# Screening Discrimination in the Allocation of Graduate Scholarships

Vincent Chandler [1]
chandlev@econ.queensu.ca
Queen's University

## Abstract

This paper shows the presence of discrimination in the allocation of doctoral scholarships using a unique dataset containing the 1901 recipients of 2004-2005 doctoral awards from the Social Science and Humanities Research Council (SSHRC). Taking advantage of the random allocation of candidates to evaluators, I provide evidence that evaluators give higher scores to candidates in their own discipline. I then proceed to distinguish between two types of discrimination: taste-based and screening. In the first type, evaluators give the same preferential treatment to all candidates in their own field. In the second type, since evaluators can better assess candidates in their own field, better candidates benefit from having an evaluator from their discipline while weaker ones suffer from such a match. This paper is the first to find support for screening discrimination: better candidates have higher scores when they are assessed by evaluators from their discipline, while there is no significant effect for weaker candidates. Moreover, the weakest candidates, those close to the funding threshold, actually have a lower probability of receiving a large award when there is an evaluator from their discipline in their evaluation committee. Finally, the key assumption underlying screening discrimination is that evaluators can better assess candidates similar to themselves. I provide support for this assumption by showing that the scores of evaluators are a good predictor of whether candidates in their own field will become assistant professor seven years after receiving the award.

JEL Codes: I23, J70

November 2014

---

[1] The author wishes to thank SSHRC for accepting to share the scores of the applicants and is particularly endebted to Margaret Blakeney and Andreas Reichert for their help with the data. Furthermore, I appreciated useful comments from participants at the CEA meeting and those from Robin Boadway, Jennifer Hunt, Mike Kottelenberg, Steve Lehrer, Lealand Morin and Vincent Pohl. Financial assistance from the Ontario Graduate Scholarship and the Fonds de Recherche du Quebec - Societe et Culture is gratefully acknowledged. The usual caveat applies.

# 1 Introduction

Identifying the best candidate for a position is always challenging. To perform this exercise, evaluators try to extract as much information as possible from each application. This task is easier when evaluators share the same background as the candidate, because they can more easily interpret the information provided by applicants (Cornell and Welch, 1996). When evaluators assess candidates who do not share their background, they may be unable to assess the quality of the applicants precisely, and therefore give more weight to prior beliefs and assign these candidates scores close to the average. Under those conditions, if only the candidates with the highest scores are chosen, they will probably share the background of the evaluator. Even if evaluators have no taste for discrimination and no prior belief about differences in productivity, they may discriminate against candidates different from themselves, so-called screening discrimination.

Even though there exists a large body of literature on gender (e.g. Bagues and Esteve-Volart, 2010) and racial (e.g. Bertrand and Mullainathan, 2004, Oreopoulos, 2011, Parsons et al., 2011 and Price and Wolfers, 2010) discrimination, there is little research showing discrimination potentially due to screening abilities. Bagues and Villadoniga (2012) and Li (2012) are the exceptions. The former find that hiring committees in the Spanish judicial system tend to hire candidates with skills similar to their own, and the latter shows that NIH committees favor researchers with whom one of the members has previously co-authored. In both cases, it is unclear what is the mechanism. Members of the committee could prefer candidate similar to themselves or they may be better at screening them. In this paper, I study the individual scoring decision of evaluators assessing doctoral applicants to award scholarships. Due to the random allocation of candidates and the multidisciplinary nature of the selection committee, this institutional framework is ideal to study discrimination against people from a different background. As in the previous literature, discrimination could be caused by the informational advantage of evaluators when assessing candidates from their discipline or by a systematic taste-based bias against candidates from other disciplines. The objective of this paper is to distinguish between these two discrimination mechanisms.

I first provide empirical evidence for discrimination: evaluators give on average 0.18 more points (12.9% of a standard deviation) to candidates in their own discipline. This bias could either be due to a belief of superiority[2] or to screening discrimination[3]. To distinguish between both mechanisms, I go through 3 steps. First, I separate the sample into relatively stronger and weaker candidates. If the bias is due to screening discrimination, stronger candidates should benefit more than weaker ones from having an evaluator in their field who can recognize their true type. If the bias is due to a belief of superiority, however, both groups should benefit equally from an evaluator from their discipline. I find that evaluators give on average 0.29 points (20.9% of a standard deviation) more to stronger candidates in their field, while there is no statistically significant effect for weaker candidates, suggesting screening discrimination.

Second, I take advantage of the allocation of two types of scholarships. Students entering first- or second-year at a Canadian university are eligible for both the Canadian Graduate Scholarship (CGS) ($35 000 p.a) and for the SSHRC Doctoral Fellowship ($20 000). Since more than 85% of eligible candidates receive the larger scholarship, recipients on the funding threshold are relatively weaker. In the presence of screening discrimination, these students could actually suffer from having an evaluator from their discipline (Phelps, 1972). Indeed, I find that students with an evaluator from their discipline in their committee have a 8 percentage point lower probability of receiving the large award. This reversal cannot be explained by a belief of superiority and is a further argument for screening discrimination.

Finally, the fundamental assumption underlying screening discrimination is that evaluators can extract more information from candidates in their own discipline. The scores assigned by evaluators to candidates in their own discipline should therefore be good predictors of future success. Indeed, I find that a one-point increase in the average score given by evaluators to candidates in their discipline raises their probability of becom-

---

[2] Similar to Arrow (1973), evaluators may think that candidates in their discipline will be better researchers, because they use a methodology preferred by the evaluator.

[3] In the presence of screening discrimination, evaluators have the same prior beliefs about the distribution of talent. Accordingly, there should not be any difference between the average score awarded to both groups. This data set, however, only contains scholarship recipients; they are on average better than non-recipients. One should therefore expect a difference in average scores.

ing assistant professor by 1.8 percentage points. The scores given by evaluators outside the discipline of the candidate are uncorrelated with their job market outcomes but are correlated with their probability of completion.

Overall, this paper makes the following contributions to the literature in personnel economics. First, I show the presence of discrimination due to similarities in skills. This paper builds on Bagues and Villadoniga (2012) and Li (2012) by studying the decisions of single evaluators and not group decisions. Second, this paper is the first to show empirically that screening abilities can cause discrimination. The heterogeneous effect of discrimination across the distribution and the fact that evaluators do extract more information from candidates in their discipline provide evidence for screening discrimination.

The rest of the paper is organized as follows. In the next section, I present a theoretical framework to understand the differences between two types of discrimination. I then describe the setting in which the model will be tested: the selection procedure used by SSHRC to allocate scholarships. Section 4 provides empirical evidence for screening discrimination. In section 5, I turn my attention to the long term impact of funding on the outcome of students. Finally, I summarize the findings and present some policy implications.

