Do recruiters prefer applicants with similar skills? Evidence from a Randomized Natural Experiment[*]

Manuel Bagues [1] and Maria J. Perez-Villadoniga [2]

Abstract

In this paper we examine the potential existence of a similar-to-me effect in terms of skills between recruiters and applicants. Using evidence from entry exams to the Spanish Judiciary, where applicants are randomly assigned across evaluation committees, we find that committee members tend to be more demanding at those stages where they are more knowledgeable. As a result, applicants who excel in the same dimensions as recruiters are more likely to be hired.

Keywords: hiring, randomized experiment, similar-to-me effect.
JEL Classification: M51, J45

[1] mfbagues@emp.uc3m.es. Universidad Carlos III, Departamento de Economía de la Empresa, Calle Madrid 126, 28903 Getafe (Madrid), Spain.
[2] mjpvilla@uniovi.es. Universidad de Oviedo, Departamento de Economía, Avenida del Cristo s/n, 33006 Oviedo (Asturias), Spain.

1 Introduction

The question of what influences employers' hiring decisions has been discussed in the Economics, Management and Psychology literature 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 may exhibit a tendency to select candidates' that are similar to themselves in some respect, such as values, background or some demographic characteristic.[1]

Two types of explanations have been proposed for the potential existence of a similar-to-me effect. On the one hand, some authors have offered a preference-based explanation. The similarity-attraction paradigm argues that there is a strong association between similarity to oneself and interpersonal attraction, which may 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. Interpersonal attraction then leads to positive bias in the recruiter's interview conduct."[2] On the other hand, other authors have pointed out that a similar-to-me effect might arise if there exist information asymmetries about the real productivity of workers and evaluators can observe more reliable signals of quality when they evaluate candidates that share the same background as themselves (Cornell and Welch 1996). According to these authors, if it "is easier for individuals to screen people of similar background ... people

---

[1]Most authors have focused on similarity in terms of demographic characteristics. With respect to race, Stoll et al. (2004) find that black agents hire a relative greater proportion of black applicants than white agents. Another example is the study conducted by Prewett-Livingston et al. (1996), who find a positive effect of same-race on promotion. In sports, Parsons et al. (2008) and Price and Wolfers (2007) find a same-race bias by referees in baseball and basketball respectively. The match between the gender of evaluators and candidates has also attracted attention from economists, although the empirical evidence on this respect is mixed. Antonovics et al. (2005) find that in the TV show "The Weakest Link" female contestants discriminate against male participants. Dillingham et al. (1994) find that, within an association of economists, female voters are more likely to vote for female candidates relatively to male voters. On the contrary, Broder (1993) finds that female applicants for grants of the National Science Foundation receive relatively higher evaluations from male evaluators than from female ones. Bagues and Esteve-Volart (forthcoming) show that candidates to positions in the Spanish Judiciary are less likely to be hired when they are randomly assigned to an evaluation committee where the majority of members belong to their own gender. A particular case where the field of specialization of candidates is considered is Hamermesh and Schmidt (2001), 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.

[2]On a related ground, the attraction-selection-attrition (ASA) theory (Schneider 1987), which is one the main theoretical foundations of the person-organization fit literature, posits that organizations attract, select, and retain similar types of people.

(will) tend to hire others of their own type even when they have no innate preference for similar people" (p. 544).

In a framework of asymmetric information, in this paper we propose a simple model that applies to a multistage selection process where at each stage candidates are evaluated on their ability at a different dimension. This type of selection process is very usual in labor markets, probably because it allows to limit the number of times that bad candidates are evaluated (U.S. Department of Labor 1999). The model predicts that evaluators will be more demanding at those dimensions where they are more accurate. As a result, if evaluators are more accurate in those dimensions where they are more skilled, candidates that excel in the same dimensions as evaluators have more chances of being hired.

We test this hypothesis using evidence from a real labor market selection process: the entry exams to judge and prosecutor that were held in Spain in the period 2003-2007. This is a selection process 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 facilitates the empirical analysis. Several evaluation committees are formed each year and candidates are allocated across evaluation committees following a random procedure. The composition of committees varies in terms of the field of specialization of their members, allowing to study how the expertise of committee members affects selection decisions.

We find that committees are significantly more demanding at those stages at which they are more skilled. An increase in one standard deviation in evaluators' relative knowledge of the field wbich is being evaluated at a given stage is associated with grades being 0.25 marks lower (out of a maximum of 25 marks). This favors candidates that have been assigned to evaluation committees with similar skills as themselves, as they will have higher chances of qualifying. As a result, while the groups of candidates that were initially assigned to each committees were similar in all respects, successful candidates tend to be similar to their recruiters in terms of their skills. In quantitative terms, an increase in one standard deviation in evaluators' relative knowledge of a field is associated to successful candidates being on average 5% of a standard deviation more knowledgeable in that

3

field, as measured by their performance in a preliminary multiple choice test.

The structure of the paper is as follows. In section 2 we present a theoretical model that illustrates the mechanisms through which, in a multistage selection process, a similar-to-me effect might arise. In section 3 we provide background information about hiring procedures in the Spanish Judiciary and we present the empirical analysis. Finally, section 4 concludes and discusses potential implications of the results for firms' recruiting practices.

