for men than for women, how can the evidence of women's underperformance on this essay (with future trainee performance held constant) be reconciled with this score being a stronger signal of job performance for women than for men? In fact, rather than contradicting each other, these two observations are completely separate. First, because the correlation between common culture essay scores and all trainee performance indicators is very low, the common culture essay, on which men happen to perform better than women, can be seen as providing almost no signal for future trainee performance. On the other hand, the trainee performance signal that the common culture essay does carry happens to be somewhat stronger for women than for men. In the selected sample of men and women who have passed the CSE, these observations manifest as women outperforming men on the trainee evaluations conditional on the CSE (essay) score or, equivalently, women underperforming on the CSE scores conditional on the trainee evaluations. That is, supposing that (1) men and women perform similarly on both the CSE and the trainee evaluations in an unselected sample, and (2) the CSE is completely unrelated to male trainee performance but perfectly correlated with female trainee performance, then if sample selection considered only those with a CSE score above a certain threshold, male performance will be the same on average in the selected as in the unselected sample because of our assumption that for men, CSE and trainee scores are uncorrelated. For women, however, the average performance in the selected sample will be much higher than in the larger unselected one because the correlation between female CSE scores and trainee performance scores is assumedly positive. It is this latter we observe in the data: the fact that the CSE (common culture essay) score acts as a stronger signal for job-related trainee quality for women than for men implies that, conditional on having passed the same cut-off point, women outperform men in the trainee programme (see Cameron and Trivedi 2005, p. 550, for the formula for truncated distribution means in a sample selection model). That is, *ceteris paribus*, a higher correlation between the variables determining sample selection (here, the CSE scores) and the outcome of interest (trainee job-performance score) results in the latter having a higher mean in the truncated sample, as indeed occurs in our analysis.

pass the CSE, receive a job offer, and enter the trainee programme with the share in the initial set of candidates (see Table 10) is complicated by some candidates voluntarily dropping out during various stages of the application process. Nonetheless, compared to the 1% gap between the share of women among total applicants (61%) and among training programme entrants who passed and accepted the job (60%), the equivalent gap among external candidates is somewhat larger at 2% (with 58% women among all applicants and 56% women among training programme entrants). Among internal candidates, in contrast, these shares are 64% and 66%, respectively, which is partly explainable by a lower share of women candidates exiting the application process after passing the written (61%) and oral (62%) tests. This smaller share of women recruited relative to the share of women in the initial candidate set provides evidence of a disparate impact among external candidates, although this effect is admittedly not large.

Given this evidence of a disparate impact (i.e. systemic discrimination), we then ask whether women hired at the margin outperform men hired at the margin at an even higher level than the average hired woman outperforms the average hired man. We address this question in Online Appendix Table A4, which provides estimates similar to those in Table 4 for the subsample of the 25% of examinees whose CSE score was so close to the passing threshold that they received a job offer to enter the trainee programme or were put on a waiting list but with no eventual job offer guaranteed (which accounts for several observations in the "Left after passing the oral" category in Table 3). Even though the smaller sample size leads to unacceptably larger standard errors, the results tentatively suggest that in this hired-at-the-margin subsample, the point estimates of women's outperformance of men are even larger than for the average candidate, especially in the second semester for both the written test and internship scores. Hence, although our separate estimates for external versus internal candidates remain inconclusive, this evidence on marginally hired candidates, albeit tentative, is consistent with our general results.

5 Procedural Discrimination

To test for procedural discrimination, we adopt an approach similar to that of Breda and Ly (2015), Breda and Hillion (2016), and Greenan et al. (2019) by testing for gender differences in the

non-anonymous orals once the anonymous written CSE scores are held constant. Contrary to those examined by Breda and colleagues, however, the oral and written tests we compare are not in the same subject areas. Nevertheless, this lack of anonymity provides an opportunity for discriminatory scoring based on gender. To probe for any tentative evidence in this regard, we employ the following estimating equation, where we can control for the written CSE scores, noting that the jury does not know these written CSE scores of the candidates sitting in front of them − only that they have passed the written test to be admitted to the oral:[17]

$$resultoral_i = \tau\, female_i + \rho_1 essay_i + \rho_2 shortquestions_i + \beta' x_i + \varepsilon_i \qquad (3)$$

In Table 11, we report the coefficients of female gender in the regressions of the standardized oral test scores on the written test results and other control variables, displaying first the results for all observations (row 1) and then the separate results for external (row 2) and internal candidates (row 3), respectively. Column (1) reports the raw differences between men and women, derived using models with female gender and a constant as the only regressors. Column (2) then reports the coefficients from models that include dummy indicators for training centre and external versus internal candidate, after which in column (3) we add the interactions of these variables with the year of the CSE. Lastly, column (4) lists results which also add the anonymous written test scores that we use as a benchmark (see equation (3) as well as interactions of these scores with the year of the CSE).

As the table shows, the coefficient for women is almost always positive and, once training centre is controlled for, remains statistically significant whether or not we control for the anonymous written scores. More specifically, when the anonymous written test scores are held constant, the point estimates are 0.07 standard deviations (significant at the 1% level) for all candidates, 0.11 standard deviations (significant at the 1% level) for external candidates, and 0.03 standard deviations (statistically insignificant) for internal candidates. Controlling for the written test scores does little to change the coefficient. Taken together, these results suggest that when we take the anonymous written

---

[17] Because the summary score was part of the CSE only in 2007 for external candidates, in the 2007 data, we control for it in addition to the variables mentioned in equation (2) (i.e., we set the summary score variable to zero and add in a dummy indicator for years other than 2007, as well as for internal candidates). Likewise, because the short-question test was abolished for internal candidates in the 2008 redesign, we set this score to zero for internal candidates in 2008 onwards and add in appropriate dummy indicators.

tests as a benchmark, no evidence emerges of negative discrimination against women in the CSE orals. Even though in our case the non-anonymous oral and anonymous written test topics are not the same, the evidence is still consistent with the Breda and Ly (2015), Breda and Hillion (2016) and Greenan et al. (2019) finding of a small positive procedural discrimination in favour of women. This finding holds particularly for external candidates, where women are systemically discriminated against by the essay on common culture.

