Do Recruiters Prefer Applicants with Similar Skills? Evidence from a Randomized Natural Experiment[*]

Manuel Bagues[†]   Maria J. Perez-Villadoniga[‡]

Universidad Carlos III   Universidad de Oviedo

October 13, 2011

Abstract

In this paper we explore whether recruiters prefer applicants who are relatively strong in the skills in which the recruiters themselves excel. We analyze evidence from all entry exams to the Spanish Judiciary held between 2003 and 2007, where applicants are randomly assigned across evaluation committees. We find that applicants who excel in the same dimensions as recruiters are significantly more likely to be hired. Our findings have important strategic implications for both public and private sector recruitment practices.

Keywords: hiring; similar-to-me effect; randomized natural experiment.

JEL Classification: M51, J2.

"It is merely human nature that we overrate the importance of our own types

of research and underrate the importance of the types that appeal to others."

J.A. Schumpeter, *History of Economic Analysis*,

London: Allen and Unwin, 1954.

1   Introduction

What influences employers' hiring decisions? The answer to this question has been discussed in the economics, management and psychology literatures for several decades. One issue that has attracted particular attention is the potential existence of a "similar-to-me" effect. According to this idea, evaluators tend to select candidates who are similar to themselves in some respect. Several authors have studied the effect of similarity in dimensions such as race (Prewett-Livingston et al. 1996, Stoll et al. 2004, Parsons et al. 2008, Price and Wolfers 2010) or gender (Broder 1993, Dillingham et al. 1994, Antonovics et al. 2005, Bagues and Esteve-Volart 2010).

Less attention has been devoted to similarity in terms of skills.[1] In this paper we test for the existence of a similar-to-me effect in skills using evidence from a real labor market selection process: the Spanish entry exams to judge and prosecutor positions. This selection process is of great importance both in terms of the number of candidates involved –about 4000 each year– and in terms of the relevance of the jobs at stake: successful candidates become judges or prosecutors for life. The structure of the selection process provides a convenient setup to test for the existence of a similar-to-me effect in skills. First, candidates take a qualifying multiple choice test. Then, successful candidates are randomly allocated across several evaluation committees and they are orally evaluated on their knowledge in several areas. The composition of committees varies in terms of the field of specialization of their members, and thus it allows us to

---

[1]An exception is the work of Hamermesh and Schmidt (2003), who examine the election of Fellows of the Econometric Society. Conditional on several objective measures of quality, the field of specialization affects significantly the probability of election. In particular, theorists are more likely to be elected than econometricians. The authors hypothesize that a potential explanation for this result is that maybe theorists at high-prestige institutions make up a large fraction of the electorate and are happier to vote for candidates like themselves.

study how the expertise of committee members affects selection decisions.

We find that recruiters tend to hire candidates that are relatively more similar to themselves in terms of their skills. In quantitative terms, we observe that committees whose knowledge of a field is above the median tend to hire candidates that are on average one tenth of a standard deviation more knowledgeable in that field – as measured by their performance in a preliminary multiple choice test.

Two types of hypothesis are consistent with the observed evidence. First, there might be a preference-based explanation. According to the similarity-attraction paradigm, there is a strong association between similarity to oneself and interpersonal attraction, and this association might affect selection decisions (Byrne 1971). As Graves and Powell (1995, p. 86) put it, "perceived similarity in attitudes and values [which in turn] leads to interpersonal attraction between the recruiter and the applicant (...) then leads to positive bias in the recruiter's interview conduct." Similarly, the attraction-selection-attrition theory (Schneider 1987), one the main theoretical foundations of the person-organization fit literature, posits that organizations attract, select, and retain similar types of people. Second, the evidence may reflect the existence of information asymmetries about the real productivity of workers. As Cornell and Welch (1996, p. 544) point out, if it "is easier for individuals to screen people of similar background (...) people (will) tend to hire others of their own type even when they have no innate preference for similar people". While we are not able to disentangle between the two hypotheses, our results unambiguously suggest that evaluation committees should be balanced.

The structure of the paper is as follows. In section 2 we provide background information about hiring procedures in the Spanish Judiciary. In section 3, we present the data and, in section 3, the empirical analysis. Finally, section 4 concludes and briefly discusses potential implications for the recruiting practices of firms.