# 2 Theoretical Framework

## 2.1 Basic Model

This model is adapted from Phelps (1972) and Bagues and Villadoniga (2013). The objective of an evaluator is to identify the candidates "i" with the most potential ($p_i$) which is assumed distributed $N(\bar{p}, 1)$. The quality of the application ($q_i$) of a candidates is a function of potential:

$$q_i = p_i + \eta_i \text{ where } \eta_i \sim N(0, \sigma) \tag{1}$$

If evaluators update their priors using Bayes rule, they will assign scores in this fashion:

$$\text{Score}_i = \text{E}[p_i|q_i] = \gamma q_i + (1-\gamma)\bar{p} \text{ where } \gamma = \frac{1}{1+\sigma} \tag{2}$$

Intuitively, when candidates send an application with a quality above the average, their scores will be above the average ($\bar{p}$). The extent to which an above-average signal leads to an above-average score depends on $\sigma$. If $\sigma$ is large, the signal is not very informative, and the score of a given candidate will be close to the average score. Conversely, if $\sigma$ is small, the signal is very informative, and a quality well above the average leads to a score well above the average.

## 2.2 Belief of Superiority

In the presence of a belief of superiority, evaluators discount the potential of candidates different (diff) from themselves:

$$\begin{aligned} q_i^{same} &\sim N(\bar{p}^{same}, 1) \\ q_i^{diff} &\sim N(\bar{p}^{diff}, 1) \\ \bar{p}^{diff} &< \bar{p}^{same} \end{aligned} \tag{3}$$

The scores are therefore determined in this fashion:

$$\begin{aligned} \text{Score}_i^{same} &= \gamma q_i + (1-\gamma)\bar{p}^{same} \\ \text{Score}_i^{diff} &= \gamma q_i + (1-\gamma)\bar{p}^{diff} \end{aligned} \tag{4}$$

The difference between the score given by an evaluator similar to the candidate and the score given by an evaluator different from the candidate is:

$$\begin{aligned} \text{Score}_i^{same} - \text{Score}_i^{diff} &= \gamma q_i + (1-\gamma)\bar{p}^{same} - \gamma q_i - (1-\gamma)\bar{p}^{diff} \\ &= (1-\gamma)(\bar{p}^{same} - \bar{p}^{diff}) \end{aligned} \tag{5}$$

This difference does not depend on $q_i$, and it is always positive, because $\sigma > 0$ (see figure 1).

## 2.3 Screening Discrimination

In the presence of screening discrimination, evaluators have no prior belief about the potential of candidates different from themselves. However, evaluators cannot extract as much information from candidates different from themselves as they can for candidates similar to themselves:

$$\begin{aligned} q_i &= p_i + \eta_i^{same} \text{ where } \eta_i^{same} \sim N(0, \sigma^{same}) \\ q_i &= p_i + \eta_i^{diff} \text{ where } \eta_i^{diff} \sim N(0, \sigma^{diff}) \\ \sigma^{same} &< \sigma^{diff} \end{aligned} \tag{6}$$

The scores are therefore determined in this fashion:

$$\begin{aligned} \text{Score}_i^{same} &= \gamma^{same} q_i + (1 - \gamma^{same})\bar{p} \\ \text{Score}_i^{diff} &= \gamma^{diff} q_i + (1 - \gamma^{diff})\bar{p} \\ \gamma^{same} &> \gamma^{diff} \end{aligned} \tag{7}$$

The difference between a score given by an evaluator similar to the candidate and a score given by an evaluator different from the candidate is:

$$\begin{aligned} \text{Score}_i^{same} - \text{Score}_i^{diff} &= \gamma^{same} q_i + (1 - \gamma^{same})\bar{p} - \gamma^{diff} q_i - (1 - \gamma^{diff})\bar{p} \\ &= q_i(\gamma^{same} - \gamma^{diff}) + \bar{p}(\gamma^{diff} - \gamma^{same}) \\ &= (\gamma^{same} - \gamma^{diff})(q_i - \bar{p}) \\ &= \frac{\sigma^{diff} - \sigma^{same}}{1 + \sigma^{diff} + \sigma^{same} + \sigma^{diff}\sigma^{same}}(q_i - \bar{p}) \end{aligned} \tag{8}$$

The first term is necessarily positive, because $\sigma^{diff} > \sigma^{same} > 0$ and because the denominator is positive. The sign of the second term, however, depends on the relative ranking of the candidate. Candidates above the average will benefit from having an evaluator who will recognize their true potential. Candidates below the average, however, suffer from having an evaluator similar to themselves. Contrary to the superiority hypothesis, the bias induced by screening discrimination depends on the relative ranking of the candidate (see figure 1).

# 3 Background and Data

## 3.1 SSHRC and the Selection Procedure

The Social Science and Humanities Research Council (SSHRC) is a Canadian federal agency that promotes and supports postsecondary-based research and training in the humanities and social sciences. One of the ways through which SSHRC achieves this goal is by awarding two types of scholarships to doctoral students. First, the Joseph-Armand Bombardier Canada Graduate Scholarship (CGS) provides recipients with three annual payments of \$35 000. Only students entering the first or second year of doctoral studies at a Canadian university are eligible for CGS. Second, the SSHRC Doctoral Fellowship (SDF) represents a yearly payment of \$20 000 per year up to the fourth year of doctoral studies [4]. All students entering fourth year or below are eligible for SDF.

Applicants submit the same application for both scholarships: a project proposal, a CV, two reference letters from faculty, and all their university transcripts. For students enrolled at a Canadian university, these applications are submitted to the university pre-selection committee. Each university is provided with a quota restricting the number of students that can be forwarded to the national competition. This quota is adjusted regularly to take into account the previous success of the university. Students enrolled at foreign universities submit their application directly to SSHRC in the preliminary competition. The top-ranked candidates of the university pre-selection and those from the preliminary competition at SSHRC are forwarded to the national competition. The scores of candidates at the university or preliminary competitions are not revealed to the evaluators at the national competition.

---

[4] If students receive the scholarship in their first year of doctoral studies, they will be awarded overall \$80 000. Similarly, students wining a scholarship in 4th year will only receive one payment of \$20 000.

At the national competition, SSHRC creates 5 multidisciplinary committees. Committee 5, for example, comprises students in economics, industrial relations, law, management, business and political science[5]. Each committee is broken down into 3 or 4 subcommittees each with 3 evaluators. Applicants are randomly allocated to a subcommittee within their own committee. In 2004, there were 16 subcommittees (5 committees with each 3 subcommittees except psychology which had 4 subcommittees) with on average 65 recipients per subcommittee. In 2005, there were 19 subcommittees (5 committees with each 4 subcommittees except the fifth committee which had 3 subcommittees) with about 48 recipients per subcommittee. On average, a subcommittee allocated \$3.7 million in scholarships.

Within each subcommittee, three evaluators each assign a score to all candidates in the subcommittee with a maximum of 10 and a minimum of 0. Evaluators are given a pre-assigned distribution of scores. For example, only one 10 and one 0 can be awarded and the average score should be close to 7.5. Evaluators do not know the identity of other evaluators when evaluating candidates[6] and are not allowed to assess candidates from their own university. The scores given by the three evaluators are then added to calculate the final score according to which students are ranked. The eligible[7] students in the top tier of a subcommittee are awarded a CGS, the second-tier obtains a SDF, and the last tier does not receive a scholarship. The thresholds are defined by civil servants based on the budget allocated to the subcommittee. In the sample, these thresholds vary from 16.4 to 20.2 across subcommittees.