2   The Model

Let us consider the case of an recruiter who must evaluate a stream of candidates. There are two types s of candidates, high productivity (H) and low productivity (L) ones. The evaluator has to make a decision y with respect to each candidate: either to *select* him or to *discard* him. The best possible action that the evaluator can take is to select highly productive candidates and to reject those of low productivity. Formally, for every candidate the evaluator's payoff function can be represented as:

$$\Pi\left(y,s\right) = \begin{cases} 1 & if\ (y,s) = (select, H)\ or\ (y,s) = (discard, L) \\ 0 & otherwise \end{cases} \tag{1}$$

The problem of the evaluator is that there exist informational asymmetries that prevent her from observing with perfect accuracy whether a given candidate belongs to the high or to the low productivity group. The evaluator knows the proportion of type H individuals (p) and type L individuals ($1-p$) in the population and, at each stage of the process, she can also obtain some imperfect signal of candidates' quality. In particular, we consider a multiple hurdle selection process where, for simplicity, the number of stages is equal to two. In order to be selected the applicant must pass both stages, failure at any step of the process disqualifies him from further consideration. The evaluator needs to set at each stage a cutoff level that maximizes the probability of selecting high productivity candidates and minimizes the probability of choosing those of low productivity.

More precisely, the structure of the process is the following. In period 0 nature picks the state of

the world, i.e., whether a candidate's productivity is high or low. This information is not observable by the evaluator. In period 1 the candidate goes through the first stage of the selection process. During the evaluation, the evaluator observes a signal x_1 of candidate's quality. If this signal surpasses a threshold level l_1 set by the evaluator, then the candidate proceeds to the second stage. In period 2, the second stage of the evaluation occurs and the evaluator sets a new threshold level l_2. If the evaluator observes a signal x_2 which is above this threshold then the candidate is selected. Figure 1 describes this sequence of events.

[Figure 1 here]

Given the evaluator's payoff function (1), an optimal selection strategy requires that the expected proportion of selected candidates is equal to the proportion of high productivity individuals in the population p and, therefore, the selection problem can be characterized as follows:

$$\max_{l_1,l_2} V = \Pr[x_1 > l_1|H]\Pr[x_2 > l_2|H] \tag{2}$$

$$s.t. \quad p = p\Pr[x_1 > l_1|H]\Pr[x_2 > l_2|H] + (1-p)\Pr[x_1 > l_1|L]\Pr[x_2 > l_2|L]$$

In order to formalize the existence of a stochastic relationship between candidates' true productivity and the signal observed by the evaluator, we characterize the evaluator's screening technology using the following linear density functions:

$$Pr(x_j = x/H) = 1 - a_j + 2a_j x \tag{3}$$

$$Pr(x_j = x/L) = 1 + a_j - 2a_j x$$

where j=1, 2; $0 \leq a_j \leq 1$ and $0 \leq x \leq 1$. Parameters a_1 and a_2 represent, respectively, evaluator's accuracy in the first and second stages. Note that these functions satisfy the monotone likelihood ratio property (MLRP). This is, a good signal is more likely be observed when the

candidate is relatively more productive. Figure 2 illustrates the case where the evaluator is relatively more accurate in the evaluation of the first stage or, equivalently, $a_1 > a_2$.

[Figure 2 here]

In this case, the probability that the evaluator observes a high signal from a high quality candidate is greater in the first stage than in the second one. Similarly, the probability of observing a low signal from a high quality candidate is lower in the first stage. The same reasoning applies for low quality candidates.

The optimal threshold level that the evaluator should set at each stage depends on the proportion of good and bad candidates in the population and on the accuracy with which the evaluator can observe candidates' quality at each stage. The first proposition deals with the evaluator's optimal strategy when her accuracy is the same at both stages.

Proposition 1 *If the evaluator is equally accurate at both stages then she should set in both stages the same threshold level.*

$$a_1 = a_2 \implies l_1^* = l_2^*$$

Proof. See Appendix ∎

However, the accuracy of the evaluator may not be necessarily the same at each stage. It may be the case that in one of the two stages the evaluator observes signals that are relatively more informative. What is the optimal strategy in this case? Proposition 2 deals with this question.

Proposition 2 *The greater the accuracy of the evaluator at a certain stage, the higher will be the threshold level required to pass that stage.*

$$\frac{dl_j^*}{da_j} > 0; \qquad where \; j\text{=1, 2 \& } i \neq j$$

Proof. See Appendix ∎

6

3   Empirical Analysis

In the theoretical model presented above it is shown that, in a multistage selection process, an evaluator who makes an optimal use of the available information will be more demanding at those stages at which she can assess quality more accurately. Recruiters' accuracy is likely to depend both on the field which is being evaluated and on evaluators' own characteristics. In this respect, a number of authors have pointed out that, in general, evaluators' accuracy tends to be higher when they evaluate dimensions at which their knowledge is greater.[3] If this is the case, according to the above model, evaluators will tend to be relatively more demanding at those dimensions where they are more skilled and a similar-to-me effect between evaluators' and successful candidates' skills will be observed.

