Career vs. Social Impact: Gendered Responses in Teacher Recruitment in Ghana*

Urša Krenk† Kobbina Awuah‡

September 29, 2026

Abstract

Organizations recruiting for jobs with a social mission must decide what to emphasize when advertising them. In a field experiment with a Ghanaian NGO, we randomly varied whether emails recruiting university graduates into a teaching fellowship emphasized its career benefits or its social impact. The career message increased male application rates by 71% relative to the social-impact message and left female rates unchanged; we reject that men and women responded equally. Suggestive evidence indicates that candidates recruited through the career message were more likely to receive offers and be hired, and that men, but not women, were less prosocial.

Keywords: recruitment, gender, field experiment, education, development.
JEL Classification: C93, D64, J16, J24, M51, I25, O15.

*This study is registered in the AEA RCT Registry, AEARCTR-0011501.
†University of Zurich. Corresponding author. Email: ursa.krenk@econ.uzh.ch
‡This research was conducted while at the University of Zurich.

1 Introduction

Organizations must decide which aspects of a job to emphasize when they advertise it. Recruitment strategies that highlight specific job attributes, such as career advancement or the social impact and purpose of work, can affect both the size and composition of the applicant pool. Understanding these effects is particularly important for organizations that try to balance multiple objectives, such as hiring workers of high ability, attracting prosocially motivated employees, and maintaining a gender-balanced workforce. For example, one concern is that emphasizing the career-related aspects of a job (including higher earnings) may attract more skilled candidates, who tend to have better outside options, but at the same time crowd out candidates who are mission-oriented (Ashraf et al., 2020; Dal Bó et al., 2013; Deserranno, 2019). Moreover, these trade-offs may differ between men and women. Men and women might differ in their preferences for certain job characteristics, such as job meaning or how competitive the job is, and in their internalized gender norms (Burbano et al., 2024; Del Carpio and Guadalupe, 2022; Delfino, 2024; Flory et al., 2015; Mas and Pallais, 2017; Oh, 2023; Palffy et al., 2023; Wiswall and Zafar, 2018). They might also face different outside options. Consequently, recruitment strategies emphasizing certain aspects of a job can result in different selection patterns, and thus different trade-offs, for men and women.

In this paper, we study these questions through a randomized field experiment conducted with a Ghanaian NGO that operates in the education sector. The NGO recruits recent university graduates as teaching fellows, providing them with training before placing them into rural schools for a two-year fellowship. We randomly vary the content of recruitment messages sent to 1,667 candidates who had registered an interest in the fellowship but had not yet applied, highlighting either the career advancement opportunities associated with the fellowship (the *career* treatment) or the fellowship's social impact and purpose (the *purpose* treatment). This design allows us to measure how increasing the salience of these distinct job aspects affects the quantity, quality, and prosociality of the applicant pool, whether there are trade-offs between these characteristics, and whether these trade-offs vary by gender. The experiment ran inside the NGO's regular annual recruitment cycle, was double-blind, and barely changed the organization's usual operations. Our empirical context is well suited to the question we study: teaching fellows work with weak performance incentives and in difficult conditions, which are common in low-resource settings, so both skill and prosocial motivation matter for how well they do the job.

We find that emphasizing career benefits changes who applies for the fellowship. Men assigned to the *career* treatment are approximately 71% more likely to apply than men assigned to the *purpose* treatment (p-value = 0.007), while for women the difference in application rates between the two treatments is small and not statistically significant (women in the *career* treatment are about 18% less likely to apply, a difference of 1.3 percentage points, p-value = 0.479). We can reject the null that men and women respond equally to the two messages (p-value = 0.015). This is what one would expect if men place more weight on the career component of the job, an idea we develop in Section 3.3. We next investigate the effect of the *career* treatment on applicant quality, as perceived by the organization. Candidates who received the career message are roughly 69% more likely to receive an offer (p-value = 0.090), and the hiring rate rises by 90% (p-value = 0.092), though both estimates are imprecise. Offers are determined by an entire day of in-person assessments run by evaluators who are blind to treatment, and the organization uses these scores for its final hiring and placement decisions. Because the organization applies a fixed quality bar and leaves positions unfilled when candidates do not meet it, these additional offers correspond to additional candidates clearing that bar, and not simply to a larger applicant pool.

We then examine the social preferences of the candidates each message attracts. Because we have this measure available only for a small number of respondents who reached the in-person assessment and responded to our survey, we treat these results as suggestive. Men in the *career* treatment score 0.83 standard deviations lower on our prosociality index (p-value = 0.012) compared to men in the *purpose* treatment, while women's prosociality is unchanged (p-value = 0.625); the gender difference is significant (p-value = 0.047). Taken together, the career message improves applicant quality, with no detectable difference between men and women, while the increase in the number of applicants and the decline in prosociality are both concentrated among men.

Our study contributes to several strands of literature. First, our findings complement the growing literature in organizational behavior and personnel economics examining how subtle changes in job advertisements or recruitment strategies can influence the size and characteristics of the applicant pool (Abebe et al., 2021; Abraham et al., 2024; Card et al., 2024; Coffman et al., 2024; Del Carpio and Guadalupe, 2022; Delfino, 2024; Flory et al., 2015, 2021; Kuhn and Shen, 2023; Marinescu and Wolthoff, 2020). For example, Del Carpio and Guadalupe (2022) show that a message countering gender stereotypes doubled applications to a coding training program in Peru, lowering average skills in the pool while raising the skills of the top applicants. Our paper belongs to a relatively small group of studies investigating whether advertising career benefits or higher earnings attracts more skilled applicants, and whether this leads to trade-offs between attracting skilled and prosocially motivated candidates (Ashraf et al., 2020; Dal Bó et al., 2013; Deserranno, 2019). These studies have produced mixed results. Ashraf et al. (2020) study the recruitment of community health assistants in rural Zambia, and show that advertising career benefits does not increase the number of applicants, but does attract individuals with better skills. They find trade-offs between skill and prosocial motivation primarily among applicants with lower levels of talent, but not among candidates who ultimately get hired. Similarly, Dal Bó et al. (2013), who experimentally vary advertised wages for Mexican civil servants, find that higher wages attract more candidates (but not statistically significant) and that those candidates have better skills, but do not find any trade-offs between skill and prosocial motivation. In contrast, Deserranno (2019) varies the earnings that applicants expect from a community health promoter position in Uganda, and finds that higher expected earnings attract more applicants, but that these applicants are less prosocially motivated. These mixed findings suggest that the presence and magnitude of such trade-offs are likely job- and context-specific.

Our candidates are highly skilled graduates from Ghana's elite universities, applying for a relatively prestigious teaching fellowship. They differ from the applicants in the two studies above: Ashraf et al. (2020) and Deserranno (2019) recruit from populations without university education in rural Zambia and Uganda. We also study the teaching profession, which at secondary levels and beyond is male-dominated in Ghana and in many other Sub-Saharan African countries.[1] Consequently, men and women interested in such positions might come from different parts of the skill distribution and face distinctly different outside employment opportunities. Indeed, we show that the effects of the career message vary by gender: application rates increase only among men, and the decline in prosocial motivation is confined to men, while the improvement in applicant quality is not detectably different for men and women. We are cautious about reading this as a direct extension of Deserranno (2019), since her applicants differ from ours in both education and outside options, and her treatment varies expected earnings rather than career emphasis. A similar gendered response appears in a very different setting: in a field experiment with a large technology firm, Fuchs et al. (2024) find that advertisements for entry-level positions that emphasized career advancement attracted more male applicants with stronger qualifications, while advertisements emphasizing flexibility increased applications from both men and women. The gendered response to

[1]In Ghana, women were 50% of primary school teachers but only 31% of secondary teachers in 2024 and 27% of tertiary teaching staff in 2023, compared with a Sub-Saharan African secondary average of 35%. Across OECD countries, by comparison, women are 81% of primary and 60% of secondary teachers (2024) (UNESCO Institute for Statistics, 2026).

career framing is therefore not specific to teaching or to a low-income context. The same gap appears within teaching: in a nationwide experiment in Peru, both altruistic and monetary appeals moved men more than women (Ajzenman et al., 2024). We also measure the prosociality of the applicants, which allows us to ask whether the additional male applicants come at a cost in prosocial motivation.

Our study also contributes to the broader literature examining strategies for improving teaching quality and attracting higher-quality teachers in low-income countries. Education systems there face persistent challenges such as high teacher absenteeism, low teacher motivation, and poor instructional quality. Learning outcomes remain very low as a result: in Sub-Saharan Africa, 89% of 10-year-olds cannot read and understand a simple text (World Bank, 2022; Akseer and Játiva, 2021). While previous research has largely focused on how performance-based incentives can directly improve the effort and attendance of existing teachers (Duflo et al., 2012; Lavy, 2009; Muralidharan and Sundararaman, 2011), comparatively less evidence exists on how recruitment strategies influence teacher quality by shaping who enters the profession. Within this smaller literature, Brown and Andrabi (2023) shows that offering performance incentives can attract more skilled and motivated people into teaching. In a nationwide experiment in Peru, Ajzenman et al. (2024) show that making either an altruistic identity or an existing monetary incentive more salient increased the share of disadvantaged schools that teachers applied to, and that both effects were larger for men than for women. Their treatments change where existing teachers apply, while ours change who enters the applicant pool in the first place. We show that emphasizing career benefits in recruitment messages can attract candidates whom the organization assesses as more skilled. We do not observe teacher value-added, so we cannot say whether these candidates are more effective in the classroom. The assessment does, however, determine the organization's own hiring and placement decisions, which makes it a meaningful measure of the quality the organization is selecting on.

The lessons here are not specific to teaching. Any organization that recruits with more than one objective in mind faces the possibility that a single message moves those objectives in different directions. Our results suggest that it can also move them differently for men and women: here, the same message that raised applicant quality also made the pool more male and, suggestively, less prosocial.

The rest of the paper is organized as follows. Section 2 describes the setting, and Section 3 the experimental design and our conceptual framework. Section 4 presents the empirical strategy, the data, and balance. Section 5 reports the results, Section 6 discusses mechanisms, and Section 7 concludes.

2 The Research Setting

We collaborate with an NGO operating in the education sector in Ghana. The NGO recruits recent university graduates, trains them as teachers and places them into rural communities across the country for a two-year teaching fellowship. To be eligible to apply for the position, candidates must have completed at least a Bachelor's degree in any field. Prior teaching experience is not required, and the organization provides extensive pre-placement training and on-the-job support throughout the two-year fellowship. Candidates can apply for the fellowship as part of the "National Service"[2], or as a regular job, if they have already completed their National Service.[3] Each year, a cohort of between 50 and 150 fellows is placed into schools located in selected regions among Ghana's 16 administrative regions. Fellows receive a stipend comparable to the average salary for recent university graduates in Ghana, which is higher than what they would typically earn completing their National Service in the public sector. The job is considered prestigious, and the selection process is competitive, with historically only between 15 and 20% of applicants being offered positions.[4] The organization sets a target number of fellows each year and leaves slots unfilled if too few candidates meet its bar. After completing the fellowship program, approximately 60% of fellows continue working in the education sector, either as teachers or with other education-focused non-profits. Many of those who leave education join other non-profit organizations or enter the public sector.