6   Gender Differences by Year

To further illustrate how a hiring procedure such as the CSE might lead to systemic discrimination, we observe the female coefficient over time in separate estimates for each of our observation years from 2007 to 2012. In Figure 2A and 2B (external versus internal candidates), we plot this female coefficient for both overall trainee and oral CSE scores, conditional on all CSE scores and anonymous written CSE scores, respectively. For external candidates, we observe a downward trend of female relative to male performance in the overall conditional trainee performance but an upward trend in the conditional oral CSE scores. Although we cannot pinpoint the exact reason for this latter (given that the 2008 redesign cannot explain the post-2008 outcomes), it may be that more positive assessments on the CSE orals for women over time decreased the systemic discrimination against them, proxied here by their outperforming men on the trainee evaluations conditional on CSE scores. No such trends emerge for internal candidates, however, nor any statistically significant evidence for systemic discrimination in the year-by-year regressions − the coefficient only becomes statistically significant when all years are combined.

On the other hand, when we conduct *reverse* year-by-year regressions of the CSE scores on the future trainee scores and control variables from Table 8, we find no clear trend over time, and the 90% confidence intervals become comparatively large in the smaller annual samples (see Figure 3). Moreover, despite a slight upward trend in the point estimates, these latter are negative in 4 out of the 6 years (2007, 2008, 2011, and 2012) for the external candidate common culture essay and in 5 out of 6 years (2007, 2008, 2009, 2011, and 2012) for the internal candidate orals.

There thus seems to be tentative evidence that the systemic disadvantage of women, conditional on future trainee performance, in the CSE common culture essay for external candidates (and in the less consequential oral CSE for internal candidates) exhibits no clear time trends. The high persistence of this disadvantage pinpoints it as a likely key component of systemic discrimination against women in the CSE for external candidates, an observation that, combined with the results from Figure 2, points to changes in the CSE design as the most effective way to address systemic discrimination. At the same time, our findings overall suggest that the *Griggs v. Duke Power Co.* presumption of limited correlation between skills taught and tested in the general education system and job performance is just as relevant today as in 1971. Hence, both the public sector and private companies, rather than relying on scores from general education testing, should pay far more attention to applicant skills that are actually relevant to success and productivity in the workplace.

7   Conclusions

According to this investigation of the French civil service entry (CSE) examination for employment in the regional public administration (*instituts régionaux d'administration*, IRA), women outperform men on both the anonymous written trainee evaluations and the non-anonymous internship programme reports. Even more notable, this outperformance still holds when we control for the CSE scores obtained through both a set of anonymous entry tests and a non-anonymous oral. We thus conclude that the CSE exhibits systemic discrimination against women through its use of recruitment criteria that slightly favour men but are not necessarily relevant for actual job performance.

This systemic discrimination applies specifically to external female candidates (those not yet working in the civil service), who, when the CSE scores are held constant, outperform males on total trainee evaluation scores by 0.24 standard deviations, compared with 0.13 standard deviations for internal candidates (with 4 years in the service). We find a partial explanation of the CSE's unsuitability for identifying the best job performers among external candidates when we run reverse regressions of CSE scores on gender, institutional controls, and trainee evaluations scores, our proxy for future job performance. When this job performance proxy is held constant, women mainly

20

underperform on the CSE common culture essay, a minimally job-related exercise required only of external candidates.

Although the overall effect of the systemic discrimination is not very large, its very presence is still surprising given an apparent lack of procedural discrimination in the CSE itself, which we assess by analysing female versus male performance on the non-anonymous CSE oral test while controlling for the anonymous written test scores. Conditional on these latter, the CSE orals, which resemble a job interview, show no apparent bias against women but instead a small bias against men. Hence, although the oral and written tests are different in content, benchmarking the non-anonymous oral scores against the anonymous written results indicates that the CSE is procedurally fair towards women or even positively discriminative in their favour to the disadvantage of men, albeit to a small extent. An alternative explanation for better female performance on the orals could be systemic discrimination against women on the written test, meaning that among all those who passed the written test, a higher share of women is suited to the job, which suitability is reflected during the job-interview-style orals.

Overall, then, our empirical evidence suggests that recruiters should pay far greater attention to their hiring criteria, as excellence on a competitive examination that does not measure job-requisite skills does not imply excellence in job performance. Because the civil service entry procedure investigated places great emphasis on fair competition with a large degree of anonymity, our finding of possible systemic discrimination even within this system has important implications for all hiring processes. In particular, our identification of an "essay on common culture" requirement as the probable tool of (probably unintended) systemic discrimination indicates that cultural background, although potentially a subtle gatekeeper in hiring or promotion, is not necessarily a signal of productivity on the job. In fact, some situations that look like gender or ethnic discrimination may at least in part be cases of cultural discrimination.

References

Apesteguia, J, Azmat G, Iriberri N (2012) The impact of gender composition on team performance and decision making: Evidence from the field, *Management Sci.* 58: 78–93.

Autor, D H, Scarborough D (2008) Does job testing harm minority workers? Evidence from retail establishments. *Quart. J. Econom.* 123: 219–277.

Bagues, MF, Esteve-Volart B (2010) Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. *Rev. Econom. Stud.* 77: 1301–1328.

Bohnet, I, van Geen A, Bazerman M (2016) When performance trumps gender bias: Joint vs. separate evaluation. *Management Sci.* 62: 1225–1234.

Bornholz R, Heckman JJ (2005) Measuring Disparate Impacts and Extending Disparate Impact Doctrine to Organ Transplantation, *Perspect. Biol. Med.* 48 Supplement: S95-S122.

Bourdieu, P. Passeron J-C (1964) *Les héritiers. Les étudiants et la culture.* (Les Editions de Minuit, Paris).

Bourdieu, P. (1979) *La distinction, critique sociale du jugement.* (Les Editions de Minuit, Paris).

Breda T, Hillion M (2016) Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in France. *Science* 353: 474–478.

Breda T, Ly ST (2015) Professors in core science fields are not always biased, against women: Evidence from France. *Amer. Econom. J.: Appl. Econom.* 7: 53–75.