We test this hypothesis using data from entry exams to judge and prosecutor in Spain. This selection process offers several convenient features for the analysis. First, the structure of the selection process is similar to the one described in the model: it consists of several qualifying stages where candidates are evaluated on their knowledge in different areas. Second, candidates are randomly allocated to evaluation committees, which vary in terms of the field of specialization of their members. Thus, we are able exploit the randomness of candidates' assignment to committees in order to test the effect of committees knowledge profile. Third, 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.

3.1   Background

In Spain, candidates to judge and prosecutor positions go through a public examination at the national level. This selection process is held yearly for both positions jointly. The structure of the process is described in Table 1.[4] The exam consists of three qualifying stages: a multiple choice

---

[3]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 as diverse as chess (Chi, 1978), physics (Chi et al., 1982) and grammar (Kruger and Dunning, 1999).

[4]The evaluation process is described in each call. See, for example, "Agreement of May 8 2006 of the Selection Commission", BOE nº3, May 13, 2006, p. 18510. A shorter description of the process is also available in Bagues &

7

test and two oral tests. The multiple choice test is held first. It 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. Questions are mainly drawn from the first of the two blocks of topics that are required in the public examination.[5] This first block includes 190 topics from 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. Candidates whose grade is above the minimum passing threshold qualify for the first oral test.[6]

Once it is known how many candidates have qualified evaluation committees are formed. 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.[7] The allocation of candidates to committees is done according to the following process: candidates are ranked in alphabetical order and then, beginning with the surname initial drawn from a lottery, they are distributed to committees in such a way that every committee is assigned the same number of candidates. Then the first oral test starts. In this test each candidate must answer orally, within 75 minutes, five questions which are drawn randomly from the pool of topics included in the first block. After the candidate has answered, the committee decides, on a majority basis, whether the candidate qualifies for the second oral test, the president deciding in case of a tie. Every successful candidate is given a numerical grade, which is computed summing up the grades proposed by each member of the committee, excluding the minimum and the maximum.

The structure of the second oral test is similar, although in this case the five questions are drawn from the second block of topics. Those candidates that manage to pass this test can then choose

Esteve-Volart (forthcoming)

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

[6]The minimum passing grade is decided each year by the Selection Commission based on the number of evaluation committees that are available.

[7]Members of the committee are appointed by the selection commission 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, proposed by the General Council of the Spanish Judiciary; the two prosecutors, proposed by the State Prosecutor; the full professor, proposed by the University Coordination Council; the state lawyer and the court secretary, proposed by the Ministry of Justice; and the private lawyer, proposed by the Bar General Council.

between becoming judge or prosecutor according to their final ranking.[8] This ranking is made in the following way: first, a final grade is computed as the sum of the grades received in both oral tests; then candidates who got the first place in each committee are ranked according to their final grade; candidates ranked in the second place in each committee are ranked, and so on until all candidates have been ranked.

3.2 Data

We have collected data from public exams to judge and prosecutor that were held in Spain in years 2003 to 2007. In the five years considered, 27 evaluation committees were formed, involving 243 evaluators and 21,405 candidates. Below we offer information on the characteristics of candidates and committees.

3.2.1 Candidates

Slightly less than half of those candidates that 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.[9] Around one third of candidates were male and a significant proportion, 19%, lived in Madrid at the time of the exam. Unsuccessful candidates tend to retake the exam. In our sample nearly 60% of candidates had already taken the exam three or more times. Success rates are rather low: only 8% of candidates passing the multiple choice test managed to pass both oral tests and obtained the position (less than 4% of those who had originally registered).

3.2.2 Committees

Table 3 displays descriptive statistics on evaluators' characteristics. On average, 70% of evaluators were male. Their mean age was 52. The average evaluator had participated 2.4 times in a similar

---

[8]Formally, passing all stages does not guarantee obtaining a position as committees cannot deliver more positions than those 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 still did not get the position.

[9]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.

committee over the previous five years.[10]

Evaluators may differ in terms of their field of specialization. In Table 4 we report the field of specialization of professors, court secretaries and judges. 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 are specialized in Civil, Administrative or Social Law. Most judges are specialized in Civil and/or Criminal Law; around one fifth of them are experts 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 committees' knowledge profile, for each committee we have computed the proportion of professors, court secretaries, and judges who are knowledgeable 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-.[11] We denominate this variable *knowledge of first block.* On average, 64% of committee members that are professors, judges or court secretaries are specialized in one of these topics (see Table 3, last row). The remaining 36% is specialized in subjects evaluated in the second oral test -Procedural, Administrative and Social Law-.

3.3 Results

The following equation models the determinants of candidates' performance:

$$y_{ijct} = \alpha + \beta x_{it} + \gamma z_{ct} + \theta d_t + \varepsilon_{ijct} \tag{4}$$

where subindex i stands for individual, j for stage, c denotes the evaluation committee and t indicates the year when the exam was held. Variable y measures candidate's performance, x is a vector of individual characteristics which includes the grade obtained by the candidate in the

---

[10]Out of a total of 243 observations there are 111 unique individuals.

[11]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.

preliminary multiple choice test, gender, geographic origin, and how many times he has taken the exam before. Variable z represents committee's *knowledge of first block*. The inclusion of a vector of year dummies d captures any variation across years in the quantity and quality of candidates, in the number of available positions or in the characteristics of evaluators. Finally, ε reflects any remaining unobservable factors that could affect candidates' success.