This setting is particularly well-suited to study our research question for two main reasons. First, the NGO's recruitment process closely resembles typical hiring practices used by many firms and organizations. Employers regularly choose how to advertise their roles, often tailoring their messages through mailing lists or targeted outreach, similar to the experimental variation used here. Second, this particular position, a teaching fellowship, is especially suitable for examining recruitment trade-offs because it combines substantial potential for social impact and meaningfulness with significant career benefits for the fellows. This al-

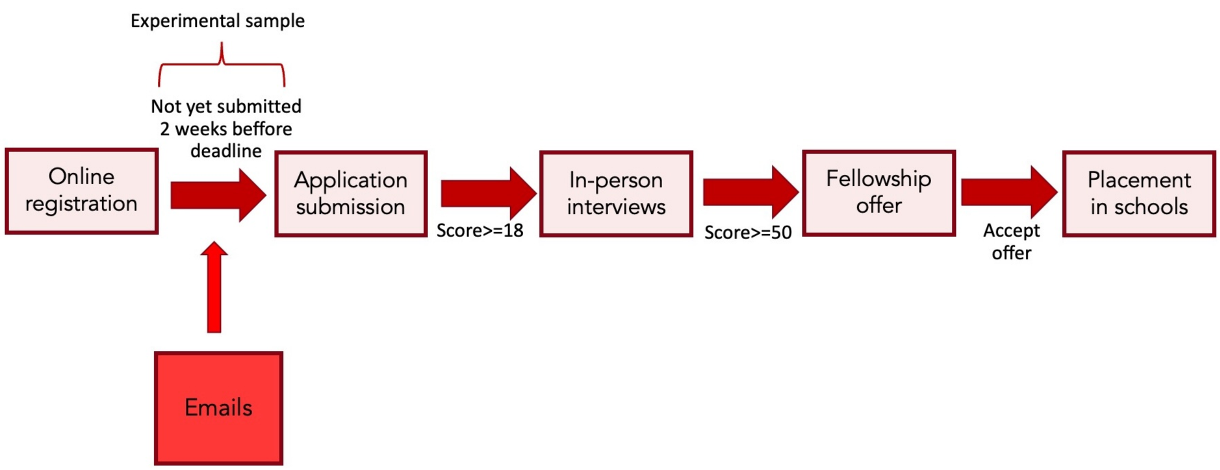
[2] All graduates from accredited tertiary institutions in Ghana are required to complete a "National Service", which involves a one-year placement typically in the public sector, although placements in the private sector or with NGOs are also possible.

[3] In our experimental sample, approximately 44% of the candidates registered for the fellowship as part of their National Service.

[4] Typically, around one half of the applicants are invited to the assessment day, and between one half and three quarters of those who attend are given offers.

lows us to study how emphasizing different aspects of the job influences candidate selection, specifically regarding the quantity, quality, and prosociality of the applicants attracted, and whether these responses differ by gender.

Figure 1: Recruitment Process



Notes: The figure shows the recruitment process of the partner NGO. Applications were open from April to June. After candidates registered online, they were randomly assigned to different recruitment emails, but these treatment emails were only sent if candidates had not yet submitted their application two weeks before the deadline. For successful candidates in our sample, the time between application and receiving a job offer was approximately three months. Those who accepted the job offer began their community placements between October of the same year and January of the following year.

Figure 1 illustrates the organization's recruitment process, which consists of several stages. First, interested applicants register online to express their interest. After registering, applicants are added to the organization's mailing list, through which they are informed when the application portal opens and may receive recruitment emails. The online application requires candidates to answer six short essay questions, each scored from 1 to 5 points. These essays assess the applicant's leadership ability, grit, and fit for the program.[5] Our experiment was conducted during this online application stage, after candidates had registered and expressed their interest. Candidates who score at least 18 points on their essays are invited to the next stage, a full day of in-person assessments. The assessment day is designed as a fresh start, in that the essay grades carry no further weight and the

[5]Questions 1 to 4 assess the candidate's motivation, educational philosophy, and alignment with the organization's mission; question 5 is a proxy for grit; and question 6 measures the ability to lead and influence others. Grading is blind: evaluators, who are current employees or program alumni, see only characteristics relevant to eligibility, such as education level, National Service status and graduation year.

evaluators differ from those who graded the applications and do not observe the online application scores. This assessment day evaluates candidates through multiple tasks, including an individual interview, a group exercise, a mock-teaching session, and a problem-solving activity. Each task is scored out of 100 points. Candidates who achieve an average score of at least 50 points across these tasks receive a fellowship offer. Assessment performance is the organization's key metric for evaluating candidates: beyond determining offers, it guides the placement of fellows across regions, with the strongest performers assigned where the need is greatest. Assessment days are centrally organized in bulk, often scheduled at short notice and on fixed, non-flexible dates, so a substantial share of invited candidates are unable to attend. Recruitment is cyclical and typically occurs between March and July. Successful candidates who accept the fellowship start their community placements between October of the same year and January of the following year.

3 Experimental Design

3.1 Sample Selection

Our experimental pool consists of a subset of individuals who expressed interest in the fellowship by creating a profile on the organization's webpage. Specifically, it includes eligible[6] candidates who had not submitted their application two weeks before the application deadline. The organization typically sends "reminder-to-apply" emails around this time, which is why our experimental sample includes only candidates who had not yet applied by that point. Thus, our sample is selected based on candidates having at least a minimum level of interest in the position and having not submitted their application by two weeks prior to the deadline. We provide a more detailed description of our experimental sample in Section 4.4.[7]

3.2 Email Randomization

In total, there were four rounds of "reminder-to-apply" emails sent as part of our experiment (twice per week during the last two weeks before the application deadline). We randomized

[6]Eligibility is defined as having completed at least a Bachelor's degree by the start date of the fellowship. We discuss this in more detail in Section 4.4.

[7]Compared to the full pool of registered candidates, our sample over-represents "late" submitters. We discuss potential differences between "late" and "early" submitters in Section 5.4.

the content of these emails by including statements emphasizing one of two distinct aspects of the fellowship: i) the impact of the fellowship on children and the community, or ii) the fellowship's benefits for the career progression of the fellows. We refer to these two treatment conditions as the *purpose* and *career* treatments, respectively. Depending on when (and whether) applicants eventually submitted their application, they received between one and four reminder emails, all corresponding to the same experimental condition.

These two experimental conditions were cross-randomized with either a male or female photograph, resulting in four distinct types of emails sent in each round. Randomization was conducted at the individual level in four separate rounds, stratified by gender. Most participants were randomized in the first round. Around 2% of participants were randomized in rounds 2 through 4 because they had registered on the webpage only after the initial randomization (i.e., within the last two weeks prior to the application deadline).[8] The experiment was conducted in a double-blind manner: participants were unaware that the reminder emails were part of a study, and the recruitment team did not know the applicants' treatment assignments. This design helps minimize experimenter demand effects and prevents any influence on the recruitment team's assessments due to the applicants' assigned treatment.

An example of a reminder email is shown in Figure 2. The photographs used in the experiment were of actual fellows from previous cohorts. The accompanying texts were general statements that the fellows agreed could be posted alongside their photographs. In total, eight distinct photographs were used (four women and four men), combined with eight different statements. As an example, the two statements included in the fourth reminder email are shown below; the purpose version is the one visible in Figure 2. Appendix Table A.8 lists all eight statements in the order in which the reminders were sent.

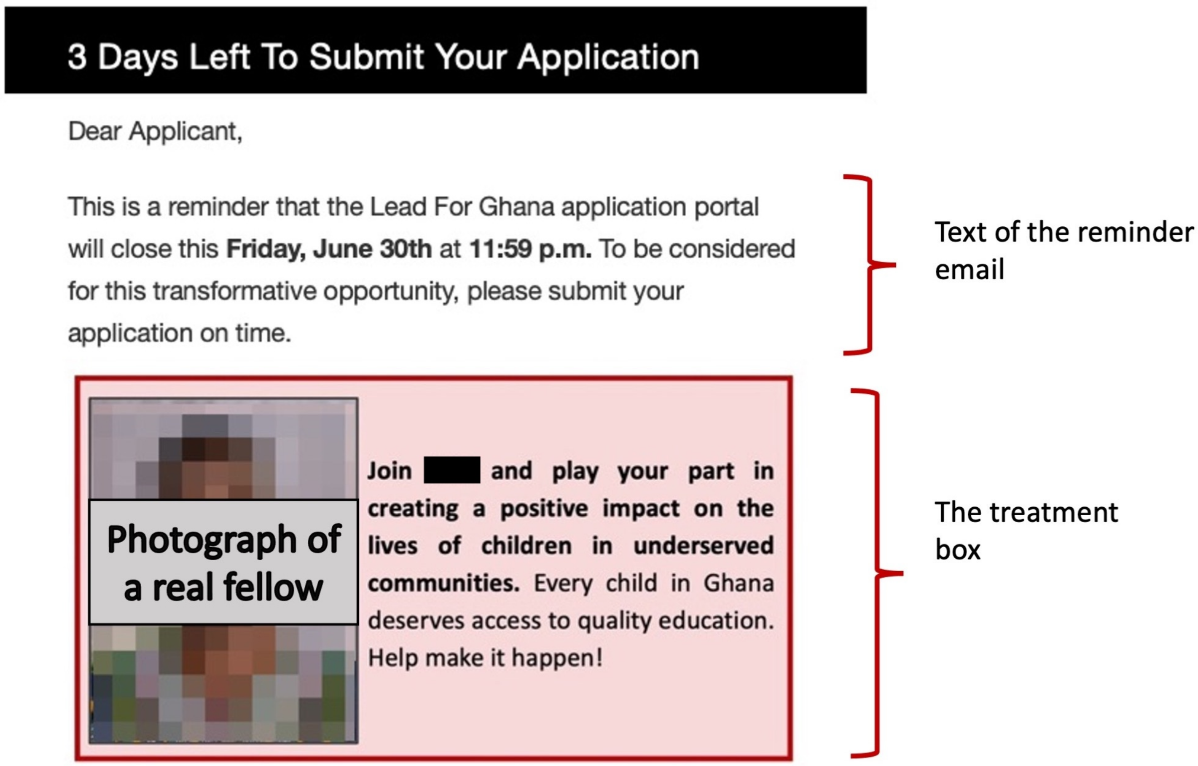
Career treatment

> Join [name of the NGO] and build a solid foundation for your professional journey. You will get to explore your inner potential and gain transferable skills that will make you stand out professionally. Seize the opportunity!

Purpose treatment

[8] A further 56 candidates were recruited through a parallel Facebook survey, in line with the preregistration plan. They were randomly shown one of the same four statement-and-photograph combinations in the survey, were kept in that group for the reminder emails, and form a fifth randomization round. Because they registered only after seeing a treatment message inside the survey, their presence in the sample may itself depend on the treatment they were shown; we therefore exclude them from the main analysis. Table A.7 reports all main outcomes including them.

Figure 2: Example Email



Notes: The figure shows an example of one of the emails used in the intervention. The treatment text was placed in a separate box below the main body of the email and was visible in the preview when viewed on a smartphone. The email footer contained the organization's logo.

> Join [name of the NGO] and play your part in creating a positive impact on the lives of children in underserved communities. Every child in Ghana deserves access to quality education. Help make it happen!

This design allows us to identify the causal effect of the *career* versus *purpose* treatments. Applicants received a different statement combined with a different photograph in each reminder round. This approach helps ensure that our findings are not driven by any single specific text or photograph. It also had practical advantages, as the organization believed that repeatedly sending reminders with the same content and photograph would seem unusual to applicants. Using the organization's email tracking system, we observe which candidates actually opened the reminder emails, enabling us to estimate both intent-to-treat (ITT) effects and local average treatment effects (LATE). We report these results in Section 5.

The aim of the two treatments was to vary the emphasis placed on two distinct aspects of the fellowship: i) the social aspect, namely making a positive difference in children's lives, and ii) the career aspect, emphasizing opportunities to boost candidates' career development. Although the fellowship was new to most applicants (few applicants reapply), we do not interpret our intervention as providing genuinely new information. Rather, it serves to highlight two important but distinct motivations associated with the fellowship.

3.3 Conceptual Framework

It is useful to think about our experimental design in the light of the following conceptual framework. A candidate applies to the fellowship if the value she attaches to the fellowship exceeds the value of her outside option. The fellowship's value has two components, its career benefits and its social purpose. Candidates differ in how much weight they place on each component, in their beliefs about how much of each the fellowship offers, and in the strength of their outside options. As described above, each message raises the salience of one of the two components at the moment the candidate decides, and with it the perceived value of that component.[9] Some candidates therefore apply under one message but not the other.