The funding decisions and the individual scores are communicated to the students in May/June. About half the applicants receive some scholarship. Students entering first or second year who are above the CGS threshold are informed that they could receive CGS should they decide to study in Canada. If they decide to study abroad, they are awarded

---

[5] Candidates are not always allocated to a committee based on their own discipline, but based on the discipline closest to their project proposal. For example, if an applicant in philosophy writes a proposal on the epistemology of economics, the application will probably be considered by committee 5 and not by committee 2, which usually evaluates candidates in philosophy.

[6] Evaluators meet once they have evaluated all candidates.

[7] Those starting first or second year of a doctoral studies at a Canadian university.

a SDF. Generally, recipients of scholarships may not cumulate other scholarships[8], but they can work a maximum of 450 hours per 12-month period. Every year recipients must complete an annual report administered by the university. SSHRC does not set any specific requirements for the renewal of the scholarship, making the non-renewal of the award unlikely.

## 3.2 Data

I have assembled data on the universe of recipients of the 2004 and 2005 national competitions to capture long term outcomes. More specifically, I have information on scholarship recipients (year of study, discipline, subcommittee, destination university and award received) and on evaluators (discipline and score for each candidate in the subcommittee)[9].

Table 2 shows a very similar distribution of disciplines for recipients and evaluators by discipline. Figure 2 presents the distribution of scores for students in the same discipline as the evaluator and for students in a different discipline. If evaluators can better assess candidates in their own discipline, there should be a greater variance in the scores awarded to students in their own discipline. There is no evidence of a difference in variance. There are two reasons to possibly explain the lack of difference in variance. First, the weakest candidates are not in the sample, because they did not receive a scholarship. They probably would have received very low scores from evaluators from their own discipline but not from evaluators outside their discipline. Second, the true scores are censored, because evaluators cannot give scores above 10. Evaluators may have wished to award extraordinary candidates in their discipline a score above 10, but it was impossible. Evaluators outside the discipline of these candidates may have recognized their excellence and given them a 9.5. Due to the truncation of the sample and institutional censoring, any large difference could have disappeared.

The scores are not only evidence of possible discrimination, they are also used to explain long term outcomes of the scholarship recipients. Jacob and Lefgren (2011) and

---

[8] SSHRC allows cumulating scholarships with non-federal funding agencies, but the two main provincial funding agencies (Ontario Graduate Scholarship and the Fonds Quebecois de Recherche et Societe et Culture) do not allow cumulation of awards.

[9] These data were received from SSHRC through the Access to Information and Privacy Act.

Li (2012) both define their outcome as the quantity/quality of publications. Such a measure would be inappropriate in this study, because the recipients come from different disciplines with different publication cultures. While it is normal for a recent graduate in economics not to have any publication, a recent graduate in psychology may have a dozen publications. Furthermore, journals are ranked differently across disciplines, and it would be very challenging to reconcile these rankings for all social sciences and humanities. For these reasons, I choose two other measures of success: completion and employment.

Completion of the doctoral program is a short-term objective. One would expect all recipients to complete the thesis for which they were funded[10]. In fact, only 34.6 percent of all recipients completed their doctoral program in 5 years or fewer, and 65.1 percent of all recipients took seven years or fewer. These results may be conservative, because the information stems from the catalogs of university libraries. Some students may have completed their dissertation, but it may not have been added to the library catalog yet. Whenever I could not find a dissertation, but a department announced the completion of the program, I used the date announced by the department.

Promoting employment in research is the long-term objective of SSHRC. I use employment as tenure-track assistant professor as a proxy to capture research productivity[11]. Assistant professors have been chosen by hiring committees on the basis of their research quality and potential, and they are expected to conduct research to receive tenure and promotion. Assistant professors are therefore very likely to make a more long-lasting and valuable contribution to research than people in other professions. To determine who is an assistant professor seven years after receiving an award, I searched for the 1901 recipients on the internet[12]. I could not find 161 applicants (8.5 percent of the sample). Of those, I could not find evidence of completion for 85 applicants at the time of the search, making it unlikely for them to be assistant professors[13]. I assume that the remaining 76 (4 percent of the sample) are not assistant professors, because I would otherwise have

---

[10] I do not study whether they complete the project submitted in the application.

[11] The results are robust to including post-docs within this definition.

[12] The search for the 2004 applicants took place in December 2011, and the search for the 2005 applicants took place in December 2012

[13] The results stay the same if I exclude missing observations or if I code them as not-professor.

found them on departmental web sites or publications[14]. Within all assistant professors, I do not distinguish between universities, mostly because it would be very challenging to rank universities for each discipline. Some smaller universities may be, for example, better than large universities in certain niche disciplines. Furthermore, most candidates are employed at small teaching universities for which departmental rankings in terms of research are less reliable.

Table 3 presents some summary statistics for covariates and for all three dependent variables. By definition, all CGS holders studied in Canada. A large share of SDF holders in first or second year were probably above the threshold but decided to study abroad, thus explaining the small share of SDF holders studying in Canada in first or second year. The majority of recipients are female which is consistent with a broad trend in social sciences and humanities. It is difficult to interpret completion rates and the probability of being assistant professor across years of study. On the one hand, students receiving the scholarships in fourth year were probably turned down three times suggesting that students applying in fourth year are probably worse students applying in first year. On the other hand, evaluators have more information for candidates in fourth year than they do for candidates in first year, making their funding decision much more precise. From the pool of weaker candidates, evaluators may be able to precisely identify the most promising ones. It is unclear which of these two effects is the strongest. The probability of completing in 5 years is highest for second-year recipients, while recipients entering fourth year have the highest probability of finishing in seven years. Finally, the probability of being assistant professor is also the greatest for students in fourth year. This last result could simply be due to a time advantage. As I searched for all candidates at the same time (ie 7 years after receiving the award), first-year recipients had 7 years following their first year in the doctoral program to get a position, while fourth-year recipients had 11 years.

---

[14] Table 1 reports the coefficients of the regression: $\text{missing}_i = \beta_0 + \beta_1 \text{gender}_i + \beta_2 \text{total score}_i + \beta_3 \text{PhD Location}_i + \beta_4 \text{Year of Study} + \beta_5 \text{Discipline Dummies} + u_i$ There is a higher probability of finding someone with a high score who would have a higher probability of being assistant professor suggesting that the assumption that missing observations are not assistant professors is reasonable.

# 4 Discrimination

## 4.1 Evidence of Discrimination on the Scores

If evaluators give on average higher scores to candidates in their own discipline, then $\beta_1$ in the following regression will be positive and statistically significant:

$$score_{ij} = \beta_0 + \beta_1 \mathbb{1}\{\text{Discipline}_i = \text{Discipline}_j\} + v_i + z_j + \epsilon_{ij} \tag{9}$$

where “i” represents the recipient and “j” the evaluator.

The term $v_i$ captures unobserved candidate characteristics including skills. I control for this unobserved heterogeneity by including candidate fixed effects. The term $z_j$ captures systematic differences between evaluators. Such differences are unlikely to occur, because SSHRC gives precise indications to evaluators. However, there could still be systematic differences due to the truncation of the sample.