Following the multiple choice test candidates are allocated to committees in an essentially random way. Therefore, by construction, in the first oral stage the quality of candidates that have been assigned to each committee is independent of evaluators' characteristics, including our variable of interest: how knowledgeable the committee is on the topics that are evaluated at each stage. Our identification strategy exploits this fact. Any systematic variation across committees in the characteristics of candidates that pass the first oral stage is attributable to the effect of committee composition. Note, however, that in the case of the second oral stage coefficients cannot be interpreted directly as a consistent estimation of the causal effect of committee's composition given that in this case the pool of candidates is no longer independent of the characteristics of the committee.

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 estimate equation (4) using two different measures of success: (i) the grade obtained by the applicant and (ii) a dummy variable that indicates whether the candidate passed the stage or not. We report results for the first and second oral stages and for the process as a whole. To facilitate the interpretation of our results in all regressions variables at the committee level have been normalized with mean equal to zero and standard deviation equal to one.

3.3.1 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. After the lottery, the list of candidates that have been allocated to each committee is made public. 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. Results in Table 5 suggest that this is actually the case.

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 and their geographical origin-.

3.3.2 The determinants of candidates' success

Table 6 displays the results of estimating model (4) for the first oral stage. In columns (1) & (2) we measure performance using the grade obtained by candidates. As we only observe a numerical grade for successful candidates, we estimate a tobit model with left censoring at 12.5, which is the minimum passing grade. As expected, candidates who performed better at the preliminary multiple choice test tend to obtain higher grades in 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 less mistakes-, tend also to obtain higher grades. Conditional on their performance in the multiple choice test, grades are slightly lower among males. This result may potentially reflect an advantage of male candidates on multiple choice tests relative to oral tests.[12] Committees' knowledge of the subjects evaluated in the first oral test has a significant negative effect on candidates' grades in this test. In quantitative terms, an increase in one standard deviation in committee's *knowledge of first block* decreases candidates' grades by 0.25 points. In order to check whether this effect is due to other committee characteristics that might be correlated with their knowledge profile, in column (2) we control for evaluators' age, gender and experience. None of these variables has any significant effect on candidates' grades and the estimated effect of committees' *knowledge of first block* remains unchanged.

In columns (3) & (4) we repeat the analysis using this time as dependent variable a dummy variable which indicates whether the candidate passed or not the first stage. Results are qualitatively similar to those reported in the first two columns. Those candidates who have obtained a higher grade in the preliminary multiple choice test have higher chances of passing the first oral test (column

---

[12]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).

(1)). Also, candidates that, conditional on their grade in the multiple choice test, had left more questions unanswered tend to perform better in the first oral test. Female and more experienced candidates tend to perform relatively better. Candidates' chances of passing the first oral test tend to be lower when evaluated by committees whose members are relatively more expert in the first block of topics. In quantitative terms an increase in one standard deviation in *knowledge of first block* decreases the number of candidates passing this test by about 4% (0.7 percentage points), although this effect is not statistically significant at standard levels.

Results for the second oral stage are reported in Table 7, although, in this case, given the sample selection issue, it is not possible to attribute a causal interpretation to the estimated coefficients. Among candidates that make it to the second oral stage, those ones that had obtained a higher grade in the preliminary multiple choice test tend to be more successful. As well, more experienced candidates tend to perform relatively worse in this stage. We do not find any significant correlation between committee characteristics and candidates' grades.

Results for the overall exam are displayed in table 8. Again, we observe that obtaining a higher grade in the multiple choice test is significantly associated with success. Candidates from Madrid have also relatively higher chances of success relative to their performance in the multiple choice test.[13] Success rates are similar across of committees, independently of their composition. This is, we do not observe a relationship between the number of candidates that are hired at each committee and the knowledge profile of its members.

3.3.3 Similar-to-me in skills effect

We have found that evaluators tend to be more demanding at those stages where they are more skilled. A straightforward implication of this result is that, if candidates' quality is not perfectly correlated across dimensions, candidates hired at each committee will tend to be relatively better at those dimensions where their evaluators excel.

In order to test whether such a similar-to-me in skills effect exists, an objective measure of candidates' knowledge across dimensions is needed. We proxy candidates' *knowledge of the first*

---

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

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.

The following equation describes the relationship between the knowledge profile of successful candidates and committee composition:

$$y_{it} = \alpha + \beta x_{it} + \gamma z_{ct} + \theta_t d_t + \varepsilon_{ict} \tag{5}$$

where y represents the grade obtained in the preliminary multiple choice test, x candidates' characteristics, z refers to committees' characteristics and the vector of year dummies d captures any year specific factor. In addition to predetermined individual characteristics, we also consider in our regression the ordinal ranking obtained by each (successful) candidate at the committee level at the end of the exam. Taking into account this ranking allows us to compare across committees candidates who obtained the same ranking, avoiding any potential distortion linked to differences in the number of candidates that have obtained the position at each committee.[14]

As expected, we observe a significant positive correlation between evaluators' and successful candidates' *knowledge of the first block* (see Table 9, column (1)). An increase in one standard deviation in evaluators' knowledge of the material which is being covered in the first oral test is associated to successful candidates in that committee having obtained 0.42 more points in the multiple choice test (5% of a standard deviation). In column (2) we include in our regression other committee characteristics such as experience, age, and gender. These variables have no significant effect on the type of candidates that are being selected and their inclusion does not affect the point estimate of the effect of committees' *knowledge of the first block*.[15]

---

[14]All committees within an exam have the same number of candidates and the same number of available positions but some committees may decide not to assign all positions.