2 Background

In Spain, candidates to judge and prosecutor positions must pass a public examination at the national level. This selection process is held yearly for both positions jointly. The structure of the process is as follows (see Table 1).[2] The exam consists of three qualifying stages: a preliminary multiple choice test and two oral tests. The multiple choice test contains one hundred questions, each one listing a set of four possible answers. For every correct answer the candidate receives one mark. If the answer is incorrect, he loses 0.33 marks. If the question is left unanswered, the candidate gets zero marks.

Candidates whose grade is above the minimum passing threshold qualify for the first oral test and are assigned to evaluation committees. The minimum passing grade is decided each year by the Selection Committee based on the number of evaluation committees that are available. The allocation of candidates to committees is done according to the following process. Candidates are ranked in alphabetical order and committees are ranked numerically. A lottery decides the initial according to which the alphabetical list of candidates will be matched with the list of committees.

In the first oral test candidates must answer orally, within 75 minutes, five randomly drawn questions. Once the candidate has finished answering these questions, the committee decides whether the candidate qualifies for the second oral test. The decision is made on a majority basis; in case of tie, the president of the committee decides. Every successful candidate is given a numerical grade, which is computed by adding the grades proposed by each member of the committee after excluding the minimum and the maximum grades. The structure of the second oral test is similar. Candidates who manage to pass this test can then choose between becoming judges or prosecutors according to their final ranking.[3] A candidate's final grade is the sum of the grades

---

[2]The evaluation process is described in detail in each call. See, for example, "Agreement of May 8 2006 of the Selection Commission", BOE n. 3, May 13, 2006, p. 18510. A shorter description of the process is also available in Bagues and Esteve-Volart (2010)

[3]Formally, passing all stages does not guarantee obtaining a position as committees cannot deliver more positions than the number of positions that was initially assigned to them. However, this constraint is rarely binding, in the period considered here there was only one case where a candidate managed to pass all stages but did not get the position.

that she has obtained in the two oral stages. Once final grades are calculated, these grades are used to rank each successful candidate among all candidates evaluated by the same committee. This ranking decides the order that successful candidates follow in order to choose which specific position they take.

In the multiple choice test questions are mainly drawn from the first of the two blocks of topics that are required in the public examination.[4] This first block includes 190 topics in the fields of Civil Law, Criminal Law, General Theory of Law and Constitutional Law. The second block includes 171 topics from Procedural Law, Administrative Law, Commercial Law and Social Law. In the first oral test questions are randomly drawn from the pool of topics included in the first block. In the second oral test the five questions are drawn from the second block of topics.

Each committee is composed of nine members. The president is a magistrate of a Superior Court of Justice or a prosecutor of similar rank. The other eight members are two magistrates, two prosecutors, one state lawyer, one court secretary, one private lawyer and a university professor.[5] Committees may differ in terms of the field of specialization of their members.

3   Data

We have collected data from public exams to judge and prosecutor positions held in Spain in years 2003 through 2007. During that period 27 evaluation committees were formed, involving 111 unique evaluators and 21,405 candidates. Below we describe the characteristics of candidates and committees.

---

[4]In 2003 and 2004, all questions in the multiple choice test came from the first block of topics. In 2005, 2006 and 2007, 80% of the multiple choice questions were drawn from the first block and 20% from the second.

[5]Members of the committee are appointed by the Selection Committee as follows: the president, by joint proposal of the president of the General Council of the Spanish Judiciary and the State Prosecutor; the two magistrates, as proposed by the General Council of the Spanish Judiciary; the two prosecutors, as proposed by the State Prosecutor; the full professor, as proposed by the University Coordination Council; the state lawyer and the court secretary, as proposed by the Ministry of Justice; and the private lawyer, as proposed by the Bar General Council.

3.1 Candidates

Slightly less than half of the candidates who had initially registered for the exam passed the multiple choice test and, therefore, qualified for the first oral stage and were evaluated by committees. In Table 2 we show descriptive statistics on these candidates.[6] Around one third of candidates were male and a significant proportion, 19%, lived in Madrid at the time of the exam. Success rates are low: only 8% of candidates passing the multiple choice test managed to pass both oral tests and obtained the position (less than 4% of candidates who had originally registered). Unsuccessful candidates tend to retake the exam. In our sample nearly 60% of candidates had already taken the exam three or more times.

We proxy for *candidates' knowledge of the first block* (mainly Civil and Criminal Law) using the grade obtained in the preliminary qualifying multiple choice test where, as explained above, at least 80% of the questions belong to the first block. Since the same multiple choice test was taken by all candidates each year, this allows us to have a homogeneous measure of quality which is independent of evaluators' actions. Unfortunately, a similar measure was not available for the second block of questions.