Cameron AC, Trivedi PK (2005) *Microeconometrics, Methods and Applications* (Cambridge University Press, Cambridge, UK).

Eberhard M, Meurs D, Simon P (2008) Construction d'une méthodologie d'observation de l'accès et du déroulement de carriére des generations issues de l'immigration dans la fonction publique [Construction of a methodology for observing the access and career development of generations of immigrants in the public service]. Final report, National Institute of Demographic Studies (INED), Paris.

Goldin C, Rouse C (2000) Orchestrating impartiality: The impact of "blind" auditions on female musicians. *Amer. Econom. Rev.* 90: 715–741.

Greenan N, Lanfranchi J, L'Horty Y, Narcy M, Pierné G (2019) Do competitive examinations promote diversity in civil service? *Public Adm. Rev.* 79: 370-382.

Kolev J, Fuentes-Medel Y, Murray F (2019) Is blinded review enough? How gendered outcomes arise even under anonymous evaluation, NBER Working Paper No. 25759, Cambridge, MA.

McCrary, J. (2007) The effect of court-ordered hiring quotas on the composition and quality of police. *Amer. Econom. Rev.* 97: 318–353.

Meurs, D. (2019) *Les épreuves écrites des concours de la fonction publique: une barrière sélective à l'entrée*, Fondation Jean Jaurès, Rubrique "Société", Paris.

OECD (2019) *Government at a Glance 2019* (OECD Publishing, Paris),
       https://doi.org/10.1787/8ccf5c38-en.

Peresie JL (2009) Toward a coherent test for disparate impact discrimination. *Indiana Law J.* 84:
       773–802.

Selmi M (2006) Was the disparate impact theory a mistake? *UCLA Law Rev.* 53: 701–782.

Sutter M, Glätzle-Rützler D (2015) Gender differences in the willingness to compete emerge early in
       life and persist. *Management Sci.* 61: 2339–2354.

Table 1
CSE Structure Over Time

	2007	2008	2009	2010	2011	2012
Internal candidate						
Dossier [factor 4]	x	x	x	x	x	x
Ten short questions [factor 5, since 2008: 0]	x					
Oral [factor 4]	xx	x	x	x	x	x
Language [factor 1]	o	o	o	o	o	o
External candidate						
Common culture essay [factor 4]	x	x	x	x	x	x
Ten short questions [factor 5, since 2008: 4]	x	x	x	x	x	x
Administrative summary [factor 3]	x					
Oral [factor 4]	xx	x	x	x	x	x
Language [factor 1]	x	x	x	x	x	x

Notes: x = the test was a mandatory component of the CSE in that year; xx = the 2007 oral test, which included a set of questions on a text studied 30 minutes previously in addition to the standard job interview (since 2008, a job interview only); o = an optional language test for internal candidates. No entry means that the component was abolished in 2008 (e.g. the short-question test for internal candidates and the administrative summary for external candidates).
Source: Communication with the *instituts régionaux d'administration*; for the most recent CSE structure, see https://www.fonction-publique.gouv.fr/score/ecoles-de-formation/ira-et-ena/ira/information-sur-dossier-raep.

Table 2
Structure of Trainee Evaluations Over Time

	2007	2008	2009	2010	2011	2012
Both internal and external candidates						
First semester						
Written administrative test [factor 5]	x	x	x	x	x	x
Internship grade [factor 7]	x	x	x	x	x	x
Orals – internship based [factor 5]	x	x	x	x	x	x
Defense [factor 3]	x	x	x	x	x	x
Second semester						
Written administrative test [factor 5]	x	x	x	x	x	x
Internship grade [factor 5]	x	x	x	x	x	x
Orals – field of specialization [factor 5]	x	x	x	x	x	x
Foreign language professional topic [factor 1]	x	x	x	x	x	x

Notes: x = the test was a mandatory component of the trainee evaluations; the trainee exams in the column headed "2007" took place in 2009. "2007" here refers to the year the recruitment process started with the CSE described in Table 1.
Source: Communication with the *instituts régionaux d'administration* and web pages: https://ira-bastia.gouv.fr/presentation-de-la-formation/; https://www.ira-lille.gouv.fr/formation-initiale/; https://www.ira-lyon.gouv.fr/; https://www.ira-metz.gouv.fr; https://www.ira-nantes.gouv.fr/index.php/presentation-formation-initiale/.

Table 3
CSE Candidate Outcomes

	2007	2008	2009	2010	2011	2012
Failed the written test	2,615	4,496	5,614	5,343	5,209	4,990
Left after passing the written	182	121	119	130	103	77
Failed the oral test	610	649	550	697	738	762
Left after passing the oral	243	284	209	217	168	116
Passed and accepted the job	674	566	605	599	575	589
Total number of candidates	4,324	6,116	7,097	6,986	6,793	6,534

Notes: The category "Left after passing the oral" also includes candidates who were put on a waiting list after the CSE orals and then either left or were not offered a position in the trainee programme.
Source: Administrative data from the French *instituts régionaux d'administration*; author calculations.

Table 4
Female Coefficient for the Trainee Scores Using Different Specifications: All Candidates

Dependent Variable	No controls		Controls: year, training centre and internal/external candidate		Controls: job entry exam scores (also interacted with year)	
	Coeff.	(s.e.)	Coeff.	(s.e.)	Coeff.	(s.e.)
Total trainee score	0.16***	(0.03)	0.20***	(0.03)	0.19***	(0.03)
Total 1st semester	0.11***	(0.03)	0.14***	(0.03)	0.14***	(0.03)
Total 2nd semester	0.17***	(0.03)	0.20***	(0.03)	0.18***	(0.03)
Written 1st semester	0.15***	(0.03)	0.16***	(0.03)	0.16***	(0.03)
Written 2nd semester	0.20***	(0.03)	0.21***	(0.03)	0.20***	(0.03)
Internship 1st semester	0.17***	(0.03)	0.18***	(0.03)	0.18***	(0.03)
Internship 2nd semester	0.15***	(0.03)	0.16***	(0.03)	0.16***	(0.03)
Orals 1st semester	-0.04	(0.03)	0.00	(0.03)	0.01	(0.03)
Orals 2nd semester	0.03	(0.03)	0.05	(0.03)	0.06*	(0.03)
Defense 1st semester	-0.01	(0.03)	-0.01	(0.03)	0.00	(0.03)
Language 2nd semester	-0.01	(0.03)	0.05	(0.03)	0.01	(0.03)
Left after CSE before training	-0.02	(0.01)	-0.01	(0.01)	-0.01	(0.01)

Notes: The estimates are based on 3,608 candidates from the 2007–2012 CSE cohorts who passed the entire exam and attended the trainee programme, 2,173 of whom were women and 1,435 were men. One exception is the estimates displayed in the last row as "Left after the CSE before training", which are based on 4,849 observations. ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively; numbers without asterisks are not statistically significant even at the 10% level.
Source: Administrative data from the French *instituts régionaux d'administration*; author calculations.