[15]Including these variables in the regression is associated to an increase in the standard errors of our variable of interest such that the coefficient is not any longer significant at standard levels. Given that the inclusion of these variables decreases the adjusted R-squared of the regression and, as there is no theoretical reason to expect any effect, our preferred specification is the one depicted in column (1).

4 Conclusion

In this paper we examine the potential existence of a similar-to-me in skills effect between recruiters and applicants. In a framework of asymmetric information, our model suggests that an evaluator who makes an optimal use of the available information will be relatively more demanding at those stages where she is more knowledgeable and, therefore, more accurate. As a result, applicants that excel in the same dimensions as recruiters have relatively higher chances of being hired.

We test the predictions of the model using data from public exams to judge and prosecutor that were held in Spain between 2003 and 2007. This selection process offers very convenient features for the empirical analysis. Candidates are randomly allocated to evaluation committees and committees vary in terms of the knowledge profile of its members. This set up allows to estimate consistently the effect of committee composition on the evaluations received by candidates. Since different committees are evaluating candidates that apply for the same position, we can rule out the possibility that a correlation between the skills of evaluators and the skills of successful candidates is observed due to the self-selection of candidates or committee members.

The empirical evidence is consistent with the model's prediction. Committee members are relatively more demanding at those stages where they are more knowledgeable. 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 that 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. As much as this behavior is relevant, our results have to be interpreted as a lower bound of its true level.[16]

In addition to an asymmetric information story, our results could also be explained by the existence of differences in evaluators' objective functions. If some evaluators were biased in favor of

---

[16]Personal conversations with several candidates suggests that this actually might have happened in a number of cases. Still, room for such strategic behavior is severely limited by the shortness of the time elapsed since candidates are allocated to committees until they are evaluated.

certain dimensions, this might lead them to become specialized in those dimensions, and this would also be consistent with these individuals taking most into account these particular dimensions in their evaluations. Unfortunately, our evidence does not allow to test between the two potential explanations.

The existence of a similar-to-me in skills effect in firms' recruitment practices has several important implications. First, firms' hiring decisions may have substantial long term effects. Once a firm becomes relatively stronger in a certain dimension, it may find more convenient, from an informational point of view, to recruit candidates that excel in this dimension. [17] Second, the existence of a similar-to-me in skills effect may generate a same-group bias. If races or genders are equally productive but differ in their distribution of ability, candidates belonging to the same gender or racial group as the evaluator may have higher chances of being hired. In a companion paper we formally show the precise conditions under which this implication holds (Bagues and Perez-Villadoniga 2008).

A   Appendix

A.1   Proof of Proposition 1

Proof. The corresponding Lagrangian to the evaluator's problem (2) is given by:

$$L = \Pr\left[x_1 > l_1 | H\right] \Pr\left[x_2 > l_2 | H\right] + \lambda \left[p - p \Pr\left[x_1 > l_1 | H\right] \Pr\left[x_2 > l_2 | H\right] + (1-p) \Pr\left[x_1 > l_1 | L\right] \Pr\left[x_2 > l_2 | L\right]\right]$$

Since, by definition, signals are bounded between zero and one, any interior solution to the evaluator's problem (2), (l_1^*, l_2^*), must satisfy:

(i) $Pr(x_1 = l_1^*/H)Pr(x_2 = l_2^*/L) \Pr\left[x_2 > l_2^*|H\right] \Pr\left[x_1 > l_1^*|L\right] = Pr(x_2 = l_2^*/H)Pr(x_1 = $

---

[17]The potential dynamic consequences of a similar-to-me effect in hiring was noted also 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.

$l_1^*/L) \Pr[x_1 > l_1^*|H] \Pr[x_2 > l_2^*|L]$

(ii) $p = p \Pr[x_1 > l_1^*|H] \Pr[x_2 > l_2^*|H] + (1-p) \Pr[x_1 > l_1^*|L] \Pr[x_2 > l_2^*|L]$

If evaluator's accuracy is the same in both stages ($a = a_1 = a_2$), given the evaluator's screening technology (4), the necessary condition (i) can be written as:

$$2a(l_1^* - l_2^*)(l_1^* - 1)(l_2^* - 1)(a^2 - 2a^2 l_1^* - 2a^2 l_2^* + 2a^2 l_1^* l_2^* + 1) = 0$$

This equation yields the following two interior solutions:

(a) $l_1^* = l_2^*$

(b) $l_1^* = \frac{1 + a^2 - 2a^2 l_2^*}{2a^2(1 - l_2^*)}$.

where (b) cannot be a solution to our problem since, for any value of a and l_1^* within their domain, l_1^* takes values above one, and then condition (ii) cannot be satisfied. Therefore, as stated in (a), the optimal strategy implies fixing the same threshold in both stages. ∎