3.2 Committees

Table 3 displays descriptive statistics on evaluators' characteristics. There are 111 unique evaluators and the average evaluator had served 2.4 times in a similar committee over the previous five years. On average, 70% of evaluators were male. Their mean age was 52.

The professors we observe in committees are expert either in Procedural Law, Criminal Law or Civil Law. In the case of court secretaries, nearly half of them are experts in Criminal Law, but some of them specialize in Civil, Administrative or Social Law. Most judges specialize in Civil and/or Criminal Law; around one fifth of them are ex-

---

[6]Candidates that were exempt from taking the multiple choice test are not included in the sample (since 2006 those candidates that have passed the first oral stage in the previous two years are exempt from taking the multiple choice test). This applies to 421 candidates in 2006 and 559 candidates in 2007.

perts in Administrative Law. We do not observe variation in the field of specialization of prosecutors, who are generally assigned to a criminal court, or state lawyers, who tend to have a general knowledge of all fields. Information on the field of specialization of private lawyers was not available on a systematic basis.

To provide an aggregate measure of the knowledge profile of evaluators, for each committee we have computed the proportion of professors, court secretaries, and judges who specialize in the main two subjects which are included in the first block (and therefore are evaluated in the first oral test, Civil and Criminal Law).[7] We call this variable *evaluators' knowledge of first block.* Approximately 64% of committee members that are professors, judges or court secretaries specialize in one of these topics (see Table 3, last row). The remaining 36% are experts in subjects that belong to the second block of topics (and, therefore, are evaluated in the second oral test, Procedural, Administrative and Social Law).

4   Empirical analysis

Following the multiple choice test candidates are allocated to committees in a random way. Therefore, by construction, the characteristics of candidates who have been assigned to each committee are not related to evaluators' characteristics, including our variable of interest: *evaluators' knowledge of first block.* Our identification strategy exploits this fact. Any systematic variation across committees in the characteristics of candidates who are selected can be attributed to the effect of committee composition.

Our empirical analysis is structured as follows. First, we show that the allocation of candidates to committees was such that there is no significant correlation between candidates' and evaluators' characteristics. Then we explore the determinants of success in these public exams. We report results for the first and second oral stages and for the process as a whole. Finally, we analyze whether there is any relationship between the knowledge profile of evaluators and candidates' chances of success. In particular,

---

[7]Court secretaries and judges that were assigned to a military court were not included in this measure, as Military Law is not evaluated in any of the oral exams.

we examine whether candidates that are knowledgeable in the first block of topics are more likely to be hired if they are evaluated by a committee which excels in those dimensions.

4.1 Is the assignment of candidates to committees truly random?

Candidates to judge and prosecutor positions are assigned to committees through a lottery which determines the surname initial to start the matching. Given the way the assignment is structured, there should not be any significant correlation between the characteristics of candidates and the composition of evaluation committees. We examine this issue in Table 4. We cluster standard errors at the committee level to account for the fact that observations within a committee are not independent.[8] None of the observable committee characteristics -evaluators' knowledge of first block, age, gender composition or experience- is correlated with any of the observable characteristics of candidates -the grade obtained by candidates in the multiple choice test, their gender, their experience or their geographical origin-.

Table 5 provides information about the grade obtained in the multiple choice test by candidates that were assigned to a committee whose knowledge of the first block of topics was above or equal to the median and for candidates who were assigned to a committee whose knowledge of the first block was below the median. As expected, the multiple choice grade is statistically similar across different committees (columns (1)-(3)).

4.2 The determinants of candidates' success

The following equation models the determinants of candidates' performance:

---

[8]When the number of clusters is relatively small the consistency of the usual cluster-robust standard error estimator may not hold (Wooldridge 2003). Given that we have only 27 committees, we bootstrap robust standard errors cluster at the committee level. These bootstrapped clustered standard errors are slightly larger than cluster-robust standard errors.