A message will matter most to those candidates who place more weight on the component it emphasizes. There is some evidence that men place more weight on career advancement than women (Wiswall and Zafar, 2018), a difference that may itself reflect gendered norms about which job attributes are appropriate to pursue (Burbano et al., 2024; Del Carpio and Guadalupe, 2022; Palffy et al., 2023). If beliefs and outside options are otherwise comparable across genders (an assumption we make), the career message should move more men than women across the application threshold.[10] We summarize this in Hypothesis 1, below.

Hypothesis 1. *The career treatment increases applications more among men than among women.*

[9] The organization reports that candidates often view the fellowship as a volunteering position, and highlights career benefits in its own materials for that reason. We therefore cannot fully separate salience from information, since the career message may also have corrected a misperception; we do not observe candidates' beliefs directly. Related work in settings where the position is new to applicants models such messages as information or signals (Ashraf et al., 2020; Deserranno, 2019; Fuchs et al., 2024). Either channel operates through the randomly assigned message, so our estimates are unaffected.

[10] Men and women can of course also differ in their beliefs about how much career or social benefit the fellowship offers, and in their outside options. We do not measure beliefs, so we cannot rule out belief differences. We are also not aware of good evidence on outside options differing between young educated men and women. But if men do have better outside options, the framework makes the same prediction: for them, the career return is what makes applying worthwhile, the logic of Hypothesis 2 below.

Because men are the majority of the applicant pool, the response among men should also raise the overall application rate, though by less than it does among men. This leads to Hypothesis 1b.

Hypothesis 1b. *The career treatment weakly increases the overall application rate.*

Candidates with strong outside options forgo more by joining, so the fellowship's social purpose alone will often not be enough to make them apply; its career return is what can compensate them. The marginal applicants the career message draws in should therefore have stronger outside options, and to the extent that the labor market and the NGO reward similar skills, they are well placed to clear the NGO's fixed hiring bar. We summarize this in Hypothesis 2, below.

Hypothesis 2. *The career treatment increases offer and hiring rates.*

The framework does not sign the effect on prosociality, and the two closest studies illustrate why it can go either way. The marginal candidates, those who apply under the career message but not the purpose message, may place little weight on the mission, which would lower the pool's average prosociality, the pattern in Deserranno (2019), where emphasizing earnings deterred the most prosocial applicants. But candidates with strong outside options who consider a teaching fellowship at all plausibly care about the mission, in which case prosociality need not fall, the pattern in Ashraf et al. (2020), where career incentives attracted talent without a trade-off among those hired. The direction is therefore an empirical question.

We test Hypothesis 1 with the gender interaction in equation 2, Hypothesis 1b with the pooled specification in equation 1, and Hypothesis 2 with the unconditional offer and hiring outcomes.

4 Empirical Strategy, Data, and Balance

4.1 Empirical Strategy

Our empirical strategy relies on the random assignment of statement types to compare the impact of these statements on entry and selection into the teaching fellowship. We estimate the following regression:

$$y_i = \alpha + \beta_1 career_i + X_i\lambda + \epsilon_i, \tag{1}$$

where $career_i$ equals one if candidate i was assigned a statement emphasizing the career aspect of the fellowship, and zero otherwise. X_i is a vector of control variables that includes strata (randomization round) fixed effects, whether the photograph in the email matched the candidate's own gender, whether the candidate had completed National Service, whether they were in their final year of study, their university, whether they previously applied for the fellowship, region of origin, GPA, the month of registration, and how they initially learned about the NGO. Since the randomization was conducted at the individual level, we use heteroscedasticity-robust standard errors. The coefficient β_1 identifies the causal effect of receiving a career-focused statement on outcome y_i.

To test whether men and women respond differently, we also estimate a specification where we interact our treatment indicator with a female dummy:

$$y_i = \alpha + \beta_1 career_i + \beta_2 career_i \times female_i + \beta_3 female_i + X_i\lambda + \epsilon_i. \tag{2}$$

In this specification, β_1 is the treatment effect for men, $\beta_1 + \beta_2$ is the effect for women, and β_2 is the difference between the two.

Not all candidates opened the reminder emails. Using the organization's email tracking system, we observe whether a candidate opened at least one reminder email, which 63% of candidates did. This enables us to estimate both intent-to-treat (ITT) effects and local average treatment effects (LATE). For the LATE, we instrument actual exposure to the treatment, defined as being assigned the career statement and opening at least one reminder email, with the treatment assignment itself. In the interacted specification, we instrument the exposure term and its interaction with the female dummy analogously. Because we are uncertain how reliable the organization's email tracking system is, our main results are the ITT estimates.

4.2 Outcome Variables

We consider three sets of outcomes y_i: application outcomes (submission rates), downstream outcomes (offer and hiring rates), and applicants' prosociality.

Application Outcomes We define a binary variable equal to one if the candidate submitted the online application, and zero otherwise.

Offer and Hiring Outcomes We use a binary indicator for whether a candidate received a job offer, and a binary indicator for whether they accepted it and were hired. We present unconditional outcomes, coded as zero if the candidate did not reach the relevant stage. We interpret these as quality outcomes, since offers are determined by performance at the in-person assessment, which starts from scratch and which the organization treats as its key measure of candidate quality. In the Appendix, we also report outcomes conditional on reaching a given stage, coded as missing otherwise.

Prosociality To measure applicants' prosociality, we administered a survey to candidates attending the in-person assessment. The survey occurred roughly two months after the application submission, with a response rate of approximately 85%. The prosociality measure is an index constructed using four different questions that are meant to capture candidates' prosocial attitudes. Two questions were hypothetical and taken from Falk et al. (2018): 1) How willing are you to give to good causes without expecting anything in return (scale 0-10); 2) Imagine the following hypothetical situation: Today you unexpectedly received 150 GHS. How much of this amount would you donate to a good cause? The third question captures real donation behavior: at the end of the survey, we asked participants how much of their survey incentive (30 GHS, equivalent at the time to approximately 2.5 USD) they would like to donate to a Ghanaian charity assisting families with school fees. The fourth question asked candidates to rank five job attributes by importance; we use how highly they ranked "making a positive difference in the community". We created the prosociality index by standardizing each of the four responses using the mean and standard deviation of men in the *purpose* treatment group and averaging them (Kling et al., 2007). Because the survey sample is limited to assessment attendees and treatment itself affects who reaches that stage, we report bounds following Lee (2009) alongside these estimates.

4.3 Identification Assumptions

The probability of a candidate progressing from one hiring stage to another should be independent of the treatment assignment. This assumption could be violated if the recruitment team's screening criteria were influenced by candidates' assigned treatment, or if the treatment itself affected applicants' effort levels. However, we believe this assumption is reasonable, given that the double-blind design should prevent recruiters from knowing or being influenced by candidates' treatment assignments. We discuss both of these potential threats to identification in Section 6.

4.4 Balance

The sample consists of 1,667 candidates: 949 (57%) men and 718 (43%) women.[11] Demographic information is missing for a small number of candidates who registered on the webpage but did not fully complete their profiles. Appendix Table A.1 provides summary statistics by gender and balance checks for the entire sample. Among candidates, 60% of men and 50% of women had already completed their National Service, while 36% of men and 46% of women were in their final university year. Most applicants identify as ethnically Akan (speaking Twi as their mother tongue; 61% of men and 65% of women), and most applicants (87% of men and 86% of women) graduated from one of the five largest universities: KNUST, University of Ghana, University for Development Studies, University of Cape Coast, and University of Education, Winneba. Treatment assignment is largely balanced on observable characteristics, as demonstrated in columns 7 and 8 of Appendix Table A.1, which report F-statistics and p-values from regressions of each variable on treatment assignment, controlling for strata fixed effects. Additionally, we fail to reject the null hypothesis of joint orthogonality across all variables reported in the table ($F(9,1653)=1.13$, p-value $= 0.339$).

5 Results

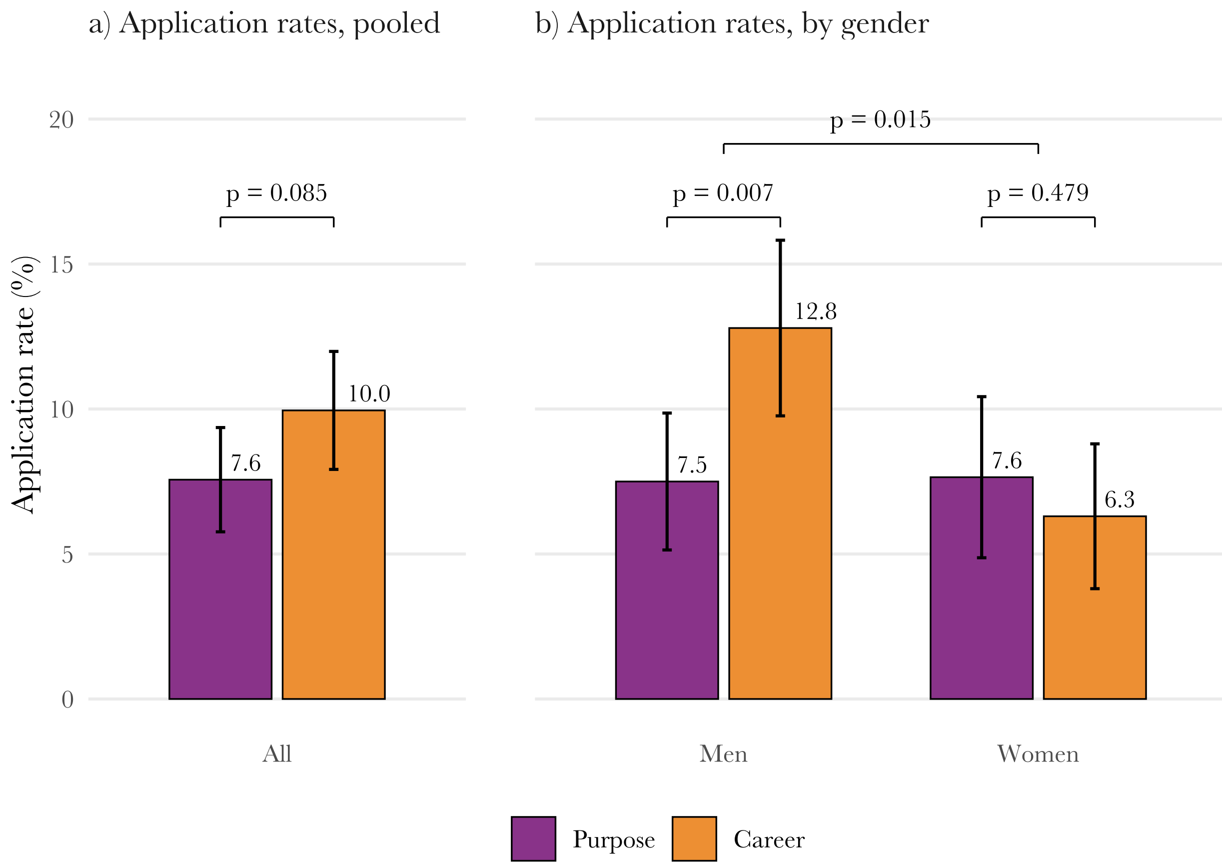
This section presents the main results of our study. We first report the results on application rates, followed by the results on applicant quality and social preferences.