Table 5
Female Coefficient for the Trainee Scores Using Different Specifications with the Third Order
Polynomial of the CSE Score as Control: All Candidates

Dependent Variable	Linear Controls: job entry exam scores (also interacted with year)		Linear Control: total job entry exam score (also interacted with year)		Third-order Polynomial Controls: total job entry exam score (also interacted with year)	
	Coeff.	(s.e.)	Coeff.	(s.e.)	Coeff.	(s.e.)
Total trainee score	0.19***	(0.03)	0.22***	(0.03)	0.21***	(0.03)
Total 1st semester	0.14***	(0.03)	0.16***	(0.03)	0.16***	(0.03)
Total 2nd semester	0.18***	(0.03)	0.21***	(0.03)	0.21***	(0.03)
Written 1st semester	0.16***	(0.03)	0.17***	(0.03)	0.17***	(0.03)
Written 2nd semester	0.20***	(0.03)	0.21***	(0.03)	0.21***	(0.03)
Internship 1st semester	0.18***	(0.03)	0.19***	(0.03)	0.19***	(0.03)
Internship 2nd semester	0.16***	(0.03)	0.17***	(0.03)	0.17***	(0.03)
Orals 1st semester	0.01	(0.03)	0.02	(0.03)	0.02	(0.03)
Orals 2nd semester	0.06*	(0.03)	0.07**	(0.03)	0.07**	(0.03)
Defense 1st semester	0.00	(0.03)	0.00	(0.03)	0.00	(0.03)
Language 2nd semester	0.01	(0.03)	0.06**	(0.03)	0.06**	(0.03)

Notes: The estimates are based on 3,608 candidates from the 2007–2012 CSE cohorts who passed the entire exam and attended the trainee programme, 2,173 of whom were women and 1,435 were men. ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively; numbers without asterisks are not statistically significant even at the 10% level.
Source: Administrative data from the French *instituts régionaux d'administration*; author calculations.

Table 6
Female Coefficient for the Trainee Scores under Different Specifications: External Candidates

Dependent Variable	No controls		Controls: year, training centre and internal/external		Controls: job entry exam scores (also interacted with year)		Difference in Female Coefficient Between External and Internal Candidates	
	Coeff.	(s.e.)	Coeff.	(s.e.)	Coeff.	(s.e.)	Coeff.	(s.e.)
Total trainee score	0.29***	(0.04)	0.29***	(0.04)	0.24***	(0.04)	0.12*	(0.06)
Total 1st semester	0.21***	(0.04)	0.20***	(0.04)	0.17***	(0.04)	0.07	(0.06)
Total 2nd semester	0.27***	(0.04)	0.28***	(0.04)	0.23***	(0.04)	0.11	(0.07)
Written 1st sem.	0.24***	(0.04)	0.25***	(0.04)	0.22***	(0.04)	0.17***	(0.07)
Written 2nd sem.	0.26***	(0.04)	0.27***	(0.04)	0.24***	(0.04)	0.12*	(0.07)
Internship 1st sem.	0.22***	(0.04)	0.20***	(0.04)	0.19***	(0.04)	0.02	(0.06)
Internship 2nd sem.	0.19***	(0.04)	0.19***	(0.04)	0.17***	(0.04)	0.03	(0.06)
Orals 1st semester	0.07	(0.04)	0.07	(0.04)	0.04	(0.04)	0.05	(0.07)
Orals 2nd semester	0.10**	(0.04)	0.10**	(0.04)	0.09**	(0.04)	0.08	(0.07)
Defense 1st sem.	-0.04	(0.04)	-0.03	(0.04)	-0.03	(0.04)	-0.07	(0.07)
Language 2nd sem.	0.13***	(0.04)	0.13***	(0.04)	0.00	(0.03)	-0.03	(0.06)

Notes: The estimates are based on 2,092 external candidates from the 2007–2012 CSE cohorts who passed the entire exam and attended the trainee programme, 1,170 of whom were women and 922 were men. The last column of results states the difference between the third column of results in Table 6 (external candidates) and Table 7 (internal candidates) together with its standard error, derived from a joint estimation for both groups with appropriate interaction terms. ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively; numbers without asterisks are not statistically significant even at the 10% level.
Source: Administrative data from the French *instituts régionaux d'administration*; author calculations.

Table 7
Female Coefficient for the Trainee Scores under Different Specifications: Internal Candidates

Dependent Variable	No controls		Controls: year, training centre and internal/external		Controls: job entry exam scores (also interacted with year)	
	Coeff.	(s.e.)	Coeff.	(s.e.)	Coeff.	(s.e.)
Total trainee score	0.08	(0.06)	0.08	(0.05)	0.13***	(0.05)
Total 1st semester	0.05	(0.06)	0.06	(0.05)	0.10**	(0.05)
Total 2nd semester	0.09	(0.06)	0.09	(0.05)	0.12**	(0.05)
Written 1st semester	0.03	(0.06)	0.05	(0.06)	0.05	(0.05)
Written 2nd semester	0.13**	(0.06)	0.13**	(0.06)	0.12**	(0.05)
Internship 1st semester	0.14**	(0.06)	0.15**	(0.05)	0.17***	(0.05)
Internship 2nd semester	0.11*	(0.06)	0.11*	(0.06)	0.14***	(0.05)
Orals 1st semester	-0.08	(0.05)	-0.06	(0.05)	-0.01	(0.05)
Orals 2nd semester	-0.01	(0.06)	-0.01	(0.06)	0.01	(0.05)
Defense 1st semester	0.04	(0.05)	0.02	(0.05)	0.04	(0.05)
Language 2nd semester	-0.07	(0.05)	-0.07	(0.05)	0.03	(0.04)

Notes: The estimates are based on 1,516 internal candidates from the 2007–2012 CSE cohorts who passed the entire exam and attended the trainee programme, 1,003 of whom were women and 513 were men. ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively; numbers without asterisks are not statistically significant even at the 10% level.
Source: Administrative data from the French *instituts régionaux d'administration*; author calculations.