$$y_{ijt} = \alpha + \beta x_{it} + \theta d_t + \varepsilon_{ijt} \tag{1}$$

where subindex i stands for individual, j for stage, and t indicates the year when the exam was held. The dependent variable y is a dummy that indicates whether the candidate qualified, x is a vector of individual characteristics which includes the grade obtained by the candidate in the preliminary multiple choice test, gender, geographic origin, and how many times she has taken the exam before. For the sake of simplicity, the multiple choice test information has been normalized to have mean equal to zero and standard deviation equal to one. The inclusion of a vector of year dummies d captures any systematic variation over years in the quantity and quality of candidates or in the number of available positions. Finally, ε reflects any remaining unobservable factors that could affect candidates' success. Table 6 displays the results of estimating model (1). In column (1) we report the results for the first oral stage. As expected, candidates who had obtained a higher grade in the preliminary multiple choice test have higher chances of passing the first oral test (column (1)). Interestingly, candidates who, conditional on the grade obtained in this multiple choice test, left more questions unanswered –or equivalently, made fewer mistakes–, also tend to perform better in the first oral test. Conditional on their performance in the multiple choice test, females tend to perform relatively better. This result may potentially reflect a relative advantage of male candidates on multiple choice tests relative to oral tests.[9]

Results for the second oral stage are reported in column (3). Among candidates that make it to the second oral stage, those who had obtained a higher grade in the preliminary multiple choice test tend to be more successful, but the relationship is weaker than in the first oral stage. This is consistent with the fact that the multiple choice test covers mainly contents from the first block of topics. Additionally, more experienced candidates tend to perform relatively worse at this stage. Results for the

---

[9]There is no clear consensus in the literature relative to the potential existence of a male advantage in the performance of fixed response tests vs constructed response tests. While several studies have found no significant gender differences (Walstad and Becker 1994, Greene 1997), other authors claim that males may have a relative advantage on multiple choice tests (Bell and Hay 1987, Bolger and Kellaghan 1990).

overall exam are displayed in column (5). Again, we observe that obtaining a higher grade in the multiple choice test is significantly associated with success. Candidates from Madrid have higher chances of success relative to their performance in the multiple choice test.[10]

4.3 Similar-to-me in skills effect

We compare the knowledge profile of successful candidates across different types of committees. The evidence is consistent with the existence of a similar-to-me effect. We observe that evaluators that are more knowledgeable in the first block of questions tend to hire candidates that also excel in this first block of topics (Table 5, columns (4)-(6)). The magnitude of this difference is approximately equal to one tenth of a standard deviation and is statistically significant at the 5% level. In order to understand better the origin of these differences we examine how evaluators' knowledge profile affects hiring decisions using the following equation:

$$y_{ijct} = \alpha + \beta x_{it} + \gamma z_c + \lambda(k_{it} * k_{ct}) + \theta d_t + \varepsilon_{ijct} \tag{2}$$

where c denotes the evaluation committee, variable k_{it} represents *candidates' knowledge of the first block*, as proxied by the grade obtained in the preliminary multiple choice test, and k_{ct} represents *evaluators' knowledge of the first block*. Note that, by construction, *evaluators' knowledge of the first block* is inversely related to evaluators' knowledge of the second block.

We observe that candidates that excel in the first block are more likely to pass the first stage if they are assigned to a committee that is relatively more knowledgeable in the first block, but the effect is small and it is not significantly different from zero (Table 6, column (2)). In the second stage we also find that candidates that are knowledgeable in the first block have relatively higher chances of success if they are evaluated by a committee which is expert in the first block (or, equivalently, which is

---

[10]This might potentially reflect the fact that candidates who live in Madrid may have better access to highly ranked members of the Judiciary.

not knowledgeable in the second block). The effect is statistically significant at the 5% level. Note that, even though we cannot statistically reject that the effect is similar in the first and in the second stages, the analysis suggests that the effect mainly arises in the evaluation of the second block of topics. Candidates that excel in the first block of topics benefit from having a committee which is relatively knowledgeable in this block mostly because of their leniency in the second oral test, and not because they are favored in the first oral test. Finally, in column (6) we consider whether the candidate obtains the position. We find that candidates that are knowledgeable in the first block are significantly more likely to be hired if the committee is expert in the first block.