A.2 Proof of Proposition 2

Proof. Given the screening technology (4), from (i) and (ii), we have:

$$(1 - a_1 + 2a_1 l_1^*)(1 + a_2 - 2a_2 l_2^*)(1 - a_1 l_1^*)(1 + a_2 l_2^*) = (1 + a_1 - 2a_1 l_1^*)(1 - a_2 + 2a_2 l_2^*)(1 + a_1 l_1^*)(1 - a_2 l_2^*)$$

$$p = p(1 - l_1^*)(1 + a_1 l_1^*)(1 - l_2^*)(1 + a_2 l_2^*) + (1 - p)(1 - l_1^*)(1 - a_1 l_1^*)(1 - l_2^*)(1 - a_2 l_2^*)$$

The high degree of this system of equations does not allow its analytical resolution. Instead, we have performed simulations for different values of parameters a_1, a_2 and p in the range 0 and 1. Our results confirm the claim. As evaluator's accuracy in one stage increases relative to her accuracy in the other, the optimal threshold level in the former (latter) increases (decreases).

Figure 3 illustrates the optimal threshold levels in stages 1 and 2 as evaluator's accuracy in stage 1 (a_1) increases from 0 to 1, fixing the proportion of type H candidates in the population (p) at 0.25 and her accuracy in stage 2 (a_2) at 0.5. ∎

17

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 (forthcoming), "Will gender parity break the glass ceiling? Evidence from a repeated randomized experiment", *Review of Economic Studies*.

[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 (2001), "Hall of Fame Voting: The Econometric Society", *NBER Working Paper Series* No. 8435.

[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 interview", *Journal of Applied Psychology*, Vol. 81(2), pp. 178-86.

[18] Price, Joseph and Justin Wolfers (2007), "Racial Discrimination among NBA Referees", *NBER Working Paper Series* No. 13206

[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] U.S. Department of Labor (1999), "Testing and Assessment: An employer's guide to good practices", Washington DC: Author.

[22] 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.

Figure 1: **Structure of the evaluation process**
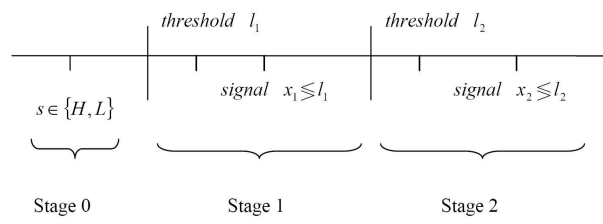
Notes: In stage 0 nature decides whether the candidate is of high (H) or low (L) productivity. Then, in each stage of the evaluation process the evaluator sets a threshold l and obtains a signal of each candidate's performance x. This signal must be above the threshold level for the candidate not to be rejected.

Figure 2: **Probability of observing signal x in each stage**
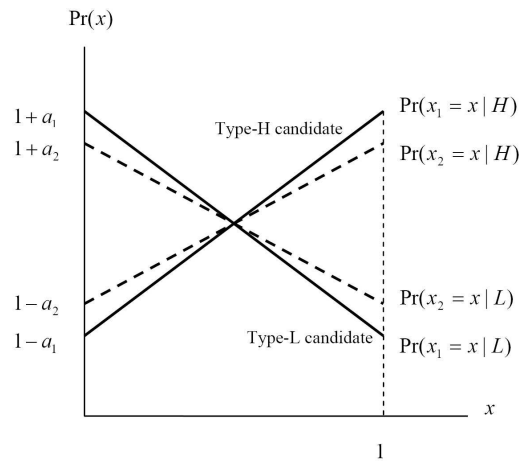
Note: The figure represents the screening technology of a recruiter that is more accurate at evaluating candidates at the first stage ($a_1 > a_2$). The smooth (dashed) lines represent the probability of observing signal x at the first (second) stage when the candidate who is being evaluated belongs to the high or to the low productivity group.

Figure 3: **Optimal threshold levels**
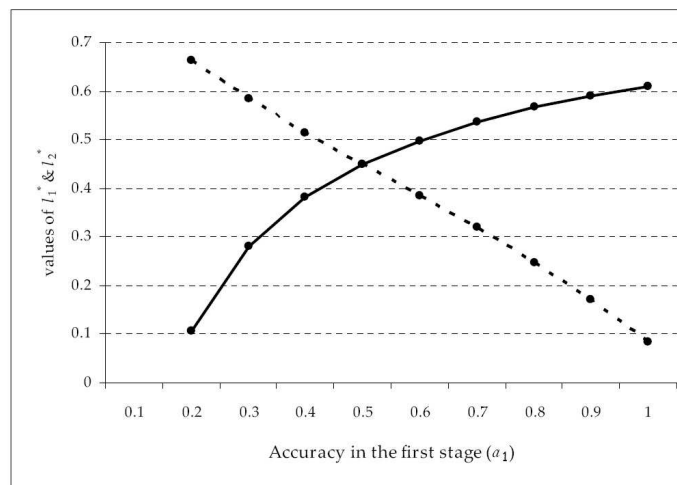
Note: The smooth (dotted) line represents the optimal threshold levels in stage 1 (stage 2) as evaluator's accuracy in stage 1 (a_1) increases from 0 to 1. The proportion of type H individuals in the population (p) is fixed at 0.25 and evaluator's accuracy in stage 2 (a_2) is fixed at 0.5.