In the presence of screening discrimination, the magnitude of $\beta_1$ should vary across applicants as illustrated in figure 3: it should be larger for the best candidates than for the weaker candidates. Conversely, if evaluators simply prefer candidates in their discipline, the magnitude of the discrimination should not vary across the distribution of students. To test this idea, I divide the sample into two categories: recipients below the median of the sample and those above the median. I then estimate equation 10 separately for each group. One could divide the sample naively by using the total score as a proxy of talent. However, such a method would would lead to inconsistent estimator. Indeed, the independent variable would be correlated with $\epsilon_{ij}$, because candidates assessed by an evaluator from their discipline who also had a large positive $\epsilon_{ij}$ would have a higher probability to be above the median. Similarly, candidates not assessed by an evaluator from their discipline who also had a large negative $\epsilon_{ij}$ would have a higher probably of being below the median. In both cases, there would be a negative correlation between $\epsilon_{ij}$ and the independent variable thus biasing the coefficient. To address this issue, I predict who will be below or above the median using the following characteristics: SSHRC scholarship for M.A., gender, discipline and location of PhD. By doing so, I disconnect the error term $\epsilon_{ij}$ from the selection into either group. Applicants whose probability of being above the median is larger than 50% are in the sample used to estimate equation 10 for above-median candidates and those whose predicted probability is below 50% would be in the sample for below-median candidates.

Discrimination may not be present when candidates and evaluators in the same discipline have little in common. In committee 1, for example, evaluators and students in fine arts could have very different interests ranging from drama to photography with some composing music. Similarly, evaluators and students in literature could be studying modern Russian literature or medieval Spanish literature. In both cases, it is unclear whether an evaluator in music composition will better assess a student in photography than would an evaluator in literature. To address this issue, equation 10 is also be estimated without the first committee which includes fine arts and literature.

Table 4 provides the estimates for the coefficients of equation 10. Models 1 includes the whole sample, and shows that evaluators tend to give approximately 0.12 points more to candidates in their discipline. Model 2 excludes candidates from the first committee, and the coefficient of interest increases to 0.18 suggesting that I may not be capturing similarity very well in this committee. Model 3 restricts the sample to above or at median candidates in the naive fashion, and the coefficient almost triples to 0.30, while the coefficient loses its significance when the sample is restricted to below median candidates. Separating the sample using the sophisticated method leads to very similar conclusions. The coefficient of interest is 0.29 when the sample is restricted to the best candidate and is also statistically insignificant when only the weaker students are considered. These results cannot be explained using only a belief of superiority; only screening discrimination can justify them.

## 4.2 Evidence of Discrimination on the Allocation

As much as the scores reflect the decisions taken by evaluators, the most relevant outcome is the final allocation of scholarships. For candidates eligible[15] for the large scholarship (CGS), the presence of an evaluator from their discipline on their subcommittee could affect their chances of receiving it. To assess this impact, I estimate the following linear probability model[16] on eligible recipients:

$$cgs_i = \alpha_0 + \alpha_1 \mathbb{1}\{\text{At Least One Evaluator from Same Discipline on Subcommittee}_i\} + \alpha_2 X_i + u_i \tag{10}$$

The control variables ($X_i$) included in the regression include gender, year of study when the candidate received the scholarship and 27 discipline dummies.

If the coefficient $\alpha_1$ is statistically significant, discrimination will not only impact scores, but will also affect the type of scholarships allocated. In the presence of a belief of superiority, $\alpha_1$ should be positive. In the case of screening discrimination, the sign is not obvious, because recipients close to the funding threshold are weak.

The coefficient $\alpha_1$ will consistently estimate discrimination if $u_i$, the unobserved skills of the candidates, is uncorrelated with the probability of having an evaluator from one's discipline on the subcommittee. Anecdotal evidence suggests that candidates are allocated alphabetically to subcommittees[17]. In such a case, the skills of candidates would be uncorrelated with their probability of being evaluated by someone from their own discipline. Even though it is impossible to verify this claim, it is possible to determine whether the probability of being evaluated by someone from one's own discipline varies by total scores. If SSHRC were to allocate the best candidates to subcommittees containing at least one evaluator from their discipline, the probability of having an evaluator from one's discipline should increase with the total score as a proxy for talent. The best applicants would have a high probability of being matched with an evaluator from their discipline,

---

[15] Recall that these would be candidates entering first or second year of their doctoral program at a Canadian university.

[16] The results are robust to a probit regression.

[17] The first candidate goes to the first subcommittee, the second, to the second one, the third, to the third one, the fourth, again to the first subcommittee and so on.

and the worst applicants should have a low probability. Table 5 shows no relationship between total score and the probability of being evaluated by someone from one's own discipline.

The coefficients of equation 11 are provided in table 6. Students facing a subcommittee composed of at least one evaluator from their discipline have a lower probability of receiving the large scholarship by 8 to 9 percentage points. This result cannot be explained by the belief of superiority, but it could stem from screening discrimination since candidates close to the funding threshold between small and large scholarships are weak[18], we would expect them to suffer from screening discrimination. Indeed, evaluators from their discipline may be better able to recognize their latent research potential, which would be below the expectation of evaluators outside their discipline. If screening discrimination leads to such a negative impact for below-average candidates, it is surprising not to find a significant negative impact when the sample is restricted to below-average candidates in table 4 (model 6). Perhaps evaluators discuss their scores when candidates are close to the perceived threshold. Contrary to above-average candidates, scores of candidates below the average matter for the allocation of scholarships. Evaluators may therefore feel the need to consult their peers to be confident of their score. If out-of-discipline evaluators learn about the score given by in-discipline evaluators and trust them, then the difference between the scores would shrink and lead to non-significant differences.

## 4.3 Predictive Power of Scores

The assumption underlying screening discrimination is that evaluators are better at assessing candidates from their own discipline. Scores of evaluators should therefore be statistically significant predictors of success for candidates within their discipline but more weakly so for those outside their discipline. In other words, the average score given by evaluators in the same discipline as the candidate should be correlated with completion or job market outcomes, while the average score given by evaluators outside the discipline of the candidates should not be. Interestingly, these two averages are not highly

---

[18] The funding thresholds are between 16.4 to 20.2, while the average total score in the sample is 21

correlated (correlation = 0.1982). This lack of correlation suggests that even though both types of evaluators appraise the same applications, they look at them very differently. Unfortunately, it is impossible to calculate such an average for candidates who were not evaluated by someone from their discipline. I therefore lose 475 observations from very small disciplines[19]. These exclusions do not create a selection bias, because the selection rule is based on observables (disciplines) for which I can control in the regression.

The following linear probability model[20] regression assesses the explanatory power of both averages on the three dependent variables (completion in 5 years, completion in 7 years and tenure-track assistant professorship):

$$\begin{aligned} \mathrm{Y}_i = &\gamma_0 + \gamma_1 \text{Average Score In-Discipline}_i \\ &+ \gamma_2 \text{Average Score Out-of-Discipline}_i \\ &+ \gamma_3 X_i + u_i \end{aligned} \tag{11}$$

The control variables ($X_i$) in the regression include gender, year of study when the candidate received the scholarship and whether the candidate was enrolled at a Canadian university.