4.4 Robustness checks

We examine the robustness of our results in several ways. First, we explore an alternative specification. Using an interaction term involves making certain assumptions about the metrics of the variables and the functional form of the effect. We estimate equation (2) interacting the variable "multiple choice grade" with a dummy variable that takes value one if "evaluators's knowledge of the first block" is above or equal to the median. As shown in Table 7, column (2), results are qualitatively very similar. Second, we explore the possibility that results are driven by some omitted committee characteristic, which might be correlated with knowledge profile. We include in our regression committees' experience, age, and gender. These characteristics do not have any significant effect on candidates' success and their inclusion does not affect results (column (3)). Third, we consider an alternative dependent variable. Instead of success, we use candidates' final grade. Since we only observe this measure for those candidates that passed the public exam, we use a tobit estimation. Again, results are qualitatively very similar (column (4)). Fourth, we extend the sample using the group candidates that was exempt from taking the multiple choice test. We proxy for their knowledge of the first block of topics using the grade obtained in the multiple choice test the previous year. The inclusion of this group of approximately 1,000 additional candidates does not affect the results (column (5)). Fifth, we explore a different identification

strategy. Most candidates retake the exam multiple times. We exploit the variation in the committee composition that an individual experiences when she is evaluated multiple times, by including an individual fixed-effect. A drawback of this strategy is that it exploits only the information provided by candidates that appear repeatedly in the database and who exhibit some variation in their outcome. This reduces the sample substantially: we have only 626 candidates (corresponding to 1547 observations) who are hired at some point of time and appear several times in the sample. The point estimates are very similar to the results obtained in our main specification but standard errors are substantially larger (column (6)). Finally, we explore how the effect has evolved over time. There was a formal change in the procedure rule such that, starting in 2005, 20% of the multiple choice test questions were taken from the second block of topics. We compare results before and after this change. We do not find any significant variation over time. If anything, the point estimate is larger in the second period (columns (7) and (8)).

5   Conclusion

In this paper we explore whether recruiters prefer applicants who are relatively strong in the skills in which the recruiters themselves excel. We test this hypothesis using data from entry exams to judge and prosecutor positions in Spain. This selection process offers several convenient features. Candidates are randomly allocated to evaluation committees, which vary in terms of the field of specialization of their members. Thus, we are able to exploit the randomness of candidates' assignment to committees in order to estimate consistently the effect of committees' knowledge profile. The subjects and the evaluations are actually taken from real life, with successful candidates receiving very substantial payoffs. Hence, while we have a setting where candidates are randomly assigned to different treatments, the setup allows to avoid the usual problems associated with artificial settings.

The empirical evidence is consistent with the existence of a *similar-to-me in skills* effect. While the groups of candidates that were initially (randomly) allocated across

committees were similar, we observe significant differences in the type of candidates that each committee hires. In particular, committees that are more knowledgeable in the topics that are evaluated in the first oral stage are more likely to hire candidates who excel in these topics.

A potential threat to the consistency of our results may exist if, following their assignment to a committee, candidates devote relatively more time to study those topics where they expect the committee to be more demanding. In this respect, our results have to be interpreted as a lower bound of the true level.[11] The lack of information about the knowledge profile of the private lawyer that sits in the committee can also affect the interpretation of our results. If private lawyers tend to be assigned to committees with a similar knowledge profile, our results might exhibit an upwards bias. If, on the contrary, private lawyers tend to be assigned to committees where their profile is relative scarce, our estimate would be a lower bound of the true effect.

Our results could be explained by the existence of differences in evaluators' objective functions. If some evaluators were biased in favor of certain dimensions, this might lead them to become specialized in those dimensions, and this would also be consistent with these individuals taking particularly into account those dimensions in their evaluations. The evidence could also reflect some type of statistical discrimination. In the presence of information asymmetries, evaluators may pay more attention to signals in dimensions in which their evaluation is more accurate. If evaluators' accuracy is higher when evaluating those dimensions in which their knowledge is greater,[12] candidates who excel in the same dimensions as evaluators may have better chances of being hired (Bagues and Perez-Villadoniga 2008). Unfortunately, our evidence does not allow us to test between these two potential explanations.

Our findings have important strategic implications for firms' recruitment practices. Once a firm becomes relatively stronger in a certain dimension, it might have a tendency

---

[11]Personal conversations with several candidates suggests that this actually might have happened in a number of cases. Such strategic behavior is severely limited by the brief time elapsed between the point at which candidates are allocated to committees and the time at which they are evaluated.

[12]Several studies have found that in a number of fields poor performers are significantly worse at evaluating quality. The lower accuracy as evaluators of bad performers has been documented in fields such as chess (Chi 1978), physics (Chi et al. 1982) and grammar (Kruger and Dunning 1999).

to recruit candidates who excel in that dimension. As a result, firms' hiring decisions may have substantial long term effects. The potential dynamic consequences of a similar-to-me effect in hiring were also noted by Schneider (1987), who argued that similarity "limits the actions of the organization owing to the fact that it occupies a constrained niche of like-minded employees sharing similar values, personalities and attitudes". Schneider calls this effect the "homogeneity hypothesis" and predicts that it may cause organizational dysfunctionality, as firms become increasingly ingrown and resistant to change.