Table 8
Female Coefficient for the Regressions of the CSE Scores on the Trainee Scores and Control Variables

	All candidates		External candidates		Internal candidates	
	Coeff.	(s.e.)	Coeff.	(s.e.)	Coeff.	(s.e.)
Essay/dossier score	-0.07*	(0.03)	-0.11**	(0.04)	-0.01	(0.05)
# observations	3,608		2,092		1,516	
Short questions score	-0.02	(0.04)	-0.03	(0.04)	-0.07	(0.12)
# observations	2,368		2,092		276	
Admin. summary score	0.05	(0.10)	0.05	(0.10)	-	-
# observations	398		398		-	-
Oral score	-0.05	(0.03)	-0.02	(0.04)	-0.11**	(0.05)
# observations	3,608		2,092		1,516	
Language score	0.05	(0.03)	0.12***	(0.03)	[-0.11*]	[(0.06)]
# observations	2,835		2,092		[743]	

Notes: All scores are standardized. The language results for internal candidates are reported in brackets, because the language exam is optional for this group. ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively; numbers without asterisks are not statistically significant even at the 10% level; author calculations.
Source: Administrative data from the French *instituts régionaux d'administration*; author calculations.

Table 9
Test-By-Test Correlation Between Trainee Performance Scores and CSE Scores
Panel A: External Candidates – Men

	Essay	Short questions	Administrative Summary	Oral	Language
Written 1st semester	0.09***	0.04	0.00	0.04	0.14***
Written 2nd semester	0.02	0.08**	0.18**	0.03	0.05
Internship 1st semester	-0.04	0.07**	-0.14*	0.07**	0.12***
Internship 2nd semester	-0.02	0.12***	-0.04	0.08**	0.13***
Orals 1st semester	0.00	0.10***	0.05	0.12***	0.07**
Orals 2nd semester	0.03	0.02	-0.03	0.18***	0.06*
Defense 1st semester	0.06*	-0.05	0.13*	0.00	-0.02
Language 2nd semester	-0.03	0.04	-0.09	0.06*	0.59***
Panel B: External Candidates – Women					
Written 1st semester	0.16***	0.13***	-0.04	0.04	0.10***
Written 2nd semester	0.06**	0.07**	0.16**	0.02	0.07**
Internship 1st semester	0.07**	0.08***	0.02	0.06**	0.15***
Internship 2nd semester	0.13***	0.12***	0.00	0.04	0.12***
Orals 1st semester	0.07**	0.08***	-0.03	0.15***	0.15***
Orals 2nd semester	0.06*	0.03	-0.04	0.16***	0.15***
Defense 1st semester	0.09***	-0.01	0.02	-0.04	0.04
Language 2nd semester	0.08***	0.05	-0.07	-0.03	0.60***
Panel C: Internal Candidates – Men					
Written 1st semester	0.07*	0.07	-	-0.02	0.13**
Written 2nd semester	0.10*	0.16*	-	0.05	0.06
Internship 1st semester	0.00	-0.09	-	0.13***	0.08
Internship 2nd semester	0.05	0.07	-	0.11**	0.09
Orals 1st semester	0.01	0.24***	-	0.16***	0.10*
Orals 2nd semester	0.00	-0.01	-	0.17***	0.11*
Defense 1st semester	0.01	0.08	-	-0.10**	0.06
Language 2nd semester	-0.01	-0.09	-	0.08*	0.57***
Panel D: Internal Candidates – Women					
Written 1st semester	0.09**	0.02	-	0.06	0.12**
Written 2nd semester	0.12***	0.23***	-	0.02	0.15***
Internship 1st semester	-0.01	0.07	-	0.10***	0.11**
Internship 2nd semester	0.02	0.03	-	0.11***	0.09
Orals 1st semester	-0.02	0.19**	-	0.20***	0.17***
Orals 2nd semester	-0.04	0.18**	-	0.22***	0.12***
Defense 1st semester	0.04	0.03	-	-0.08**	0.01
Language 2nd semester	0.02	0.11	-	0.02	0.63***

Table 10
Share of Women by CSE Outcome

	All	External	Internal
Failed the written test	61	59	63
Left after passing the written	58	57	61
Failed the oral test	58	53	64
Left after passing the oral	58	56	62
Passed and accepted the job	60	56	66
Total	61	58	64
Total number of candidates	37,850	18,905	18,945

Notes: Shares of women in the respective categories are given in percentages.
Source: Administrative data from the French *instituts régionaux d'administration*; author calculations.

Table 11
Female Coefficient for the Regressions of the Non-anonymous Oral Test Scores on the Anonymous Written Test Scores and Control Variables

	No controls		Controls: training centre and internal/external		Controls: inter-actions of year and training centre, internal/ external		Controls: year and exam score inter-actions	
	Coeff.	(s.e.)	Coeff.	(s.e.)	Coeff.	(s.e.)	Coeff.	(s.e.)
All observations	0.03	(0.02)	0.05**	(0.02)	0.07***	(0.02)	0.07***	(0.02)
# observations	8,838		8,838		8,838		8,838	
External cand.	0.08***	(0.03)	0.09***	(0.03)	0.09***	(0.03)	0.11***	(0.03)
# observations	5,270		5,270		5,270		5,270	
Internal cand.	-0.01	(0.04)	0.02	(0.04)	0.02	(0.04)	0.03	(0.04)
# observations	3,568		3,568		3,568		3,568	