5.1 Application Rates

Figure 3 presents application rates, pooled (panel a) and separately for men and women (panel b). Appendix Table A.2 reports the corresponding regressions: columns (1) to (3) estimate the pooled effect and columns (4) to (6) add the gender interaction, in each case without controls, with the randomization strata variables only, and with the full control vector. The table also reports LATE estimates, which instrument exposure to the treatment, defined as being assigned the career statement and opening at least one reminder email, with the treatment assignment itself. Pooled across genders (panel a), the *career* treatment

[11] 302 individuals included in the randomization were not eligible to apply, and their applications were not graded. This primarily occurred because these candidates held a Higher Education Diploma (298 cases) rather than a Bachelor's degree. As specified in our pre-analysis plan, we exclude these individuals from the analysis. Additionally, 308 people, about two thirds of all applications, submitted their application before our experiment took place.

Figure 3: Application Rates



Notes: The outcome is a binary indicator of whether a candidate submitted an application, expressed in percent. Panel a plots treatment-arm means estimated from equation 1, and panel b from equation 2, both without control variables. Purple bars are means under the *purpose* treatment and orange bars under the *career* treatment. Error bars are 95% confidence intervals, computed from heteroscedasticity-robust standard errors of the relevant linear combination of coefficients. The bracket above each pair of bars reports the p-value of the treatment effect for that group, and the bracket spanning men and women reports the p-value on the interaction term β_2 in equation 2. Appendix Table A.2 reports the underlying regressions.

raises the application rate from 7.6% to 10.0%, an increase of 2.4 percentage points. The pooled effect is only marginally significant and noisily estimated (p-value = 0.085), and it is not significant once controls are added. This pooled figure masks two very different responses. Panel b shows that among men the application rate rises from 7.5% under the *purpose* treatment to 12.8% under the *career* treatment, a 5.3 percentage point increase and a relative increase of 71% (p-value = 0.007). Among women it moves in the opposite direction, from 7.6% to 6.3%, a decline of 1.3 percentage points that is statistically insignificant (p-value = 0.479). The two treatment effects differ from one another (interaction p-value =

0.015), so men and women respond differently to the same recruitment message.[12] This is the pattern predicted by Hypothesis 1 (Section 3.3). Consistent with Hypothesis 1b, the pooled effect is roughly the male response diluted by the unresponsive female share. Notably, the gender difference comes from men responding to the career message rather than from women responding to the purpose message, even though women on average seem to report stronger preferences for meaning at work (Burbano et al., 2024). Our results also contrast with Ashraf et al. (2020), who find no effect of emphasizing career versus community benefits on the size of the applicant pool.

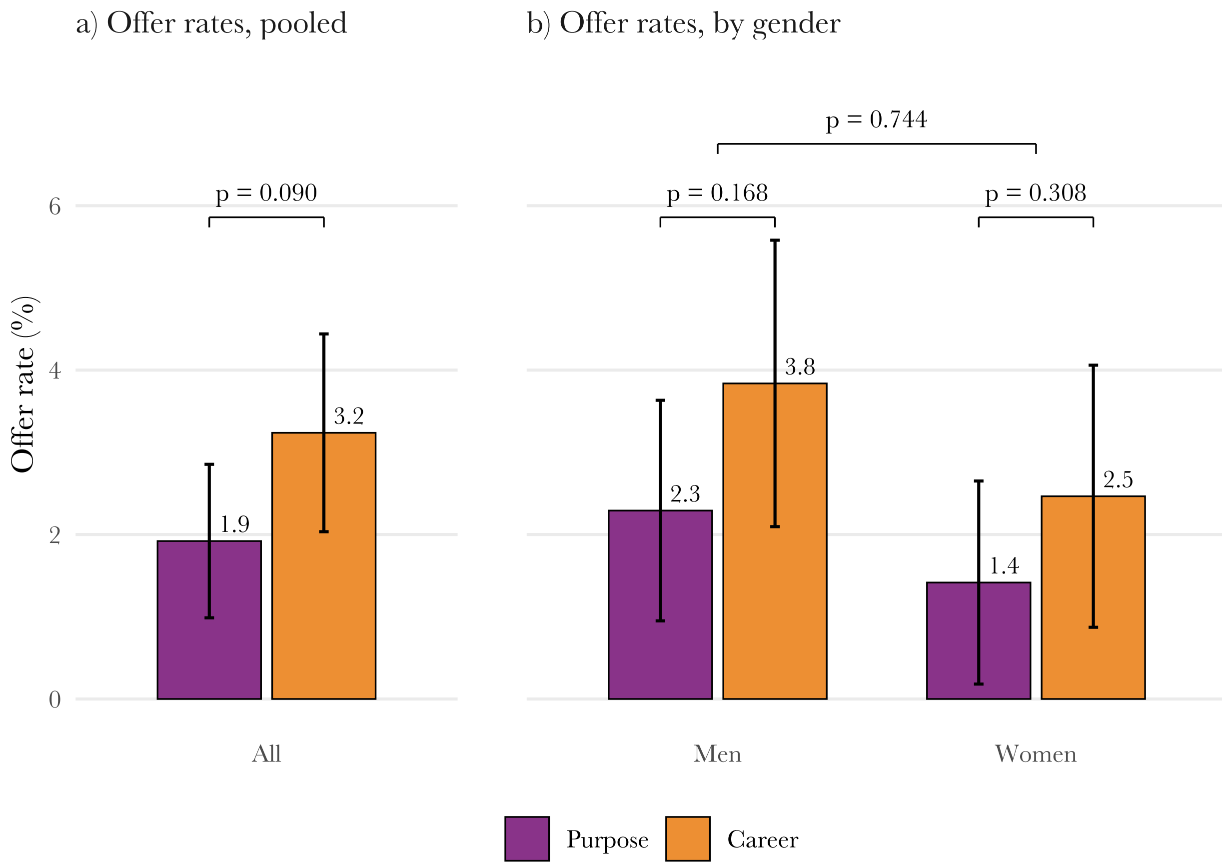
5.2 Offer and Hiring Outcomes

We next examine whether the two messages differ in the quality of the candidates they attract, proxied by offer rates. We use unconditional outcomes, coded zero for candidates who did not apply or did not reach the in-person assessment stage, because these are the outcomes the randomization actually assigns; outcomes conditional on reaching a later stage are reported in the appendix. We treat these outcomes as suggestive evidence: offer and hire events are rare in our sample (43 offers and 29 hires among 1,667 candidates), so we have limited power to detect effects. Panel a of Figure 4 shows that the offer rate rises from 1.9% under the *purpose* treatment to 3.2% under the *career* treatment, an increase of 69% (p-value = 0.090). Panel b shows the same direction for men (1.5pp, p-value = 0.168) and women (1.1pp, p-value = 0.308), with no detectable gender difference (interaction p-value = 0.744); both gender-specific estimates are imprecise given the small number of offers. Appendix Figure B.1 shows the same pattern for hiring: the hiring rate rises from 1.2% to 2.3% (p-value = 0.092), again with no detectable gender difference (interaction p-value = 0.424). Both sets of estimates are reported in Appendix Table A.4.

An increase in unconditional offer rates could in principle reflect nothing more than a larger applicant pool. In our setting this interpretation is limited: the offer threshold is absolute (an average assessment score of at least 50 points), candidates from both treatment arms were evaluated at the same assessment centers by evaluators blind to treatment, and the organization leaves slots unfilled when candidates do not meet the bar. The additional offers in the *career* arm therefore correspond to additional candidates clearing a fixed quality threshold: had the marginal applicants attracted by the career message been weak, they would not have received offers. Consistent with this, conditional on attending the assessment

[12] Appendix Figure B.2 splits application rates by the gender of the photograph included in the email. The treatment effects look similar across male and female photographs.

Figure 4: Job Offer Rates



Notes: The outcome is a binary indicator of whether a candidate received a fellowship offer, expressed in percent. It is unconditional, coded zero for candidates who did not apply or did not reach the in-person assessment, so the sample is the full set of randomized candidates. Panel a plots treatment-arm means estimated from equation 1, and panel b from equation 2, both without control variables. Purple bars are means under the *purpose* treatment and orange bars under the *career* treatment. Error bars are 95% confidence intervals, computed from heteroscedasticity-robust standard errors of the relevant linear combination of coefficients. The bracket above each pair of bars reports the p-value of the treatment effect for that group, and the bracket spanning men and women reports the p-value on the interaction term β_2 in equation 2. Appendix Table A.4 reports the underlying regressions.

center, candidates in the *career* arm scored 4.5 points higher on average (p-value = 0.053) and were 26 percentage points more likely to receive a fellowship offer (p-value = 0.019, Appendix Table A.6), although these comparisons condition on attendance, which treatment may itself affect. The effect also does not come from the written stage of the application. Among candidates who submitted, essay scores are similar across treatments for both men (19.3 versus 19.1, p-value = 0.771) and women (18.9 versus 19.4, p-value = 0.633), and the probability of being invited to the assessment center conditional on submitting does not differ either (p-value = 0.219 for men and 0.916 for women). Whatever the career message changes about the composition of the pool therefore becomes visible only at the in-person assessment. This is the pattern behind Hypothesis 2: the career message should draw in candidates with better outside options, and because the organization applies a fixed bar, more of them clear it. The framework also allows this to happen without a change in the number of applicants, which is the relevant case for women. Their application rate does not respond to the message, but among those who attend the assessment, their scores are higher.

In summary, compared to the *purpose* treatment, the *career* treatment appears to attract higher-quality applicants, with no evidence that the improvement differs by gender. It also increases application rates among men, but not among women. These results suggest that emphasizing career benefits leads to positive selection on quality into the teaching fellowship, a result similar to the one found by Ashraf et al. (2020). Next, we examine the preferences of candidates recruited under these two treatment conditions.

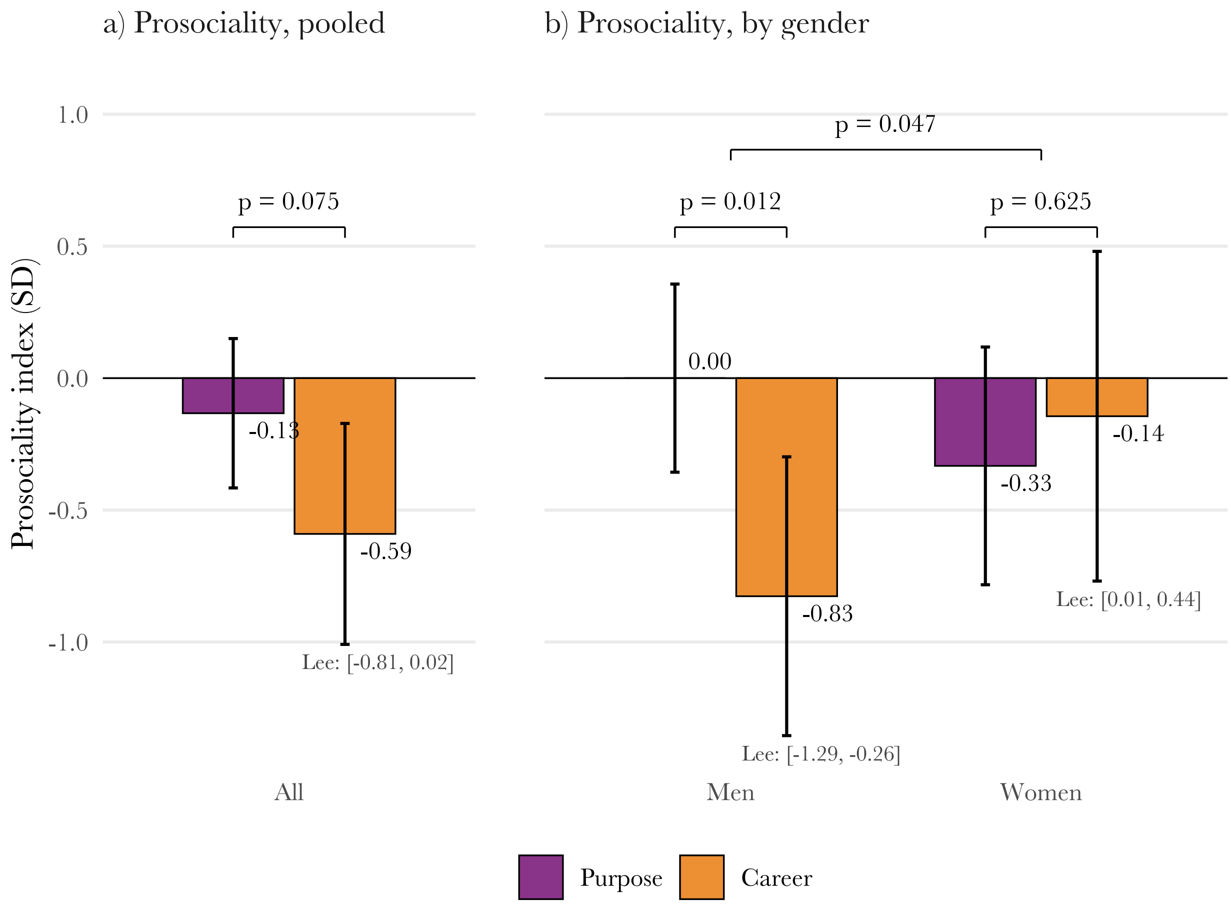
5.3 Applicant Social Preferences

In Sections 5.1 and 5.2, we established that emphasizing career benefits increases the number of male applicants while leaving the number of female applicants unchanged. It also appears to attract higher-quality candidates, at least as perceived by the NGO, since those applicants are more likely to receive offers and ultimately get hired, though as noted above these estimates are imprecise. We now examine whether candidates recruited through these two different pipelines differ in terms of their preferences, particularly their social preferences.