References

[1] Antonovics, Kate, Peter Arcidiacono, and Randall Walsh (2005), "Games and Discrimination: Lessons From The Weakest Link", *Journal of Human Resources*, Vol. 40(4), pp. 918-47.

[2] Bagues, Manuel and Berta Esteve-Volart (2010), "Will gender parity break the glass ceiling? Evidence from a repeated randomized experiment", *Review of Economic Studies*, Vol. 77(4), pp. 1301-1328.

[3] Bagues, Manuel and Maria J. Perez-Villadoniga (2008), "Why do I like people like me?", *Universidad Carlos III Working Paper* 08-06.

[4] Bell, R. C. and J. A. Hay (1987), "Differences and Biases in English Language Examination Formats", *British Journal of Educational Psychology*, Vol 57, pp. 212-20.

[5] Bolger, Niall and Thomas Kellaghan (1990), "Method of Measurement and Gender Differences in Scholastic Achievement", *Journal of Educational Measurement*, Vol. 27, pp. 165-74.

[6] Broder, Ivy E. (1993), "Review of NSF Economics Proposals: Gender and Institutional Patterns", *American Economic Review*, Vol. 83(4), pp. 964-70.

[7] Byrne, Donn (1971), The Attraction Paradigm (New York: Academic Press)

[8] Chi, Michelene T. H. (1978), "Knowledge structures and memory development", in R. Siegler ed., Children's thinking: What develops?, pp. 7396, Hillsdale, NJ: Erlbaum.

[9] Chi, Michelene T. H., Glaser, Robert and Ernest Rees (1982), "Expertise in problem solving", in R. Sternberg ed., Advances in the psychology of human intelligence, Vol. 1, pp. 1776. Hillsdale, NJ: Erlbaum.

[10] Cornell, Bradford and Ivo Welch (1996), "Culture, Information, and Screening Discrimination", *Journal of Political Economy*, Vol. 104(3), pp. 542-71.

[11] Dillingham, Alan E., Marianne E. Ferber and Daniel S. Hamermesh (1994), "Gender Discrimination by Gender: Voting in a Professional Society", *Industrial and Labor Relations Review*, Vol. 47(4), pp. 622-33.

[12] Graves, Laura M. and Gary N. Powell (1995), "The Effect of Sex Similarity on Recruiters' Evaluations of Actual Applicants: A Test of the Similarity-Attraction Paradigm", *Personnel Psychology*, Vol. 48(1), pp. 85-98.

[13] Greene, Benjamin (1997), "Verbal Abilities, Gender and the Introductory Economics Course: A New Look at Old Assumptions", *Journal of Economic Education*, Vol. 28(Winter), pp. 13-30.

[14] Hamermesh, Daniel S. and Peter Schmidt (2003), "The Determinants of Econometric Society Fellows Election", *Econometrica*, Vol. 71(1), pp. 399-407.

[15] Kruger, Justin and David Dunning (1999), "Unskilled and Unaware if It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments", *Journal of Personality and Social Psychology*, Vol. 6, pp 1121-34.

[16] Parsons, Christopher, Sulaeman, Johan, Yates, Michael and Daniel S. Hamermesh (2008), "Strike Three: Umpires' Demand for Discrimination", *IZA Discussion Paper* No. 3899.

[17] Prewett-Livingston, Amelia J., Feild, Hubert S., Veres, John G. and Philip M.

Lewis (1996), "Effects of race on interview ratings in a situational panel inter-view", *Journal of Applied Psychology*, Vol. 81(2), pp. 178-86.

[18] Price, Joseph and Justin Wolfers (2010), "Racial Discrimination among NBA Referees", *Quarterly Journal of Economics*, Vol. 125(4), pp. 185987.

[19] Schneider, Benjamin (1987), "The people make the place", *Personnel Psychology*, Vol. 40, pp. 437-53.

[20] Stoll, Michael A.; Raphael, Steven and Harry J. Holzer (2004), "Black job applicants and the hiring officer's race", *Industrial and Labor Relations Review*, Vol. 57(2), pp. 267-287.

[21] Walstad, William B. and William H. Becker (1994), "Achievement Differences on Multiple-Choice and Essay Tests in Economics", *The American Economic Review Papers and Proceedings*, Vol. 84(2), pp. 193-96.