Notes: All scores are standardized. **, and * indicate statistical significance at the 5 and 10% levels, respectively; numbers without asterisks are not statistically significant even at the 10% level.
Source: Administrative data from the French *instituts régionaux d'administration*; author calculations.
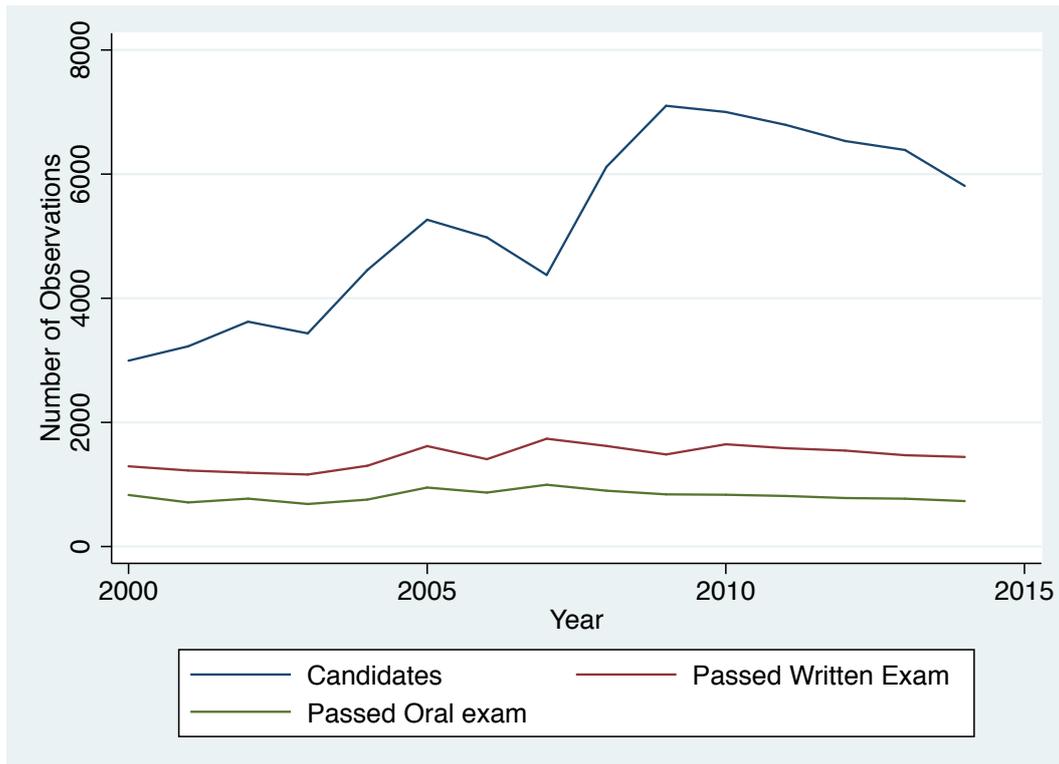
Fig. 1. Number of Candidates, Number Passing Written Test, Number Passing Orals, 2000–2014
Source: Administrative data from the French *instituts régionaux d'administration*.