22

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 ranking	19.5	11.5	1	46

Notes: Information provided in the table refers to the 9299 candidates who passed the multiple choice test and qualified for the first oral stage, except for variables "Passed the second oral stage" and "Final ranking", 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: 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: Evaluators' field of expertise

Committee #	2003	2004	2005	2006	2007
	University Professors				
1	Procedural	Procedural	Procedural	Procedural	Procedural
2	Procedural	Procedural	Procedural	Procedural	Procedural
3	Civil	Criminal	Procedural	Civil	Criminal
4	Criminal	Criminal	Procedural	Procedural	Procedural
5	-	Criminal	Criminal	Procedural	Civil
6	-	Procedural	-	Criminal	Procedural
	Court Secretaries				
1	Military	Social	Social	Criminal	Social
2	Criminal	Criminal	Criminal	Social	Criminal
3	Civil	Criminal	Administrative	Administrative	Criminal
4	Criminal	Criminal	Criminal	Administrative	Administrative
5	-	Civil	Criminal	Administrative	Civil
6	-	Administrative	-	Criminal	Administrative
	Judges				
1	Civ/Crim	Civ/Crim	Civ/Crim	Civ/Crim	Civ/Crim
	Civ/Crim	Civ/Crim	Crim	Crim	Crim
	Military	Military	-	Crim	-
2	Adm	Crim	Adm	Crim	Adm
	Crim	Crim	Crim	Crim	Crim
	-	-	Crim	-	Crim
3	Adm	Adm	Adm	Adm	Civil
	Crim	Crim	Civil	Adm	Crim
	Civil	Crim	-	Crim	-
4	Civil	Civil	Military	Adm	Civil
	Civil	Civil	Crim	Civil	Adm
	-	-	Crim	-	Crim
5	-	Adm	Civil	Adm	Adm
	-	Crim	Crim	Military	Civil
	-	Crim	-	Crim	-
6	-	Crim	-	Civil	Military
	-	Crim	-	Crim	Crim
	-	-	-	-	Crim

Notes: The table indicates the field of expertise of committee members. For instance, the cell situated at the upper left corner of the table indicates that the university professor that was in committee number one in the exam held in year 2003 was an specialist in Procedural Law. Information on the field of specialization of judges and court secretaries was obtained from professional ranks, published periodically in the Official Spanish Gazette. Professors' field of specialization was provided in the official call for each public exam.

Table 5: Random assignment of candidates to committees

	Multiple choice grade	Candidate's experience	Male candidate	Candidate lives in Madrid
	(1)	(2)	(3)	(4)
Evaluator's knowledge of first block	-0.014	0.003	0.002	0.003
	(0.080)	(0.007)	(0.006)	(0.007)
Experience in committees	0.086	-0.002	-0.010	-0.014
	(0.201)	(0.011)	(0.013)	(0.013)
Evaluators age	0.007	-0.005	0.010	0.014
	(0.213)	(0.012)	(0.014)	(0.013)
Male share in committee	0.047	-0.001	-0.003	-0.004
	(0.135)	(0.009)	(0.009)	(0.008)
Adjusted/Pseudo-R2	0.579	0.003	0.002	0.001
Predicted P		0.599	0.325	0.194
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. Bootstrapped standard errors, clustered by committee, are in parentheses.

Table 6: Performance in the first oral stage

	Grade		Success	
	Tobit estimates		Probit estimates	
	(1)	(2)	(3)	(4)
Multiple choice grade	5.569***	5.561***	0.287***	0.287***
	(0.398)	(0.401)	(0.015)	(0.015)
Multiple choice blanks	0.925***	0.924***	0.041***	0.041***
	(0.142)	(0.143)	(0.006)	(0.006)
Male	-0.295*	-0.297*	-0.021**	-0.021**
	(0.179)	(0.179)	(0.009)	(0.009)
Candidates experience=1	0.177	0.162	0.023	0.022
	(0.491)	(0.483)	(0.025)	(0.024)
Candidates experience=2	0.496	0.479	0.048**	0.048**
	(0.463)	(0.458)	(0.025)	(0.024)
Candidates experience$\geq$3	0.091	0.084	0.033*	0.033*
	(0.384)	(0.383)	(0.019)	(0.019)
Lives in Madrid	0.193	0.191	0.011	0.011
	(0.174)	(0.174)	(0.011)	(0.010)
Evaluator's knowledge of first block	-0.249**	-0.266*	-0.007	-0.007
	(0.117)	(0.137)	(0.005)	(0.007)
Experience in committees		-0.052		-0.009
		(0.274)		(0.013)
Evaluators age		0.246		0.014
		(0.335)		(0.015)
Male share in committee		-0.112		-0.004
		(0.213)		(0.010)
Pseudo-R2	0.073	0.073	0.130	0.130
Pred. P			0.170	0.170
N	9299	9299	9299	9299

Notes: The dependent variable in columns (1) and (2) is the grade obtained by the candidate at the first oral stage. Oral test grades range between 0 and 25 but as we only observe it for candidates whose grade was above 12.5, tobit regression is left-censored at 12.5, The dependent variable in columns (3) and (4) is a dummy variable that indicates whether the candidate passed the stage. The coefficients reported in these columns correspond to the marginal effects of a probit regression. Candidates exempt from the multiple choice test have not been included in the sample as there is no information on their performance in that stage. Bootstrapped standard errors, clustered by committee, are in parenthesis. *significant at 10%; **significant at 5%; ***significant at 1%.