The preference measure is available only for candidates who applied, were invited to the in-person assessment, attended it, and completed the survey (response rate approximately 85%). Given the small and selected sample, we regard the estimates in this section as suggestive, and we probe their sensitivity to selection below. We do need to note that the sample we do observe appears relatively prosocial: in the real-donation component of the index, 76% of respondents donated at least part of their survey payment, and 24% donated all of it.

Turning to the treatment effects (Figure 5, panel a), candidates recruited through the *career* treatment are overall 0.46 standard deviations less prosocial than those recruited through the *purpose* treatment (p-value = 0.075). Panel b shows that this effect is driven entirely by men, who are 0.83 standard deviations less prosocial under the *career* treatment

Figure 5: Social Preferences



Notes: The outcome is the prosociality index constructed as described in Section 4.2, with each component standardized using the mean and standard deviation of men in the *purpose* treatment, which is therefore the reference group, and averaged. Panel a plots treatment-arm means estimated from equation 1, and panel b from equation 2, both without control variables. Purple bars are means under the *purpose* treatment and orange bars under the *career* treatment. Error bars are 95% confidence intervals, computed from heteroscedasticity-robust standard errors of the relevant linear combination of coefficients. The bracket above each pair of bars reports the p-value of the treatment effect for that group, and the bracket spanning men and women reports the p-value on the interaction term β_2 in equation 2. The index is observed only for candidates who applied, were invited to the in-person assessment, attended it, and responded to the survey, so the sample is selected and treatment may itself affect selection. The range printed beside each career-arm whisker gives Lee (2009) bounds on the treatment effect.

(p-value = 0.012). Women show no difference across treatments, and we reject equality of the two effects (interaction p-value = 0.047). The male effect and the gender difference persist when restricting the sample to candidates who received a job offer, although the pooled effect is no longer statistically significant. Our framework does not predict the direction of this effect. Lee (2009) bounds address the concern that treatment affects who reaches the survey. For men the bounds run from −1.29 to −0.26 standard deviations, remaining negative under any selection pattern consistent with monotonicity; the pooled bounds (−0.81 to +0.02) do not exclude zero. The bounds are on point estimates rather than confidence intervals, so we read them as a check on the sign of the male effect, not as a significance test.

In summary, we find suggestive evidence that emphasizing career benefits involves a trade-off: it attracts more male applicants and more candidates who meet the organization's hiring bar, but the men it attracts are less prosocial. This finding contrasts with Ashraf et al. (2020), who find no such trade-off, and aligns more closely with Deserranno (2019), who finds that higher expected earnings deter the most prosocial applicants. Our results add the gender dimension: the same message produces a trade-off for one gender and not the other.

5.4 Additional Analyses

Early versus Late Appliers The application portal was open for three months,[13] allowing for substantial variation in the time taken to register on the website and submit a completed application. Appendix Figure B.3 shows the distribution of completion times among candidates who submitted before our intervention began. We use this group because our experimental sample consists, by construction, of candidates who had not yet submitted two weeks before the deadline, and because the reminder emails may themselves have affected when candidates submitted. Approximately 28% of both men and women completed the application on the same day they started, and the median completion time was 5 days (4.5 for men and 7.5 for women). In our experimental sample the median was 46.5 days (52.5 for men and 37.5 for women), though this partly reflects how the sample was constructed, since candidates who had already submitted were excluded.

We next investigate the demographic correlates of applying late, since our experimental sample consists of late appliers by construction. We define *late appliers* as candidates who took longer than the median of five days to complete their application. Appendix Figure B.4 displays coefficients from an OLS regression of this indicator on demographic characteristics. We find no substantial or systematic observable differences between early and late appliers. The only coefficients distinguishable from zero are those for the month of registration: candidates who registered closer to the deadline had less time available and therefore submitted

[13] The NGO briefly reopened the application portal for an additional three weeks due to on-campus recruitment events. We do not include outcomes from this later period in our analysis.

faster, which mechanically makes them less likely to be classified as late appliers.

Applicant Use of LLMs Our experiment took place in the months after ChatGPT became widely available. Using Pangram Text (Emi and Spero, 2024), which estimates the probability that a given text was AI-generated and which Jabarian and Imas (2025) find outperforms other available detectors, we find that 52% of submitted applications contained at least one answer with a probability of AI generation above 99%. Candidates in the *career* treatment were 15.7 percentage points more likely to submit at least one such answer (p-value = 0.062), with no difference between men and women (p-value = 0.96).[14] We do not take a stance on whether using AI in applications is positive or negative; among non-experimental applicants, those submitting within a day were more likely to have used AI, which could reflect lower effort or higher productivity.

6 Discussion

Why might the *career* treatment attract higher-quality applicants, and why is there a trade-off only for men? The framework in Section 3.3 left the direction of selection on prosociality open, and our results point to different answers for men and women. For men, the pattern is the one in Deserranno (2019): the *career* treatment increased applications by 71%, and the men it attracted appear substantially less prosocial. The logic of Ashraf et al. (2020) suggests why. If the returns to skill are higher in the outside option, the most able candidates who still apply for a prosocial job do so because they value its mission. In their setting this implies no trade-off among those actually hired. Our male results differ: we observe prosociality only among candidates who reached the in-person assessment, so the men we compare have already cleared one of the organization's screening bars, and there the *career* treatment still attracts candidates who are substantially less prosocial. One reason may be that the fellowship's career value is large relative to these candidates' outside options, so once the fellowship itself advances a career, able men no longer need prosocial motivation to find it worthwhile. For women, the pattern is closer to Ashraf et al. (2020): the number of applicants did not change, but the women who applied under the *career* treatment appear abler, while their prosociality was, if anything, higher.

[14] Alternative measures, such as the average probability across all questions or thresholds of 90% or 50%, give qualitatively similar results.

Why the message relaxes the participation margin only for men we cannot test directly. The prosociality results are consistent with the weight-based reading behind Hypothesis 1: for women, the able candidates who apply still do so partly out of prosocial motivation, while for men the career message alone made the fellowship worth applying to. But more men may simply sit close to the application margin, where a small change in perceived value is enough to tip them, or a two-year rural placement may carry different career value or personal costs for men and women. These differences need not be fixed: Ashraf et al. (2025) find that an intervention making purpose more salient among employees of a large firm narrowed gender gaps in stated priorities and in behavior. Our treatments act at an earlier stage, on who enters the applicant pool rather than on those already working inside the organization.

Can observed differences in quality be caused by changes in effort? The higher quality of applicants is consistent with positive selection induced by the *career* treatment. Alternatively, it could be that the *career* treatment motivates people to put more effort in the selection process. While it is impossible to fully distinguish higher effort from better selection, it is unlikely for our results to be driven purely by changes in effort. First, essay scores among candidates who submitted do not differ across treatments, and the quality difference appears only at the in-person assessment months later, the opposite of what a story based on effort induced by the emails would predict. Second, applicants in the *career* treatment are more likely to use AI in their application, which is consistent with spending less rather than more time on it. Third, our preference measures are consistent with selection effects, and it is unlikely that these measures would be driven solely by applicants remembering which recruitment email they received, since the preference survey was administered two months after they submitted the application.

Can observed differences in quality be caused by changes in employer screening? The experiment was conducted in a double-blind manner, so it is unlikely that the treatment influenced recruiters' screening criteria. To assess this quantitatively, we check whether observable characteristics are equally predictive of receiving an offer across treatment groups (Ashraf et al., 2020; Delfino, 2024). We regress offer rates on observable characteristics and strata variables separately by treatment arm and test whether the coefficients differ. Figure B.5 plots these coefficients. Observable characteristics are similarly predictive of offer rates in both treatment arms, with no systematic differences.

7 Conclusion

Recruitment messages shape who applies to teach, and therefore who ends up in the classroom. We conducted a randomized field experiment with a Ghanaian education NGO, varying whether recruitment emails emphasized the career benefits of a teaching fellowship or its social impact. The same message produced different responses by gender. Emphasizing career benefits increased male applications by roughly 70%, while female applications were unchanged. We also find suggestive evidence that it increased the number of candidates who met the organization's fixed hiring bar, and that the men it attracted, but not the women, were less prosocial.

These results matter for organizations that recruit with several objectives in mind. A recruiter chooses one message, but cares about the size of the applicant pool, the quality of applicants, their motivation, and often the gender composition of the workforce. Our findings show that a single message can move these objectives in different directions, and differently for men and women, so the choice of a recruitment message is also a choice about the composition of the applicant pool. Organizations that want the quality gains of career messaging without changing the gender composition of the pool could test messages targeted by audience, or mixed messages.

One limitation of our study is that we cannot observe on-the-job outcomes for teachers, so we cannot say whether the candidates who look best at the assessment stage are also the best teachers in practice. Our sample is one cohort of applicants to one organization, recruited among university graduates, and the downstream outcomes rest on small numbers of offers and hires. The two studies most related to ours, Ashraf et al. (2020) and Deserranno (2019), conducted on very different populations in Zambia and Uganda, disagree on whether attracting talent comes at the cost of prosociality; our results add a gender dimension that their pooled or all-female samples could not show. Further research could follow recruits into the classroom to test whether the selection induced by recruitment messages carries through to teacher performance and retention.

References

Abebe, Girum, A. Stefano Caria, and Esteban Ortiz-Ospina. 2021. "The Selection of Talent: Experimental and Structural Evidence from Ethiopia." *American Economic Review* 111 (6): 1757–1806. 10.1257/aer.20190586.

Abraham, Lisa, Johannes Hallermeier, and Alison Stein. 2024. "Words Matter: Experimental Evidence from Job Applications." *Journal of Economic Behavior & Organization* 225 348–391. 10.1016/j.jebo.2024.06.013.

Ajzenman, Nicolás, Eleonora Bertoni, Gregory Elacqua, Luana Marotta, and Carolina Méndez Vargas. 2024. "Altruism or Money? Reducing Teacher Sorting Using Behavioral Strategies in Peru." *Journal of Labor Economics* 42 (4): 1049–1091. 10.1086/725166.

Akseer, Spogmai, and Ximena Játiva. 2021. "Time to Teach: Teacher Attendance and Time on Task in Ghana."Technical report, UNICEF Office of Research - Innocenti, Florence.

Ashraf, Nava, Oriana Bandiera, Edward Davenport, and Scott S. Lee. 2020. "Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services." *American Economic Review* 110 (5): 1355–1394. 10.1257/aer.20180326.