A. External Candidates
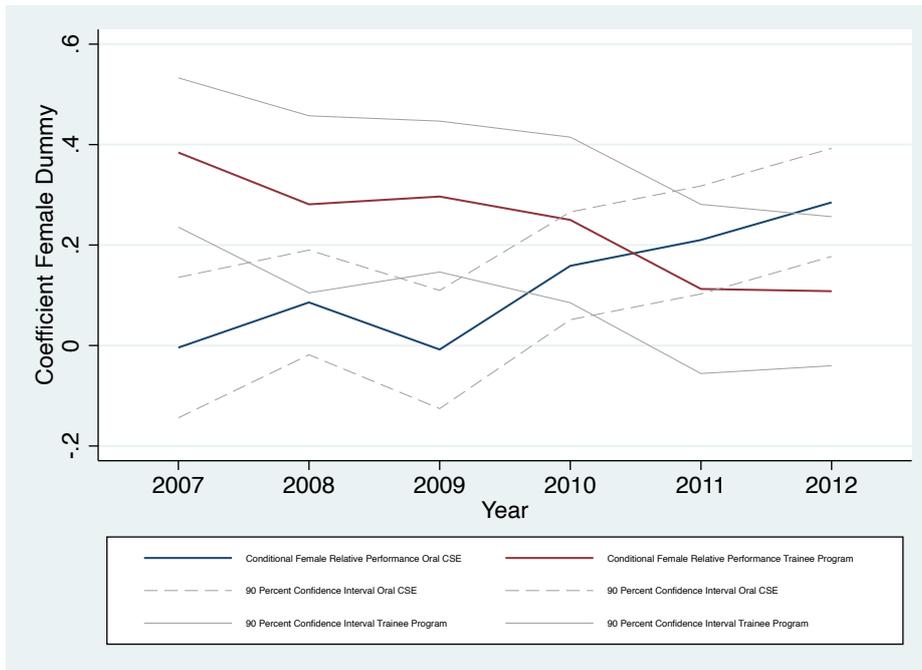
B. Internal Candidates
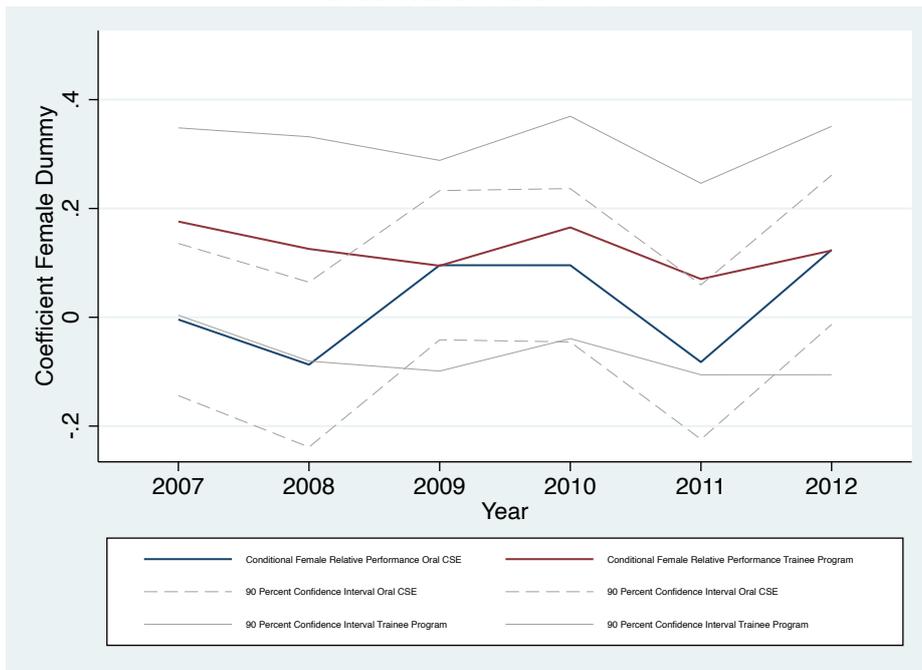
Fig. 2. Female Coefficient Over Time in CSE and Trainee Score Regressions
Source: Administrative data from the French *instituts régionaux d'administration*.
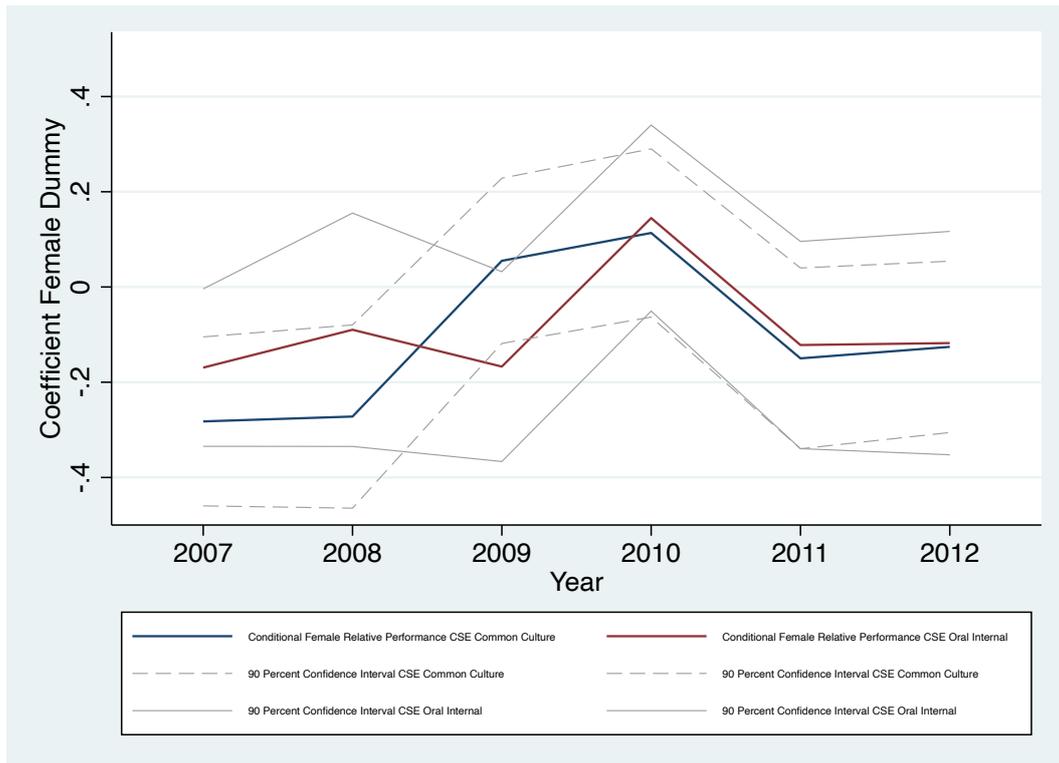
Fig. 3. Female Coefficient Over Time in the Regressions of Selected CSE Scores on the Trainee Scores and Control Variables

Source: Administrative data from the French *instituts régionaux d'administration*.

Online Appendix

Table A1
Trainee Sample Means

	External			Internal		
	Women		Men	Women		Men
Trainee evaluations						
Total trainee score	0.13	***	-0.16	0.03		-0.05
Total 1st semester	0.09	***	-0.12	0.02		-0.04
Total 2nd semester	0.12	***	-0.16	0.03		-0.06
Written 1st semester	0.11	***	-0.14	0.01		-0.02
Written 2nd semester	0.12	***	-0.15	0.04	**	-0.09
Internship 1st semester	0.10	***	-0.12	0.04	**	-0.09
Internship 2nd semester	0.09	***	-0.11	0.03	*	-0.07
Orals 1st semester	0.03		-0.04	-0.02		0.04
Orals 2nd semester	0.05	**	-0.06	0.00		0.00
Defense 1st semester	-0.02	***	0.02	0.01		-0.02
Language 2nd semester	0.06	***	-0.08	-0.02		0.04
CSE scores						
CSE Oral score	0.01		-0.01	-0.03	*	0.06
CSE Language score	0.10	***	-0.13	-0.03		0.04
CSE Share no language	0.00		0.00	0.54	***	0.46
CSE Essay/Dossier score	0.04	**	0.05	0.00		0.01
CSE Short-question test score	0.00		0.00	-0.03		0.03
CSE Administrative summary score	0.03		-0.04	-		-
Internal	0.00		0.00	1.00		1.00
Training Centre						
Bastia	0.21	*	0.18	0.20		0.22
Lille	0.20		0.21	0.19		0.20
Lyon	0.18		0.21	0.20		0.20
Metz	0.20		0.22	0.22		0.18
Nantes	0.21		0.19	0.19		0.20
Year						
2007	0.20		0.18	0.15	***	0.25
2008	0.16		0.16	0.16		0.15
2009	0.17		0.16	0.16		0.18
2010	0.15	**	0.18	0.18	**	0.14
2011	0.17		0.15	0.17		0.15
2012	0.16		0.16	0.18	***	0.13
Observations	1,170		922	1,003		513

Notes: All test scores are standardized on the sample of candidates participating in the trainee programme. Because of differences between external and internal candidates and 2008 changes in the CSE structure, the numbers of observations differ from those stated at the bottom of the table for the CSE language score for internal candidates (465 and 278 for women and men, respectively), for the short-question test for internal candidates (150 and 126 for women and men, respectively), and for the administrative summary for external candidates (232 and 166 for women and men, respectively). ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively, for the test of equality of means between women and men for the respective variable; means without asterisks are not statistically significant between men and women, even at the 10% level.
*Source*: Administrative data from the French *instituts régionaux d'administration*.