Table 7: Performance in the second oral stage

	Grade		Success	
	Tobit estimates		Probit estimates	
	(1)	(2)	(3)	(4)
Multiple choice grade	2.304***	2.313***	0.186***	0.187***
	(0.402)	(0.401)	(0.035)	(0.035)
Multiple choice blanks	0.081	0.091	-0.022	-0.022
	(0.296)	(0.303)	(0.029)	(0.030)
Male	0.189	0.203	0.022	0.022
	(0.284)	(0.282)	(0.029)	(0.029)
Candidates experience=1	-0.711	-0.695	0.009	0.008
	(0.700)	(0.704)	(0.052)	(0.052)
Candidates experience=2	-1.343**	-1.326**	-0.038	-0.039
	(0.654)	(0.662)	(0.052)	(0.052)
Candidates experience≥3	-2.429***	-2.436***	-0.128***	-0.131***
	(0.590)	(0.595)	(0.047)	(0.047)
Lives in Madrid	0.193	0.212	0.037	0.038
	(0.241)	(0.242)	(0.028)	(0.028)
Evaluator's knowledge of first block	-0.045	-0.048	0.006	0.007
	(0.141)	(0.192)	(0.009)	(0.012)
Experience in committees		0.101		-0.020
		(0.396)		(0.021)
Evaluators age		-0.186		0.028
		(0.514)		(0.023)
Male share in committee		-0.104		-0.024*
		(0.370)		(0.014)
Pseudo-R2	0.017	0.018	0.035	0.036
Pred. P			0.383	0.383
N	1919	1919	1919	1919

Notes: The dependent variable in columns (1) and (2) is the grade obtained by the candidate at second oral stage. Oral test grades range between 0 and 25. Tobit regression is left-censored at 12.5, as we only observe grade for candidates that obtained a grade above 12.5. In columns (3) and (4) the dependent variable is a dummy that indicates whether the candidate passed that stage. The coefficients reported correspond to the marginal effects of a probit regression. The sample includes only candidates that passed the first oral stage. Bootstrapped standard errors, clustered by committee, are in parenthesis. *significant at 10%; **significant at 5%; ***significant at 1%.

Table 8: Performance in the overall exam

| | Grade | | Success | |
| | Tobit estimates | | Probit estimates | |
	(1)	(2)	(3)	(4)
Multiple choice grade	35.577***	35.587***	0.118***	0.118***
	(1.815)	(1.813)	(0.005)	(0.005)
Multiple choice blanks	3.148**	3.141**	0.009**	0.009**
	(1.337)	(1.339)	(0.005)	(0.005)
Male	-0.554	-0.573	-0.002	-0.002
	(1.761)	(1.758)	(0.006)	(0.006)
Candidates experience=1	3.764	3.746	0.015	0.015
	(3.898)	(3.864)	(0.015)	(0.015)
Candidates experience=2	3.104	3.092	0.013	0.013
	(3.481)	(3.456)	(0.013)	(0.013)
Candidates experience≥3	-2.480	-2.483	-0.006	-0.006
	(2.924)	(2.930)	(0.010)	(0.010)
Lives in Madrid	2.922*	2.899*	0.011*	0.011*
	(1.670)	(1.670)	(0.006)	(0.006)
Evaluator's knowledge of first block	-0.760	-0.690	-0.002	-0.002
	(0.558)	(0.644)	(0.002)	(0.002)
Experience in committees		-1.622		-0.007
		(1.702)		(0.006)
Evaluators age		2.296		0.009
		(1.807)		(0.006)
Male share in committee		-1.455		-0.005
		(1.080)		(0.003)
Pseudo-R2	0.061	0.061	0.127	0.127
Predicted P			0.054	0.054
N	9299	9299	9299	9299

Notes: The dependent variable in columns (1) and (2) is the final grade obtained by the candidate, which is the sum of grades obtained in stage one and in stage two. Grades range between 0 and 50. Tobit regression is left-censored at 25, as we only observe grade for candidates that obtained a grade above 25. The dependent variable in columns (3) and (4) is a dummy that indicates whether the candidate obtained the position. In the probit regressions reported in columns (3) and (4) we provide the marginal effects. Candidates exempt from the multiple choice test have not been included in the sample as we have no information on their performance in that stage. Bootstrapped standard errors, clustered by committee, are in parenthesis. *significant at 10%; **significant at 5%; ***significant at 1%.

Table 9: Similar-to-me effect

	OLS estimates	
	(1)	(2)
Evaluator's knowledge of first block	0.417**	0.412**
	(0.167)	(0.194)
Experience in committees		-0.073
		(0.307)
Evaluators age		0.171
		(0.344)
Male share in committee		0.068
		(0.293)
Adjusted-R2	0.495	0.493
N	743	743

Notes: The dependent variable in both columns is the grade obtained by the candidate in the initial multiple choice test. Both regressions include as controls candidates' gender, experience, geographical origin, the final ranking within the committee and a set year dummies, Robust standard errors, clustered by committee, are in parentheses. * indicates significant at 10%; ** significant at 5%; *** significant at 1%.