The explanatory power of scores given by evaluators to candidates in their discipline may vary according to the year of study of the candidate. Candidates applying for funding before starting graduate school submit very tentative research proposals. It is therefore very difficult for evaluators to predict success on the basis of this information even if the candidate is in their own discipline. Conversely, students in their upper-years can send a stronger signal about their research potential and therefore enable evaluators within their discipline to make a better judgment. The predictive power of the score given by evaluators to candidates in their own discipline should therefore increase as students progress. In other words, $\gamma_1$ should increase as the sample is restricted to upper-year students. In contrast, evaluators assessing candidates outside their discipline may not

---

[19] Folklore, for example, was not represented in the evaluation committee.

[20] The results are robust to a probit regression.

gain as much information when students progress. The magnitude of the coefficient $\gamma_2$ should therefore not change significantly when the sample is restricted to students in upper-years. To test both ideas, I perform the regression on three different samples: all students, second to fourth year students, and third and fourth year students.

Table 7 shows that a one-point increase in the average given by evaluators from one's own discipline increase the probability of being assistant professor seven years after receiving the award by 1.8 percentage point. This effect increases to 2.8 percentage point when first-year students are excluded from the sample. Even though this difference is not statistically significant, it does suggest that upper-year students provide more information to evaluators than first-year students who have very little research experience and whose research proposals are probably very preliminary. The scores of evaluators outside the discipline have no predictive power. Overall, this result shows that evaluators are better at assessing the research potential of candidates in their own discipline, which is a key assumption of the screening discrimination model.

The scores given by evaluators to candidates outside their discipline may not be a good predictor of labor market outcomes, but they are not devoid of information. Table 8 shows that the scores of out-of-discipline evaluators are statistically significant when predicting whether candidates finish their dissertation in five or seven years. Interestingly, the coefficient does not change when the sample is restricted to upper-year students. This invariance in the magnitude of the effect could suggest that the quality of the information used by evaluators for candidates outside their discipline does not vary as students progress. The only element of the application package that remains constant across years is the undergraduate academic transcript, and I would postulate that out-of-discipline evaluators rely on these to score students. Furthermore, certain skills required to complete a doctoral program in a timely fashion such as persistence and hard work are probably similar to those necessary to keep excellent marks at the undergraduate level.

Overall, these findings show that discrimination in the allocation of scholarships positively affects strong candidates, but it is detrimental for weaker candidates. The evidence also suggests that evaluators are better at forecasting labor market outcomes for candidates within their discipline. Both results are consistent with screening discrimination.

# 5 Conclusion

This paper tests for the presence of two types of discrimination in the allocation of graduate scholarships and shows that evaluators behave differently with candidates from their own discipline. Strong applicants benefit from having an evaluator from their discipline, while weak candidates suffer. These results can be explained by screening discrimination.

The results of this paper have some policy implications for agencies granting graduate scholarships. The composition of the selection committee matters. Evaluators are better at accessing the research potential of candidates in their own field. Single-discipline committees would increase the precision of the selection procedure: candidates with the most research potential will receive the highest scores and consequently the scholarship. Ultimately, reducing the noisiness of the selection procedure means creating incentives for students to perform better ex-ante. If students think that the selection procedure is very noisy, there is no incentive to work hard, because it may not be recognized by an evaluator outside their discipline. Unfortunately, this increase in precision comes at a price. Single-discipline committees would involve more evaluators and a greater burden for the granting agency. A compromise may be to insure all candidates one evaluator from their discipline and weigh the scores differently.

Future research will investigate the role of gender in the scoring decision. A burgeoning literature has documented some evidence of gender discrimination at the committee level (e.g. Bagues and Esteve-Volart, 2010), but little work has been done at the individual level. Gender differences could affect the allocation of scholarships and then possibly explain gender differences in academia. Moreover, the data set will make it possible to study the explanatory power of scores given by gender. The style in which project proposals are written could be gender dependent making it easier for evaluators to assess candidates of the same gender. More research can also be conducted on the impact of funding. The discontinuity between recipients and non-recipients could provide more intuition on the role of money on the productivity of graduate students. Moreover, using the cohort of 2003 (pre-policy), it would be possible to compare the outcomes of SDF recipients in 2003 with the ones of CGS recipients in 2004-2005. Such an analysis would study the impact of funding throughout the distribution, and not only at the cutoff.

# 6 Bibliography

Angrist, J., Lang, D. and Oreopoulous, P., 2009. Incentives and services for college achievement: Evidence from a randomized trial. American Economic Journal: Applied Economics 1(1) , 136-163(28).

Arrow, K.J., 1973. The Theory of Discrimination. In Orley Ashenfelter and Albert Rees, eds. Discrimination in Labor Markets. Princeton, N.J.: Princeton University Press. 3-33.
Bagues, M. and Esteve-Volart, B., 2010. Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. Review of Economic Studies 77, 1301-1328.

Bagues, M. and Villadoniga, M.J.P., 2012. Do recruiters prefer applicants with similar skills? Evidence from randomized natural experiments. Journal of Economic Behavior & Organization 82, 12-20.

Becker, G., 1957. The Economics of Discrimination. University of Chicago Press, Chicago.

Bertrand, M. and Mullainathan, S., 2004. Are Emily and Gregg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. American Economic Review 94(4), 991-1013.

Cornell, S. and Welch, I., 1996. Culture, information and screening discrimination, Journal of Political Economy 104(3), 542-71.

List, J.A. 2004. The nature and extent of discrimination in the marketplace: Evidence from the field. Quarterly Journal of Economics 119(1), 48-89.

Jacob, B and Lefgren, L., 2011. The impact of NIH research grant funding on scientific productivity. Journal of Public Economics 45 (9-10), 1168-1177.

Levitt, S. 2004. Testing theories of discrimination: Evidence from weakest link. Journal of Law and Economics 47(2), 431-453.

Li, D., 2012. Information, Bias and efficiency in expert evaluation: Evidence from the NIH, Working Paper.

Mobius, M. and Rosenblatt, T. 2006. Why beauty matters. American Economic Review

96, 222-235.