[22] Wooldridge, Jeffrey (2003), "Cluster-Sample Methods in Applied Econometrics", *The American Economic Review*, Vol. 93(2), pp. 133-138.

Table 1: Structure of the exam

	Stage 1	Stage 2	Stage 3
Format	Multiple choice	Oral	Oral
Duration	150 minutes	75 minutes	75 minutes
No. of questions	100 questions	5 questions	5 questions
Subjects evaluated	mostly first block	first block	2nd block
Grade range	0-100	0-25	0-25
Passing grade	varies	12.5	12.5

Notes: The set of all possible topics is divided into two blocks. The first block includes Civil Law (40%), Criminal Law (40%) and General Theory and Constitutional Law (20%). The second block includes Civil Procedural and Criminal Procedural Law (60%), Commercial Law (20%) and Administrative and Social Law (20%).

Table 2: Candidates' characteristics

	Mean	Std. dev.	Min.	Max.
Male	0.325	0.469	0	1
Lives in Madrid	0.194	0.396	0	1
# of times the candidate has taken the exam:				
None	0.071	0.257	0	1
Once	0.155	0.362	0	1
Twice	0.175	0.380	0	1
Three or more times	0.599	0.490	0	1
Multiple choice test grade	71.3	11.8	40.4	99
Multiple choice test unanswered questions	7.89	7.14	0	47
Passed the first oral stage	0.206	0.405	0	1
Passed the second oral stage	0.387	0.487	0	1
Obtained the position	0.080	0.271	0	1
Final grade	34.2	5.47	25.02	50

Notes: Information provided in the table refers to the 9,299 candidates who passed the multiple choice test and qualified for the first oral stage, except for variables "Passed the second oral stage" and "Final grade", which are only available for those candidates that passed, respectively, the first and the second oral stage.

Table 3: Evaluators' characteristics

	N	Mean	Std. Dev.	Min.	Max.
Male	243	0.70	0.46	0	1
Experience	243	2.45	1.90	0	5
Age	189	51.9	11.25	27	70
Knowledge of first block	115	0.64	0.48	0	1

Notes: This table reports information for 111 unique evaluators that participated in evaluation committees in public exams for Judge and Prosecutor in the period 2003-2007. Information on age is not available for lawyers and university professors. The variable *knowledge of first block* is not defined for private lawyers, state lawyers and prosecutors.

Table 4: Assignment of candidates to committees

Evaluators:	Multiple choice grade (1)	Candidates' experience (2)	Male candidate (3)	Candidate lives in Madrid (4)
Knowledge of first block	-0.014	0.003	0.002	0.003
	(0.083)	(0.007)	(0.006)	(0.007)
Experience in committees	0.086	-0.002	-0.010	-0.014
	(0.203)	(0.011)	(0.013)	(0.013)
Age	0.007	-0.005	0.010	0.014
	(0.212)	(0.012)	(0.014)	(0.013)
Male share in committee	0.047	-0.001	-0.003	-0.004
	(0.133)	(0.009)	(0.009)	(0.008)
Adjusted/Pseudo-R2	0.579	0.003	0.002	0.001
N	9299	9299	9299	9299

Notes: The dependent variable in column (1) is the grade obtained by the candidate in the preliminary multiple choice test, in column (2) the dependent variable is a dummy that indicates whether the candidate took the exam three or more times, in column (3) whether the candidate was male and (4) whether the candidate lives in Madrid. Column (1) shows the results of an OLS estimation. Columns (2), (3) and (4) report marginal effects from a probit estimation. All regressions include year fixed effects (not reported). Bootstrapped standard errors, clustered by committee, are in parentheses.

Table 5: Candidates' multiple choice grade, by year and by type of committee

Committee knowledgeable in:	All candidates			Successful candidates		
	1st block (1)	2nd block (2)	p-value (3)	First block (4)	Second block (5)	p-value (6)
2003	0.00	-0.01	0.82	0.34	0.49	0.56
	[1052]	[375]		[65]	[22]	
2004	0.00	-0.00	0.97	0.85	0.77	0.52
	[1408]	[682]		[161]	[79]	
2005	-0.00	0.01	0.84	0.91	0.70	0.08
	[1105]	[731]		[111]	[77]	
2006	0.03	-0.02	0.34	0.93	0.84	0.63
	[628]	[1292]		[39]	[76]	
2007	-0.02	0.02	0.40	1.02	0.59	0.02
	[1004]	[1022]		[52]	[61]	
All years	0.00	-0.00	0.98	0.82	0.71	0.03
	[5197]	[4102]		[428]	[315]	