Table A2
CSE Candidate Sample Means for Candidates Having Passed the Written and Taking the Oral Tests

	External			Internal		
	Women		Men	Women		Men
Passed oral test	0.57	**	0.54	0.54		0.53
CSE Oral score	0.09	***	0.01	-0.08		-0.07
CSE Language score	0.07	***	-0.08	-0.01		0.02
CSE Share no language	0.00		0.00	0.57	***	0.49
CSE Essay/Dossier score	0.98		1.01	1.28	***	1.18
CSE Short-question test score	0.93	***	0.99	0.84	***	1.06
CSE Administrative summary score	0.71	***	0.55	-		-
Internal	0.00		0.00	1.00		1.00
Training Centre						
Bastia	0.22	***	0.19	0.19	***	0.23
Lille	0.20		0.20	0.21		0.21
Lyon	0.21		0.19	0.21		0.19
Metz	0.17	***	0.21	0.20		0.17
Nantes	0.20		0.21	0.20	**	0.20
Year						
2007	0.19	**	0.17	0.13	***	0.22
2008	0.17		0.17	0.18	**	0.15
2009	0.16		0.14	0.15	**	0.18
2010	0.15	***	0.19	0.19	***	0.15
2011	0.18	**	0.15	0.18	**	0.15
2012	0.15	***	0.18	0.18		0.16
	2,874		2,396	2,319		1,249

Notes: All test scores are standardized on the sample of candidates taking the written tests, with the exception of the oral test score, which is standardized based on the sample of candidates taking it. Because of differences between external and internal candidates and 2008 changes in the CSE structure, the numbers of observations differ from those stated at the bottom of the table for the CSE language score for internal candidates (1,005 and 642 for women and men, respectively), for the short-question test for internal candidates (301 and 269 for women and men, respectively), and for the administrative summary for external candidates (554 and 407 for women and men, respectively). ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively, for the test of equality of means between women and men for the respective variable; means without asterisks are not statistically significant between men and women, even at the 10% level.
Source: Administrative data from the French *instituts régionaux d'administration*.

Table A3
CSE Candidate Sample Means for Candidates Taking the Written Tests

	External			Internal		
	Women		Men	Women		Men
CSE Essay/Dossier score	-0.05	***	0.04	0.06	***	-0.06
CSE Short-question test score	-0.06	***	0.06	-0.03	***	0.35
CSE Administrative summary score	0.07	***	-0.10	-		-
Internal	0.00		0.00	1.00		1.00
Training Centre						
Bastia	0.20		0.20	0.19	***	0.20
Lille	0.19		0.19	0.21		0.21
Lyon	0.22	***	0.20	0.21	***	0.19
Metz	0.17	***	0.19	0.20		0.19
Nantes	0.22		0.22	0.20		0.21
Year						
2007	0.16	*	0.15	0.07	***	0.08
2008	0.16	**	0.15	0.16	*	0.17
2009	0.18		0.18	0.20		0.19
2010	0.17	**	0.18	0.20		0.19
2011	0.16		0.17	0.20	*	0.19
2012	0.17		0.17	0.18		0.17
	10,911		7,997	12,040		6,908

Notes: All test scores are standardized on the sample of candidates taking the written tests. Because of differences between external and internal candidates and 2008 changes in the CSE structure, the numbers of observations differ from those stated at the bottom of the table for the short-questions test for internal candidates (790 and 573 for women and men, respectively), and for the administrative summary for external candidates (1,755 and 1,210 for women and men, respectively). ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively, for the test of equality of means between men and women for the respective variable; means without asterisks are not statistically significant between women and men, even at the 10% level.
Source: Administrative data from the French *instituts régionaux d'administration*.

Table A4
Female Coefficient for the Trainee Scores Using Different Specifications: 25% of Candidates above the CSE Pass Threshold

Dependent Variable	No controls		Controls: year, training centre, and internal/external		Controls: job entry examination scores (also interacted with year)	
	Coeff.	(s.e.)	Coeff.	(s.e.)	Coeff.	(s.e.)
Total trainee score	0.27***	(0.09)	0.26***	(0.09)	0.26***	(0.09)
Total 1st semester	0.09	(0.09)	0.04	(0.09)	0.07	(0.09)
Total 2nd semester	0.31***	(0.09)	0.33***	(0.09)	0.32***	(0.10)
Written 1st semester	0.10	(0.09)	0.07	(0.09)	0.07	(0.09)
Written 2nd semester	0.24***	(0.09)	0.28***	(0.09)	0.23**	(0.09)
Internship 1st semester	0.16*	(0.09)	0.11	(0.08)	0.13	(0.08)
Internship 2nd semester	0.41***	(0.10)	0.33***	(0.11)	0.34***	(0.11)
Orals 1st semester	-0.06	(0.09)	-0.08	(0.09)	-0.07	(0.09)
Orals 2nd semester	0.09	(0.09)	0.13	(0.09)	0.15	(0.10)
Defense 1st semester	0.08	(0.09)	0.03	(0.09)	0.07	(0.09)
Language 2nd semester	-0.04	(0.08)	0.01	(0.08)	0.01	(0.08)

Notes: The estimates are based on 562 candidates from the 2007–2012 CSE cohorts who passed the entire exam and attended the trainee programme, 351 of whom were women and 211 were men. The observations selected for these estimates were selected by year and training centre to be the 25% with the lowest scores among the group of candidates who are above the passing threshold in the respective year and training centre, where above the threshold may imply being put on a waiting list. The number of observations here is below a quarter of the observations used in the main part of the paper, because some candidates decide not to join the trainee programme and some candidates on the waiting list are not offered a place in the trainee programme. ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively; numbers without asterisks are not statistically significant even at the 10% level.
Source: Administrative data from the French *instituts régionaux d'administration*; author calculations.