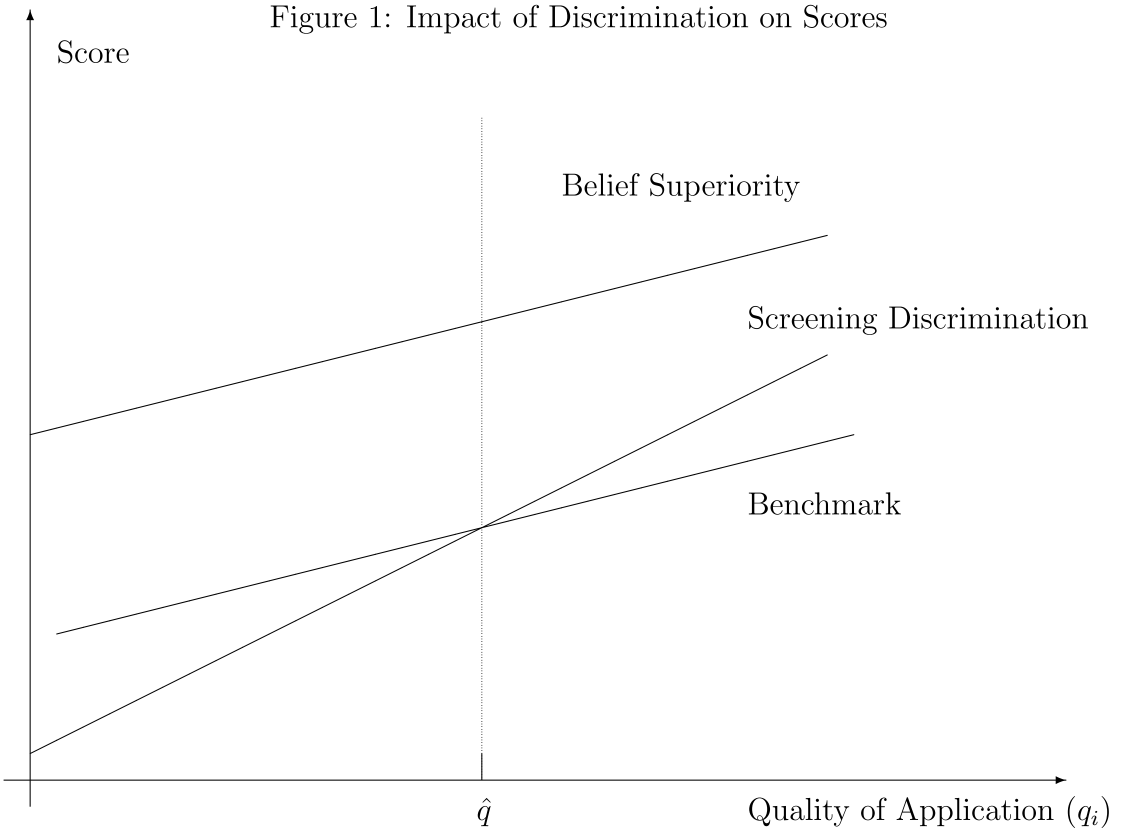
Oreopoulos, P., 2011. Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes. AEJ: Economic Policy 3(4), 148-171

Parsons, C.A. et al., 2011. Strike three: discrimination, incentives, and evaluations. American Economic Review 101(4), 1410-1435.

Phelps, E.S., 1972. The Statistical Theory of Racism and Sexism. American Economic Review 62(4), 659-661.

Price, J., and Wolfers, J., 2010. Racial discrimination among NBA referees. Quarterly Journal of Economics 125(4), 1859-1887.

Figure 1: Impact of Discrimination on Scores



**Description:** The benchmark case shows graphically the relationship: $\text{Score}_i = \gamma q_i + (1 - \gamma)\hat{p}$. In the presence of taste-based discrimination, the prior beliefs on the quality of students is different for both groups. The relationship between the quality of the proposal and the score is therefore shifted upwards for students who benefit from this bias. In the presence of screening discrimination, the noise in the signaling function shrinks for candidates who benefit from this bias. The slope of there function therefore becomes steeper and closer to 1.

# 7 Figures

Figure 2: Distribution of Scores

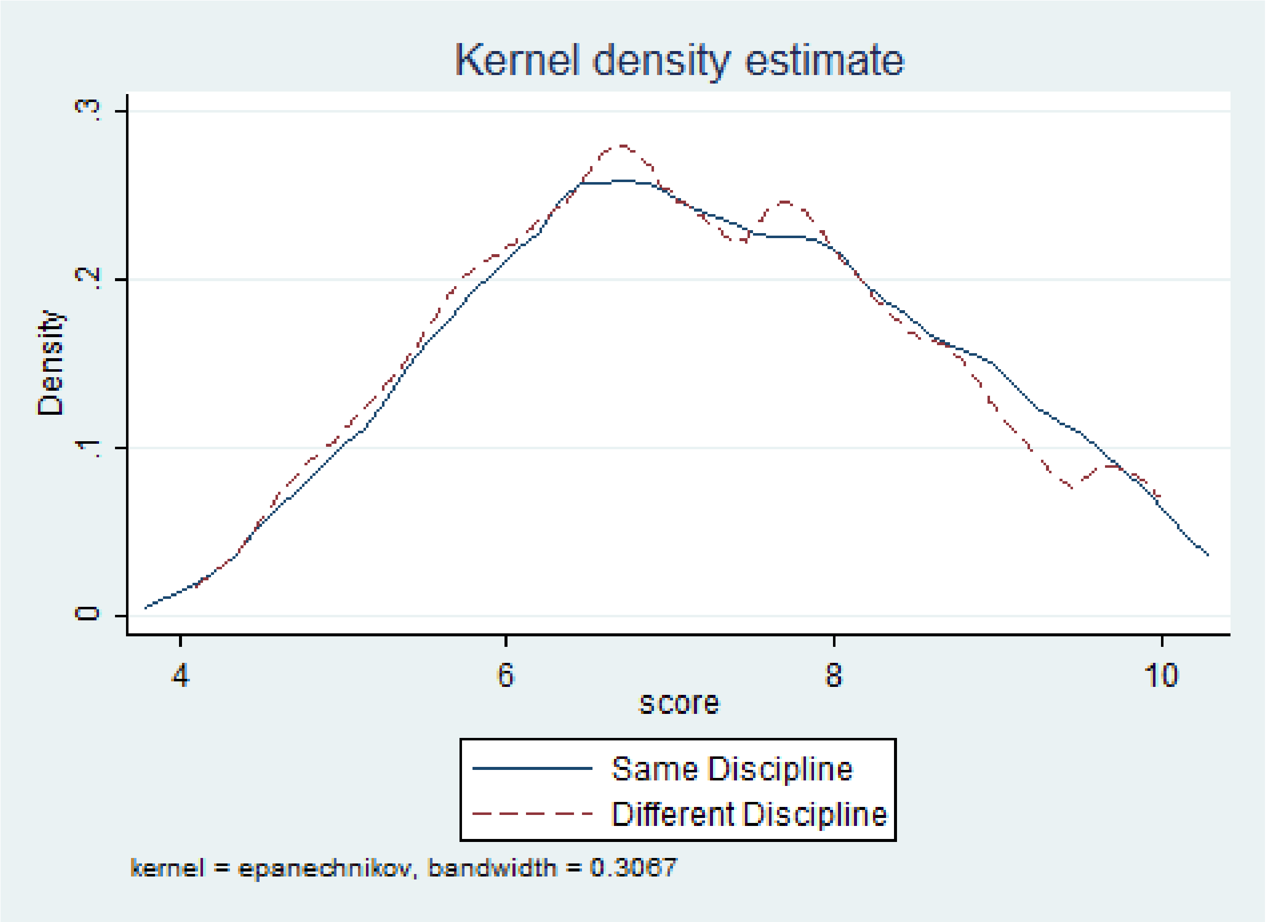


Figure 3: Relationship between the Quality of Applications and Scores in Truncated Sample