Note: This table includes information on the grade that candidates obtained in the multiple choice test. This grade is normalized with mean equal to zero and standard deviation equal to one for all candidates that passed the multiple choice test a given year. The number of observations is in brackets. Column (1) provides information for candidates that were assigned to a committee whose knowledge of the first block was above or equal to the median (0.60). Column (2) provides information for candidates that were assigned to a committee whose knowledge of the first block was below the median. Column (3) reports the p-value of a test on the equality of means, conditional on year. Columns (4), (5) and (6) report similar information for the sample of candidates that was hired.

Table 6: Determinants of candidates' success

	First stage		Second stage		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)
Multiple choice grade	0.287***	0.276***	0.140***	0.003	0.115***	0.086***
	(0.014)	(0.030)	(0.045)	(0.087)	(0.005)	(0.015)
Multiple choice blanks	0.041***	0.041***	-0.001	-0.003	0.014***	0.014***
	(0.006)	(0.006)	(0.026)	(0.026)	(0.004)	(0.004)
Male	-0.021**	-0.021**	0.020	0.018	-0.002	-0.002
	(0.009)	(0.009)	(0.026)	(0.026)	(0.005)	(0.005)
Experience=1	0.024	0.023	-0.018	-0.021	0.009	0.009
	(0.025)	(0.025)	(0.052)	(0.052)	(0.014)	(0.014)
Experience=2	0.049**	0.049**	-0.060	-0.058	0.009	0.009
	(0.026)	(0.026)	(0.054)	(0.054)	(0.012)	(0.012)
Experience≥3	0.034*	0.034*	-0.150***	-0.148***	-0.010	-0.010
	(0.020)	(0.020)	(0.054)	(0.054)	(0.010)	(0.010)
Lives in Madrid	0.011	0.011	0.038	0.039	0.011*	0.011*
	(0.011)	(0.011)	(0.029)	(0.029)	(0.007)	(0.007)
Evaluators' knowledge of 1st block		-0.027		-0.235**		-0.055**
		(0.052)		(0.110)		(0.024)
Evaluators' knowledge of 1st block*Multiple choice grade		0.001		0.013**		0.003**
		(0.003)		(0.006)		(0.001)
Pseudo R-2	0.129	0.130	0.025	0.028	0.112	0.114
Pred. P	0.170	0.170	0.384	0.384	0.058	0.057
N	9299	9299	1919	1919	9299	9299

Notes: The dependent variable is a dummy variable that indicates whether the candidate passed the stage. The coefficients reported correspond to the marginal effects of a probit regression. Bootstrapped standard errors, clustered by committee, are in parenthesis. *significant at 10%; **significant at 5%; ***significant at 1%.

Table 7: Robustness checks

	Main result	Alternative interaction	Committee characteristics	Final grade	Extended sample	Fixed effects	Years 2003-2004	Years 2005-2007
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Evaluators' knowledge of 1st block	-0.055**	-0.035**	-0.054**	-8.059**	-0.052*	-0.040	-0.014	-0.048*
	(0.024)	(0.017)	(0.026)	(3.170)	(0.028)	(0.169)	(0.098)	(0.025)
Evaluators' knowledge of 1st block*Multiple choice grade	0.003**	0.032**	0.003**	0.422**	0.003*	0.002	0.001	0.003*
	(0.001)	(0.016)	(0.001)	(0.177)	(0.002)	(0.009)	(0.005)	(0.001)
N	9299	9299	9299	9299	10279	1547	3517	6762

Notes: Column (1) reports the marginal coefficients from a probit estimation of equation (1). Column (2) reports results from an alternative specification where the variable "evaluators's knowledge of the first block" is a dummy variable that takes value one if committee's knowledge of the first block is above or equal to the median. Column (3) includes among regressors (not reported) committees' experience, age, and gender. Column (4) reports the results from a tobit regression where the left hand side variable is candidates' final grade. In column (5) the sample includes candidates that were exempt from taking the multiple choice test. Column (6) includes individual fixed effects. Column (7) and column (8) report results from the main specification for two different time periods. Bootstrapped standard errors, clustered by committee, are in parenthesis. *significant at 10%; **significant at 5%; ***significant at 1%.