Ashraf, Nava, Oriana Bandiera, Virginia Minni, and Luigi Zingales. 2025. "Meaning at Work." NBER Working Paper 33843, National Bureau of Economic Research. 10.3386/w33843.

Brown, Christina, and Tahir Andrabi. 2023. "Inducing Positive Sorting through Performance Pay: Experimental Evidence from Pakistani Schools." January. 10.35489/BSG-RISE-WP_2023/123.

Burbano, Vanessa, Nicolas Padilla, and Stephan Meier. 2024. "Gender Differences in Preferences for Meaning at Work." *American Economic Journal: Economic Policy* 16 (3): 61–94. 10.1257/pol.20220121.

Card, David, Fabrizio Colella, and Rafael Lalive. 2024. "Gender Preferences in Job Vacancies and Workplace Gender Diversity." *The Review of Economic Studies* rdae085. 10.1093/restud/rdae085.

Coffman, Katherine B., Manuela R. Collis, and Leena Kulkarni. 2024. "Whether to Apply." *Management Science* 70 (7): 4649–4669. 10.1287/mnsc.2023.4907.

Dal Bó, Ernesto, Frederico Finan, and Martín A. Rossi. 2013. "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service*." *The Quarterly Journal of Economics* 128 (3): 1169–1218. 10.1093/qje/qjt008.

Del Carpio, Lucía, and Maria Guadalupe. 2022. "More Women in Tech? Evidence from a Field Experiment Addressing Social Identity." *Management Science* 68 (5): 3196–3218. 10.1287/mnsc.2021.4035.

Delfino, Alexia. 2024. "Breaking Gender Barriers: Experimental Evidence on Men in Pink-Collar Jobs." *American Economic Review* 114 (6): 1816–1853. 10.1257/aer.20220582.

Deserranno, Erika. 2019. "Financial Incentives as Signals: Experimental Evidence from the Recruitment of Village Promoters in Uganda." *American Economic Journal: Applied Economics* 11 (1): 277–317. 10.1257/app.20170670.

Duflo, Esther, Rema Hanna, and Stephen P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102 (4): 1241–1278. 10.1257/aer.102.4.1241.

Emi, Bradley, and Max Spero. 2024. "Technical Report on the Pangram AI-Generated Text Classifier." July. 10.48550/arXiv.2402.14873.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences*." *The Quarterly Journal of Economics* 133 (4): 1645–1692. 10.1093/qje/qjy013.

Flory, Jeffrey A., Andreas Leibbrandt, and John A. List. 2015. "Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions." *The Review of Economic Studies* 82 (1): 122–155. 10.1093/restud/rdu030.

Flory, Jeffrey A., Andreas Leibbrandt, Christina Rott, and Olga Stoddard. 2021. "Increasing Workplace Diversity: Evidence from a Recruiting Experiment at a Fortune 500 Company." *Journal of Human Resources* 56 (1): 73–92.

Fuchs, Larissa, Matthias Heinz, Pia Pinger, and Max Thon. 2024. "How to Attract Talent? Field-Experimental Evidence on Emphasizing Flexibility and Career Opportunities in Job Advertisements." Discussion Paper 332, ECONtribute Discussion Paper Series.

Jabarian, Brian, and Alex Imas. 2025. "Artificial Writing and Automated Detection." NBER Working Paper 34223, National Bureau of Economic Research. 10.3386/w34223.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75 (1): 83–119. 10.1111/j.1468-0262.2007.00733.x.

Kuhn, Peter, and Kailing Shen. 2023. "What Happens When Employers Can No Longer Discriminate in Job Ads?" *American Economic Review* 113 (4): 1013–1048. 10.1257/aer.20211127.

Lavy, Victor. 2009. "Performance Pay and Teachers' Effort, Productivity, and Grading Ethics." *American Economic Review* 99 (5): 1979–2011. 10.1257/aer.99.5.1979.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *The Review of Economic Studies* 76 (3): 1071–1102. 10.1111/j.1467-937X.2009.00536.x.

Marinescu, Ioana, and Ronald Wolthoff. 2020. "Opening the Black Box of the Matching Function: The Power of Words." *Journal of Labor Economics* 38 (2): 535–568. 10.1086/705903.

Mas, Alexandre, and Amanda Pallais. 2017. "Valuing Alternative Work Arrangements." *American Economic Review* 107 (12): 3722–3759. 10.1257/aer.20161500.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119 (1): 39–77. 10.1086/659655.

Oh, Suanna. 2023. "Does Identity Affect Labor Supply?" *American Economic Review* 113 (8): 2055–2083. 10.1257/aer.20211826.

Palffy, Patricia, Patrick Lehnert, and Uschi Backes-Gellner. 2023. "Countering Gender-Typicality in Occupational Choices: An Information Intervention Targeted at Adolescents."

UNESCO Institute for Statistics. 2026. "Percentage of Teachers Who Are Female, by Level of Education."

Wiswall, Matthew, and Basit Zafar. 2018. "Preference for the Workplace, Investment in Human Capital, and Gender*." *The Quarterly Journal of Economics* 133 (1): 457–507. 10.1093/qje/qjx035.

World Bank. 2022. "The State of Global Learning Poverty: 2022 Update."Technical report, World Bank, Washington, DC.

"Career vs. Social Impact: Gendered Responses in Teacher Recruitment in Ghana"

Online Appendix

Ursa Krenk
Kobbina Awuah

Table of Contents

A Tables

Table A.1: Summary Statistics and Balance

	Men			Women			Joint	
	(N)	(Mean)	(SD)	(N)	(Mean)	(SD)	F-stat	p-value
Female	949	0.000	0.000	718	1.000	0.000	0.282	0.595
National Service	949	0.604	0.489	718	0.504	0.500	1.226	0.268
Top 5 University	949	0.870	0.336	718	0.862	0.345	0.161	0.688
Still in Education	949	0.360	0.480	718	0.464	0.499	0.575	0.448
Mother tongue								
Twi	949	0.613	0.487	718	0.653	0.476	3.477	0.062
Ewe	949	0.098	0.297	718	0.075	0.264	0.131	0.718
Ga/Dangme	949	0.076	0.265	718	0.084	0.277	0.128	0.720
Northern lang.	949	0.213	0.410	718	0.188	0.391	3.021	0.082
Home region								
Northern Regions	949	0.298	0.458	718	0.270	0.444	2.276	0.132
Application & Registration								
Word of mouth	949	0.296	0.457	718	0.273	0.446	0.249	0.618
Applied before	949	0.142	0.349	718	0.077	0.266	0.150	0.698
Registered in Junae	949	0.163	0.370	718	0.160	0.367	4.609	0.032

Notes: Table shows summary statistics and balance checks of baseline variables from the registration form. Columns 7 and 8 report the F-statistic and p-value from a joint test of significance of the treatment dummy on each of the row variables, using robust standard errors. "National Service" is a dummy for having completed National Service, "Top-5 University" is a dummy for graduating from one of Ghana's five largest universities (KNUST, University of Ghana, University of Cape Coast, University of Education, Winneba, and University for Development Studies), "Still in Education" is a dummy for currently being in the final university year, as opposed to already having graduated, "Northern Regions" is a dummy for coming from one of Ghana's northern regions, "Word of Mouth" is a dummy equalling one if the person found out about the program through word of mouth (was told by a friend, or through a campus event), as opposed to finding out about the program through social or traditional media.

Table A.2: Application Rates

	Applied					
	(1)	(2)	(3)	(4)	(5)	(6)
Career	0.024	0.022	0.014	0.053	0.051	0.041
	(0.014)	(0.014)	(0.013)	(0.020)	(0.019)	(0.019)
Career × Female				-0.066	-0.067	-0.062
				(0.027)	(0.027)	(0.027)
Female				0.001	0.001	0.007
				(0.019)	(0.018)	(0.018)
Inference (p-values)						
Career	0.085	0.111	0.282	0.007	0.009	0.031
Career × Female				0.015	0.013	0.020
LATE (β)						
Career	0.038	0.035	0.023	0.088	0.085	0.068
Career × Female				-0.109	-0.110	-0.100
Mean (Purpose)	0.076	0.076	0.076	0.076	0.076	0.076
Controls	None	Strata	Full	None	Strata	Full
N	1,667	1,667	1,667	1,667	1,667	1,667

Notes: The table presents OLS regression coefficients. The outcome is a dummy equaling one if the candidate submitted an application. Columns 1–3 report the pooled treatment effect (equation 1). Columns 4–6 add a female dummy and its interaction with treatment (equation 2), so the *Career* coefficient is the effect among men and the sum of *Career* and *Career* × *Female* is the effect among women. The first column in each block includes no control variables, so the reported effects match Figure 3 exactly. The second adds the strata variables: randomization-round fixed effects and a female dummy, the two variables the randomization was stratified on; in columns 4–6 the female dummy is already part of the specification. The third adds the full control vector: randomization-round fixed effects and controls for whether the photograph in the email matched the candidate's own gender, completion of national service, being in the final year of study, attending one of Ghana's five largest universities, coming from a northern region, being in the top GPA band, having applied to the NGO before, having heard about it through a personal referral, and having registered in June. Robust (HC1) standard errors are in parentheses. *Mean (Purpose)* is the mean of the outcome among candidates who received the purpose advertisement. *LATE* rows report coefficients from IV regressions in which assignment to the career advertisement instruments for having been assigned that advertisement and having opened at least one reminder email; only point estimates are shown.

Table A.3: Assessment Center Invitation and Attendance

	Invited				**Attended**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Career	0.008	0.004	0.023	0.016	0.006	0.002	0.009	0.004
	(0.011)	(0.011)	(0.016)	(0.016)	(0.009)	(0.008)	(0.012)	(0.012)
Career × Female			-0.032	-0.028			-0.007	-0.004
			(0.022)	(0.022)			(0.017)	(0.016)
Female			-0.003	0.003			-0.011	-0.007
			(0.016)	(0.016)			(0.011)	(0.011)
Inference (p-values)								
Career	0.464	0.729	0.160	0.317	0.488	0.800	0.454	0.743
Career × Female			0.148	0.210			0.664	0.797
LATE (β)								
Career	0.013	0.006	0.038	0.026	0.010	0.003	0.015	0.007
Career × Female			-0.053	-0.044			-0.012	-0.007
Mean (Purpose)	0.053	0.053	0.053	0.053	0.029	0.029	0.029	0.029
Controls	No	Yes	No	Yes	No	Yes	No	Yes
N	1,667	1,667	1,667	1,667	1,667	1,667	1,667	1,667

Notes: The table presents OLS regression coefficients. Columns 1–4 report unconditional invitation rates and columns 5–8 unconditional attendance rates at the in-person assessment center, both dummies coded zero for candidates who did not reach the stage, so both are defined on the full randomized sample. Within each outcome, the first two columns report the pooled treatment effect (equation 1) and the last two add a female dummy and its interaction with treatment (equation 2), so the *Career* coefficient is the effect among men and the sum of *Career* and *Career* × *Female* is the effect among women. Odd-numbered columns include no control variables. Even-numbered columns add randomization-round fixed effects and the controls listed in the notes to Table A.2. Robust (HC1) standard errors are in parentheses. *Mean (Purpose)* is the mean of the outcome among candidates who received the purpose advertisement. *LATE* rows report coefficients from IV regressions in which assignment to the career advertisement instruments for having been assigned that advertisement and having opened at least one reminder email; only point estimates are shown.