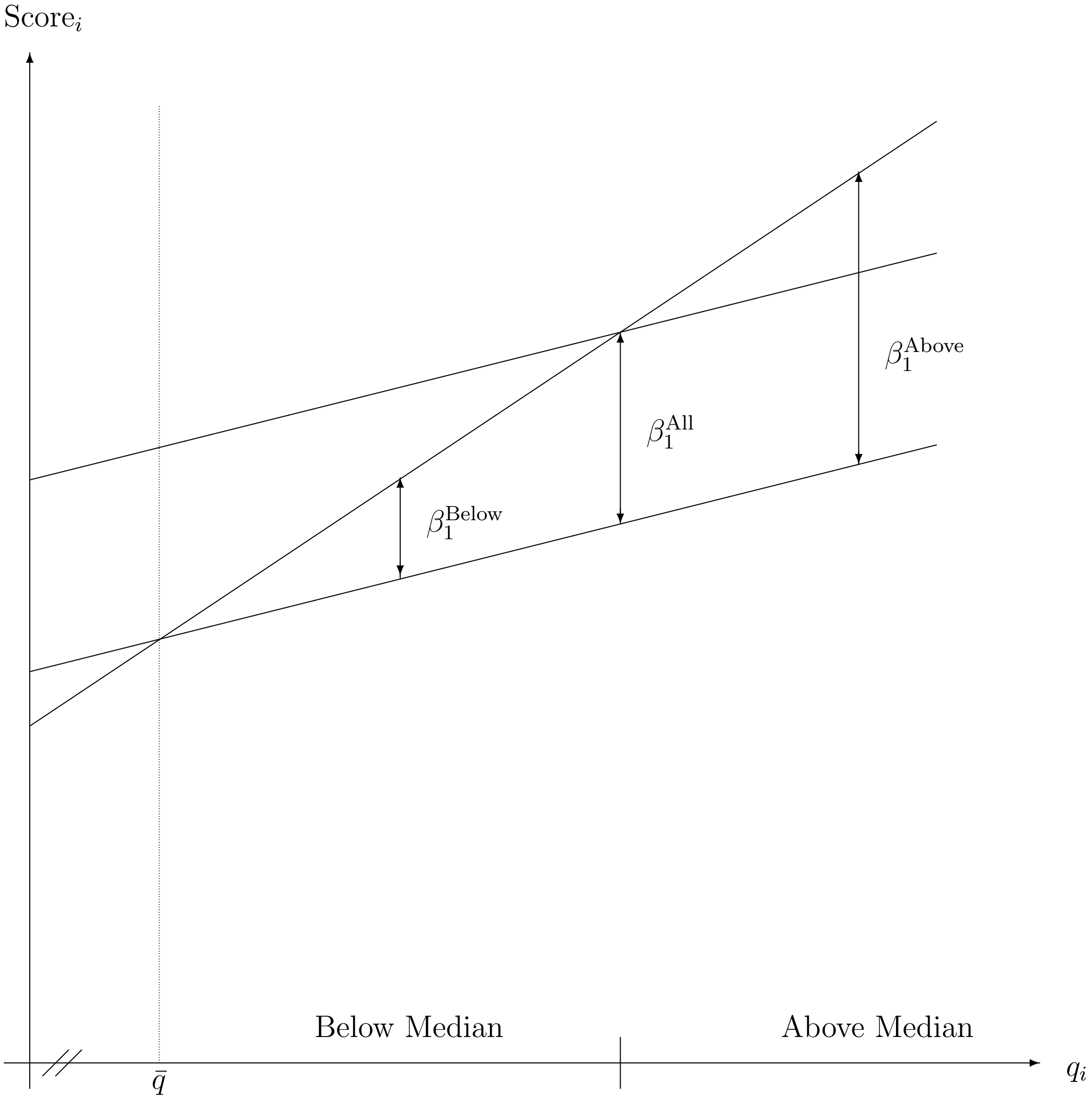


Table 1: Linear Probability Model Explaining Missing Observations

| | |
|---|---|
| Male | -0.0070 |
| | (0.0096) |
| Total Score | $-0.0037^{**}$ |
| | (0.001) |
| PhD in Canada | $0.0230^{**}$ |
| | (0.0099) |
| Year of Study Dummies | Yes |
| Discipline Dummies | Yes |
| $N$ | 1901 |
| $R^2$ | 0.021 |

(Robust standard errors)

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

**Note**: This table shows the coefficients of the equation predicting a missing observation (=1).

# 8 Tables

Table 2: Discipline Representation

| Discipline | Share of Recipients | Share of Evaluators |
|---|---|---|
| Anthropology | 4.94 | 4.76 |
| History | 11.15 | 8.57 |
| Communications | 2.84 | 3.81 |
| Psychology | 16.78 | 10.48 |
| Geography | 3.10 | 2.86 |
| Economics | 2.10 | 3.81 |
| Education | 7.10 | 6.67 |
| Fine Arts | 5.31 | 6.67 |
| Sociology | 7.15 | 3.81 |
| Management | 3.16 | 5.71 |
| Interdisciplinary | 2.52 | 1.9 |
| Law | 2.37 | 4.76 |
| Linguistics | 2.10 | 6.67 |
| Literature | 13.05 | 12.38 |
| Philosophy | 5.16 | 7.62 |
| Political Science | 6.79 | 4.76 |
| Religious Studies | 2.47 | 3.81 |
| Urban Studies | 1.89 | 0.95 |
| Sample | 1901 | 105 |

**Note**: Overall, there are 27 disciplines. To simplify this table, some disciplines were merged. Anthropology also contains archeology. History also contains classics and medieval studies. Psychology contains social work and criminology. Geography also contains demography. Sociology also contains folklore. Management also contains library science and industrial relations.

Table 3: Summary Statistics

| | Year of Study | | | | | | |
|---|---|---|---|---|---|---|---|
| | All | First-Year | | Second-Year | | Third-Year | Fourth-Year |
| | | SDF | CGS | SDF | CGS | SDF | SDF |
| *C*ovariates | | | | | | | |
| PhD in Canada | 84.0 | 35.1 | 100 | 43.0 | 100 | 88.6 | 81.7 |
| Male | 37.7 | 43.6 | 33.9 | 36.1 | 39.6 | 38.3 | 36.3 |
| *O*utcomes | | | | | | | |
| PhD in 5 years | 34.8 | 24.2 | 34.4 | 40.5 | 39.4 | 35 | 30.8 |
| PhD in 7 years | 66.1 | 59.4 | 62.7 | 62.6 | 66.7 | 65.8 | 75.8 |
| Assistant Professor 7 yrs after receiving award | 31.0 | 21.8 | 19.3 | 29.1 | 30.4 | 34 | 48.7 |
| Sample | 1901 | 165 | 357 | 158 | 444 | 500 | 273 |

**Note**: All numbers are percentages. The sample consists of all 2004 and 2005 recipients of SSHRC doctoral awards. SDF stands for SSHRC Doctoral Fellowship, and CGS stands for Canadian Graduate Scholarship.

Table 4: Fixed Effect Model Explaining the Scores

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | | | Naive | | Sophisticated | |
| | All Sample | Exclude 1st Committee | Above Median | Below Median | Above Median | Below Median |
| Own Field | 0.119** | 0.180*** | 0.296*** | 0.0707 | 0.287*** | 0.0468 |
| | (0.0482) | (0.0564) | (0.0822) | (0.0795) | (0.104) | (0.0960) |
| Constant | 7.057*** | 7.060*** | 7.826*** | 6.323*** | 7.140*** | 6.970*** |
| | (0.0342) | (0.0368) | (0.0529) | (0.0460) | (0.0414) | (0.0522) |
| Candidate FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Evaluator FE | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 5703 | 4599 | 2340 | 2325 | 2412 | 2187 |
| $R^2$ | 0.017 | 0.019 | 0.040 | 0.035 | 0.046 | 0.033 |

(Robust standard errors)

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

**Note**: Models 1 is estimated using all candidates. Models 2-6 exclude candidates from the first committee, because the discipline of evaluators in this committee was unclear. Models 3 and 5 restricts the sample to candidates above the median, and models 4 and 6 restricts it to candidates below the median. The naive estimation method simply uses the total score to distinguish between below and above median recipients, while the sophisticated method first predicts who will be above or below the median. All regressions include candidate and evaluator fixed effects.

Table 5: Probability of Having an Evaluator from Own Discipline by Total Score

| Score | Probability | Sample Size |
|---|---|---|
| 27 or more | 73.6 | 68 |
| 26 | 73.8 | 61 |
| 25 | 71.7 | 99 |
| 24 | 74.1 | 112 |
| 23 | 76.3 | 181 |
| 22 | 74.1 | 191 |
| 21 | 75.8 | 223 |
| 20 | 73.4 | 252 |
| 19 | 74.6 | 273 |
| 18 | 74.7 | 257 |
| 17 or less | 80.4 | 184 |

**Note**: This table shows the probability of being evaluated by at least one evaluator from one's own discipline for different total scores. There is no evidence suggesting that high-skilled applicants as proxied by high scores have a higher probability of being evaluated by someone from their own discipline.

Table 6: Linear Probability Model Explaining the Probability of Receiving the Higher Award

| | (1) | (2) | (3) |
|---|---|---|---|
| At Least One In-Discipline Evaluator | -0.0883** | -0.0885** | -0.0935*** |
| | (0.0349) | (0.0349) | (0.0360) |
| Male | -0.00140 | | |
| | (0.0244) | | |
| First Year | -0.00316 | | |
| | (0.0233) | | |
| Constant | 0.845*** | 0.842*** | 0.842*** |
| | (0.0871) | (0.0849) | (0.0852) |
| $N$ | 927 | 927 | 753 |
| $R^2$ | 0.063 | 0.063 | 0.071 |

(Robust standard errors)

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

**Note**: Models 1 to 2 are estimated using all candidates eligible for CGS (first- or second-year students in Canada). Models 3 omits candidates in the first committee, because the discipline of evaluators in this committee was unclear. There are 27 discipline dummies

Table 7: Linear Probability Model Using the Scores Given by Evaluators within/outside the Candidate's Discipline to Explain the Probability of being Assistant Professor 7 years after Receiving the Award

| | (1)<br>All | (2)<br>Excluding 1st Year | (3)<br>3rd and 4th Year |
|---|---|---|---|
| Avg Score In Discipline | 0.0180** | 0.0278** | 0.0271* |
| | (0.00873) | (0.0108) | (0.0148) |
| Avg Score Outside Discipline | 0.00808 | 0.00585 | -0.000533 |
| | (0.0100) | (0.0124) | (0.0176) |
| Male | 0.113*** | 0.105*** | 0.129*** |
| | (0.0262) | (0.0318) | (0.0437) |
| First Year | -0.244*** | | |
| | (0.0384) | | |
| Second Year | -0.170*** | -0.168*** | |
| | (0.0386) | (0.0385) | |
| Third Year | -0.166*** | -0.164*** | -0.160*** |
| | (0.0402) | (0.0401) | (0.0405) |
| PhD in Canada | -0.131*** | -0.165*** | -0.194*** |
| | (0.0354) | (0.0464) | (0.0655) |
| Constant | 0.416*** | 0.404** | 0.433** |
| | (0.129) | (0.163) | (0.188) |
| $N$ | 1425 | 1020 | 570 |
| $R^2$ | 0.137 | 0.148 | 0.186 |

(Robust standard errors)

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

**Note**: Model 1 contains the whole sample. Model 2 includes only recipients above first year. Models 3 is restricted to recipients entering third or fourth year. All models contain 27 discipline dummies.

Table 8: Linear Probability Model Using the the Scores Given by Evaluators within/outside the Candidate's Discipline to Explain Completion of Program in Five or Seven Years

| | Completion in Five Years | | | Completion in Seven Years | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | All | Excluding 1st Year | 3rd and 4th Year | All | Excluding 1st Year | 3rd and 4th Year |
| Avg Score In Discipline | 0.0125 | 0.00814 | 0.0177 | 0.00869 | 0.0133 | 0.0165 |
| | (0.00944) | (0.0116) | (0.0162) | (0.00939) | (0.0111) | (0.0150) |
| Avg Score Outside Discipline | 0.0181$^{*}$ | 0.0223$^{*}$ | 0.00850 | 0.0262$^{**}$ | 0.0263$^{**}$ | 0.0167 |
| | (0.0109) | (0.0131) | (0.0183) | (0.0109) | (0.0127) | (0.0162) |
| Male | 0.0817$^{***}$ | 0.0860$^{***}$ | 0.0904$^{**}$ | 0.0600$^{**}$ | 0.0580$^{*}$ | 0.0713$^{*}$ |
| | (0.0270) | (0.0324) | (0.0429) | (0.0265) | (0.0312) | (0.0412) |
| First Year | 0.0237 | | | -0.147$^{***}$ | | |
| | (0.0400) | | | (0.0380) | | |
| Second Year | 0.0897$^{**}$ | 0.0906$^{**}$ | | -0.112$^{***}$ | -0.115$^{***}$ | |
| | (0.0398) | (0.0402) | | (0.0369) | (0.0370) | |
| Third Year | 0.0323 | 0.0362 | 0.0269 | -0.114$^{***}$ | -0.115$^{***}$ | -0.115$^{***}$ |
| | (0.0412) | (0.0417) | (0.0423) | (0.0390) | (0.0391) | (0.0396) |
| PhD in Canada | -0.0294 | -0.00268 | 0.0509 | -0.0606$^{*}$ | -0.0376 | -0.0650 |
| | (0.0360) | (0.0447) | (0.0617) | (0.0351) | (0.0426) | (0.0555) |
| $N$ | 1425 | 1020 | 570 | 1425 | 1020 | 570 |
| $R^2$ | 0.050 | 0.050 | 0.047 | 0.050 | 0.043 | 0.050 |

(Robust standard errors)

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

**Note**: All standard errors are robust. Models 1 to 3 explain whether a student completed the doctoral program in 5 years, and models 4 to 6 explain completion in seven years. Models 1 and 4 contain the whole sample. Models 2 and 5 include only recipients above first year. Models 3 and 6 include only recipients above second year. All models contain 27 discipline dummies.