Table A.4: Offer and Hiring Rates

	Offer				Hired			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Career	0.013	0.010	0.015	0.011	0.011	0.009	0.015	0.012
	(0.008)	(0.007)	(0.011)	(0.011)	(0.006)	(0.006)	(0.009)	(0.009)
Career × Female			-0.005	-0.002			-0.010	-0.008
			(0.015)	(0.015)			(0.013)	(0.012)
Female			-0.009	-0.006			-0.001	0.001
			(0.009)	(0.009)			(0.008)	(0.008)
Inference (p-values)								
Career	0.090	0.175	0.168	0.298	0.092	0.165	0.096	0.165
Career × Female			0.744	0.868			0.424	0.497
LATE (β)								
Career	0.021	0.016	0.026	0.019	0.017	0.014	0.025	0.020
Career × Female			-0.010	-0.005			-0.018	-0.015
Mean (Purpose)	0.019	0.019	0.019	0.019	0.012	0.012	0.012	0.012
Controls	No	Yes	No	Yes	No	Yes	No	Yes
N	1,667	1,667	1,667	1,667	1,667	1,667	1,667	1,667

Notes: The table presents OLS regression coefficients. Columns 1–4 report unconditional offer rates and columns 5–8 unconditional hiring rates, both dummies coded zero for candidates who did not reach the stage, so both are defined on the full randomized sample. Within each outcome, the first two columns report the pooled treatment effect (equation 1) and the last two add a female dummy and its interaction with treatment (equation 2), so the *Career* coefficient is the effect among men and the sum of *Career* and *Career* × *Female* is the effect among women. Odd-numbered columns include no control variables, so the reported effects match the corresponding figures exactly. Even-numbered columns add randomization-round fixed effects and the controls listed in the notes to Table A.2. Robust (HC1) standard errors are in parentheses. *Mean (Purpose)* is the mean of the outcome among candidates who received the purpose advertisement. *LATE* rows report coefficients from IV regressions in which assignment to the career advertisement instruments for having been assigned that advertisement and having opened at least one reminder email; only point estimates are shown.

Table A.5: Prosociality

	Prosociality index		Prosociality index	
	(1)	(2)	(3)	(4)
Career	-0.457	-0.826	-0.439	-0.867
	(0.251)	(0.316)	(0.279)	(0.346)
Career × Female		1.014		1.131
		(0.495)		(0.555)
Female		-0.332		-0.416
		(0.285)		(0.357)
Inference (p-values)				
Career	0.075	0.012	0.125	0.017
Career × Female		0.047		0.050
Lee (2009) bounds on Career				
Lower	-0.812	-1.292	-1.158	-1.658
Upper	0.022	-0.256	0.323	-0.033
Sample	Full	Full	Offer recipients	Offer recipients
Mean (Purpose)	-0.133	-0.133	-0.153	-0.153
N	46	46	37	37

Notes: The table presents OLS regression coefficients. The outcome is an index of prosociality built from four standardized measures: the candidate's stated willingness to give to good causes, a hypothetical donation of 150 GHS, the amount actually donated out of the survey participation incentive, and how highly the candidate ranked making a positive difference in the community among five job attributes. Each measure is standardized using the mean and standard deviation of men in the purpose arm and the four are averaged, so coefficients are in standard deviations of that group. The estimation sample is candidates who attended the in-person assessment and responded to the follow-up survey. Columns 1–2 use the full survey sample and columns 3–4 restrict it to candidates who received an offer. Within each, the first column reports the pooled treatment effect (equation 1) and the second adds a female dummy and its interaction with treatment (equation 2), so the *Career* coefficient is the effect among men and the sum of *Career* and *Career* × *Female* is the effect among women. No column includes control variables: with 46 respondents the sample cannot support the control vector of Table A.2. Robust (HC1) standard errors are in parentheses. *Mean (Purpose)* is the mean of the outcome among purpose-arm candidates in that column's estimation sample. Because the survey is answered only by assessment attendees, and treatment may affect who reaches that stage, these are not experimental comparisons. The Lee (2009) panel reports trimming bounds on the *Career* coefficient of its own column, so the pooled columns bound the pooled effect and the interacted columns the effect among men; the bounds for women are [0.01, 0.44] in the full survey sample and [-0.30, 0.82] among offer recipients. They are wider in columns 3 and 4 because selection into the offer-recipient sample runs through the offer itself: the career arm's higher offer rate means nearly half of its observations are trimmed.

Table A.6: Conditional Outcomes

	Attended | invited		**Assessment score**		**Offer | attended**	
	(1)	(2)	(3)	(4)	(5)	(6)
Career	0.023	-0.060	4.534	3.445	0.264	0.212
	(0.103)	(0.129)	(2.290)	(2.912)	(0.109)	(0.139)
Career × Female		0.215		3.543		0.163
		(0.218)		(4.739)		(0.226)
Female		-0.171		0.449		-0.063
		(0.154)		(3.975)		(0.215)
Inference (p-values)						
Career	0.823	0.644	0.053	0.243	0.019	0.133
Career × Female		0.327		0.458		0.476
Mean (Purpose)	0.545	0.545	55.263	55.263	0.667	0.667
N	95	95	53	53	53	53

Notes: The table reports OLS regression coefficients with no control variables. Columns 1–2 report attendance at the in-person assessment center among the 95 candidates invited to it; columns 3–4 the average score across the four assessment tasks and columns 5–6 an indicator for receiving an offer, both among the 53 candidates who attended. Within each outcome, the first column reports the pooled treatment effect (equation 1) and the second adds a female dummy and its interaction with treatment (equation 2), so the *Career* coefficient is the effect among men and the sum of *Career* and *Career × Female* is the effect among women. Robust (HC1) standard errors are in parentheses. *Mean (Purpose)* is the mean of the outcome among purpose-arm candidates in that column's estimation sample. Because all three outcomes are measured only among assessment attendees, and treatment may affect who attends, they are not experimental comparisons and should be read as descriptive.

Table A.7: Robustness: Survey-Recruited Candidates

	Application	**Offer**	**Hired**	**Prosociality**
	(1)	(2)	(3)	(4)
Career	0.031	0.011	0.012	-0.863
	(0.019)	(0.010)	(0.009)	(0.306)
Career × Female	-0.050	-0.003	-0.009	1.051
	(0.027)	(0.014)	(0.012)	(0.489)
Female	-0.003	-0.005	0.001	-0.332
	(0.018)	(0.009)	(0.007)	(0.284)
Inference (p-values)				
Career	0.102	0.285	0.160	0.007
Career × Female	0.058	0.828	0.471	0.037
Controls	Yes	Yes	Yes	No
Mean (Purpose)	0.080	0.019	0.012	-0.133
N	1,723	1,723	1,723	47

Notes: The table re-estimates the main outcomes on all 1,723 randomized candidates, including the 56 recruited through a parallel survey advertised on social media. Those 56 were shown one of the four statement-and-photograph combinations inside the survey and were held in the same arm when they later registered on the NGO's webpage, so whether they appear in the data at all is affected by treatment. Every other table and figure in the paper excludes them; comparing this table with Tables A.2, A.4 and A.5 shows what the exclusion changes. All columns report the gender-interaction specification (equation 2), so the *Career* coefficient is the effect among men and the sum of *Career* and *Career* × *Female* is the effect among women. The outcomes are dummies for submitting an application, receiving an offer, and being hired, all unconditional and coded zero for candidates who did not reach that stage, and the prosociality index of Table A.5. Controls are randomization-round fixed effects and the controls listed in the notes to Table A.2. An indicator for survey recruitment adds nothing to them, since the survey recruits are exactly randomization round 5 and the round fixed effects absorb it. Column 4 includes no controls, following Table A.5: 47 respondents cannot support a 14-parameter control vector. Robust (HC1) standard errors are in parentheses. *Mean (Purpose)* is the mean of the outcome among candidates who received the purpose advertisement.

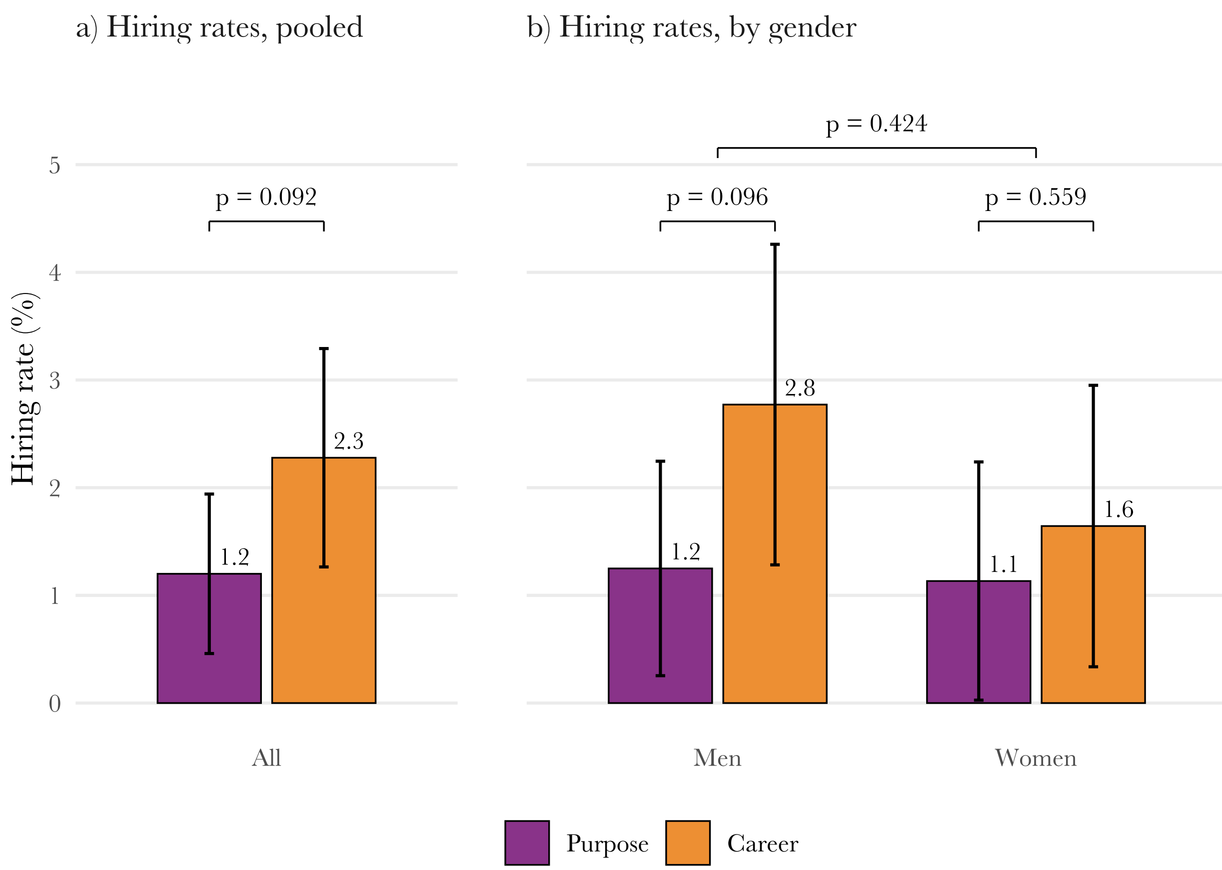
Table A.8: Treatment Statements

Career treatment	
1.	Join [name of the NGO] and invest in your future. The fellowship program offers a unique opportunity to develop your leadership abilities, enhance your skill set, and distinguish yourself in the job market. Embrace the opportunity!
2.	Join [name of the NGO] and unlock your potential for career growth. Through the fellowship program, you'll gain practical experience, develop transferable skills, and create a strong foundation for long-term success in your chosen field.
3.	Join [name of the NGO] and gain valuable skills and experiences that will enhance your career journey. You will receive practical training and support that will empower you to excel in your chosen field. Seize the opportunity!
4.	Join [name of the NGO] and build a solid foundation for your professional journey. You will get to explore your inner potential and gain transferable skills that will make you stand out professionally. Seize the opportunity!
Purpose treatment	
1.	Join [name of the NGO] and help empower children from underprivileged communities across Ghana to reach their potential. Let's ensure that every child receives a quality education. Together we can make it happen!
2.	Join [name of the NGO] and become part of a team that works tirelessly to provide quality education to children in disadvantaged areas. All children in Ghana deserve equal educational opportunities-help make it happen!
3.	Every child in Ghana should have access to good quality education. Join [name of the NGO] and help us make this a reality!
4.	Join [name of the NGO] and play your part in creating a positive impact on the lives of children in underserved communities. Every child in Ghana deserves access to quality education. Help make it happen!

Notes: The table lists the eight statements used in the treatment boxes of the reminder emails. Statement k was included in the k-th reminder email of the corresponding treatment arm, so candidates received the statements of their assigned arm in the order shown.

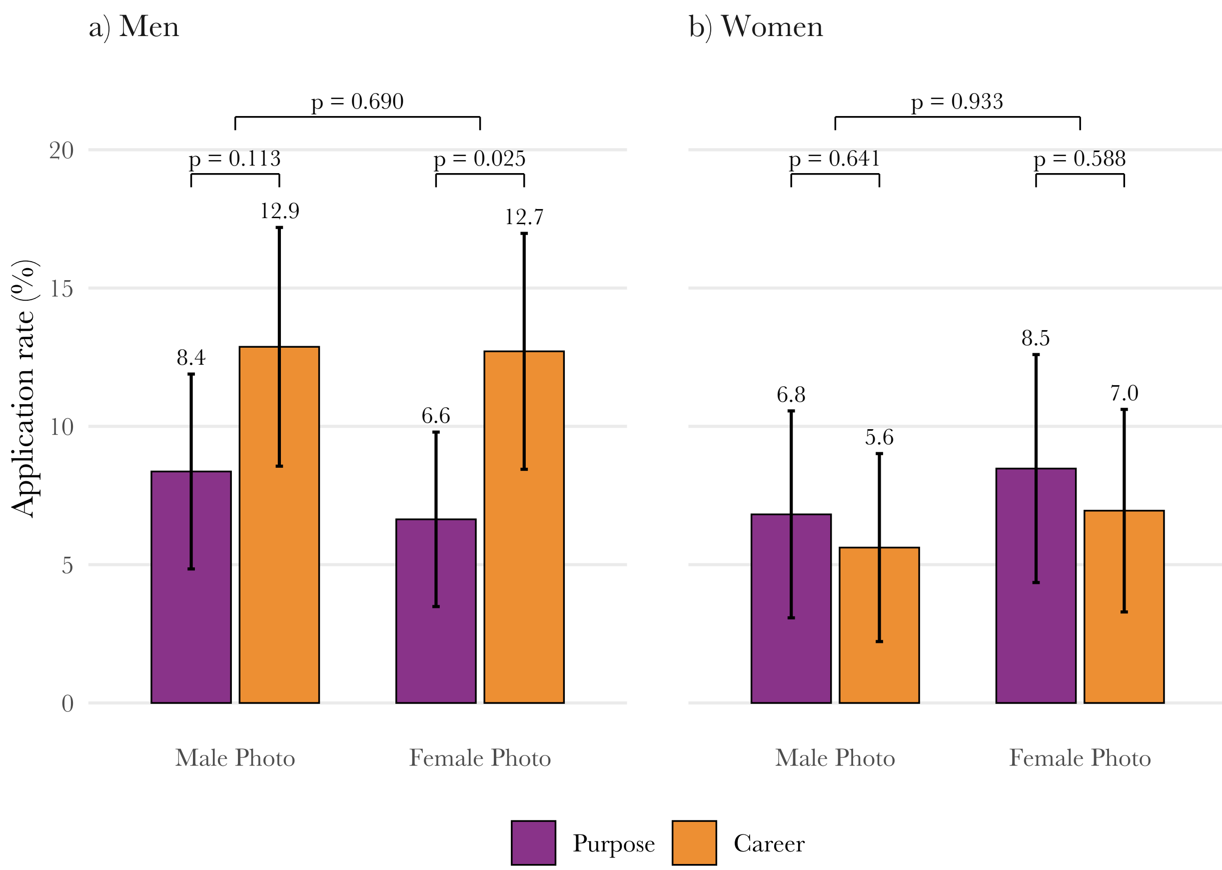
B Figures

Figure B.1: Job Hiring Rates



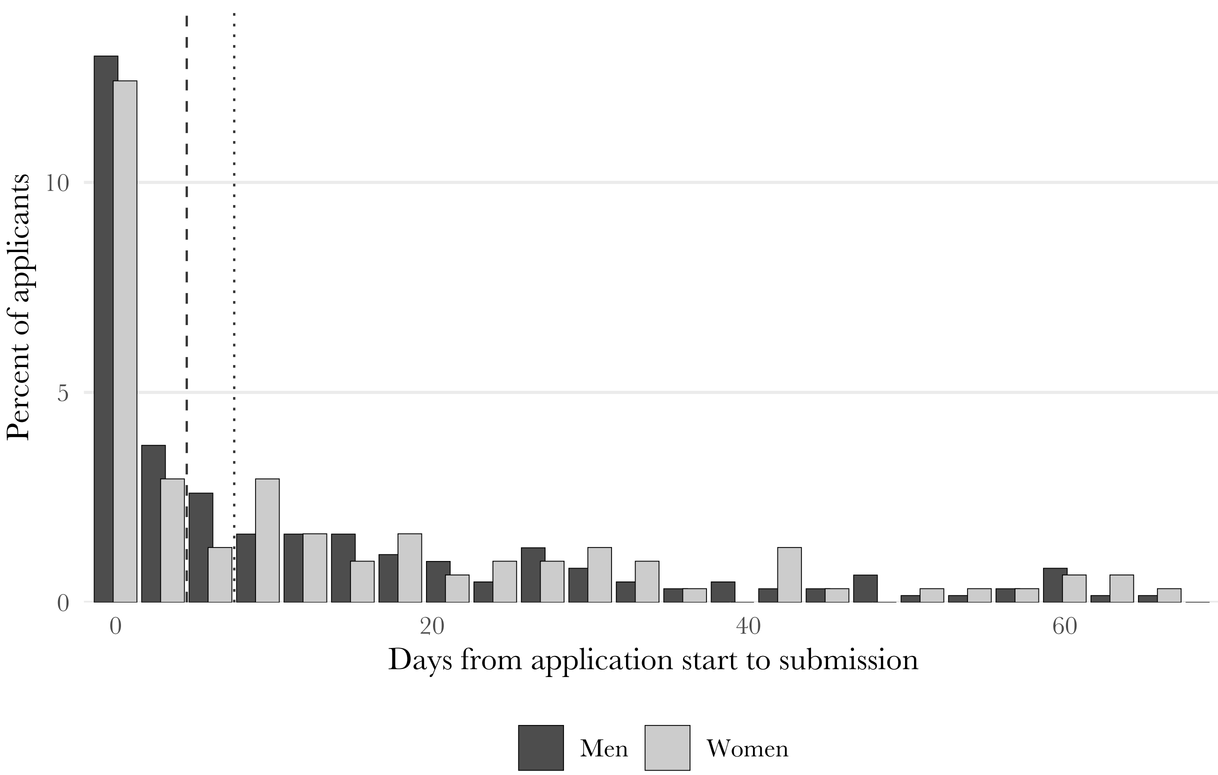
Notes: The outcome is a binary indicator of whether a candidate was hired, expressed in percent. It is unconditional, coded zero for candidates who did not apply, did not receive an offer, or declined it, so the sample is the full set of randomized candidates. Panel a plots treatment-arm means estimated from equation 1, and panel b from equation 2, both without control variables. Purple bars are means under the *purpose* treatment and orange bars under the *career* treatment. Error bars are 95% confidence intervals, computed from heteroscedasticity-robust standard errors of the relevant linear combination of coefficients. The bracket above each pair of bars reports the p-value of the treatment effect for that group, and the bracket spanning men and women reports the p-value on the interaction term β_2 in equation 2. Table A.4 reports the underlying regressions.

Figure B.2: Application Rates by Gender of the Photograph



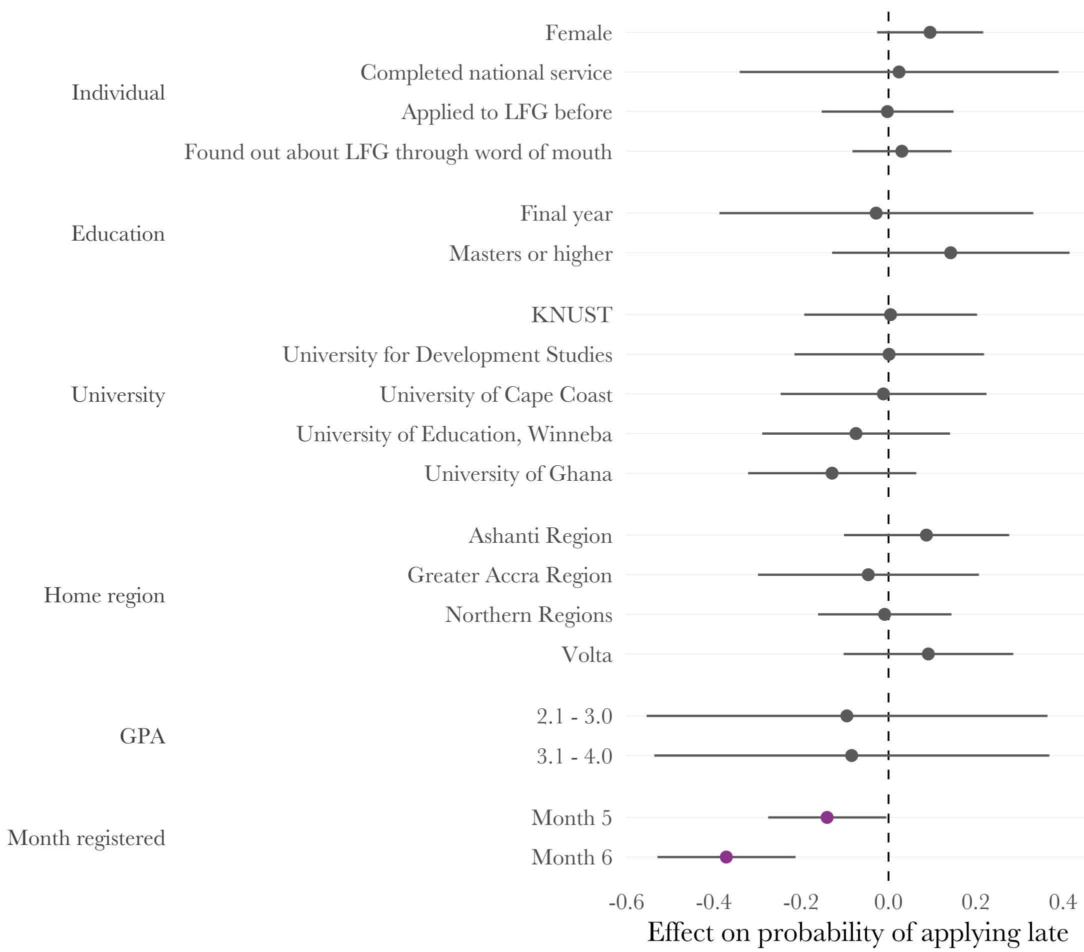
Notes: The outcome is a binary indicator of whether a candidate submitted an application, expressed in percent. Bars are treatment-arm means estimated separately for men (panel a) and women (panel b) from a regression of the outcome on the treatment indicator interacted with an indicator for the email containing a female photograph, without control variables. Purple bars are means under the *purpose* treatment and orange bars under the *career* treatment. Error bars are 95% confidence intervals, computed from heteroscedasticity-robust standard errors of the relevant linear combination of coefficients. The bracket above each pair of bars reports the p-value of the treatment effect for that photograph, and the bracket spanning the two pairs reports the p-value on the interaction between treatment and the gender of the person in the photograph.

Figure B.3: Time to Submit an Application



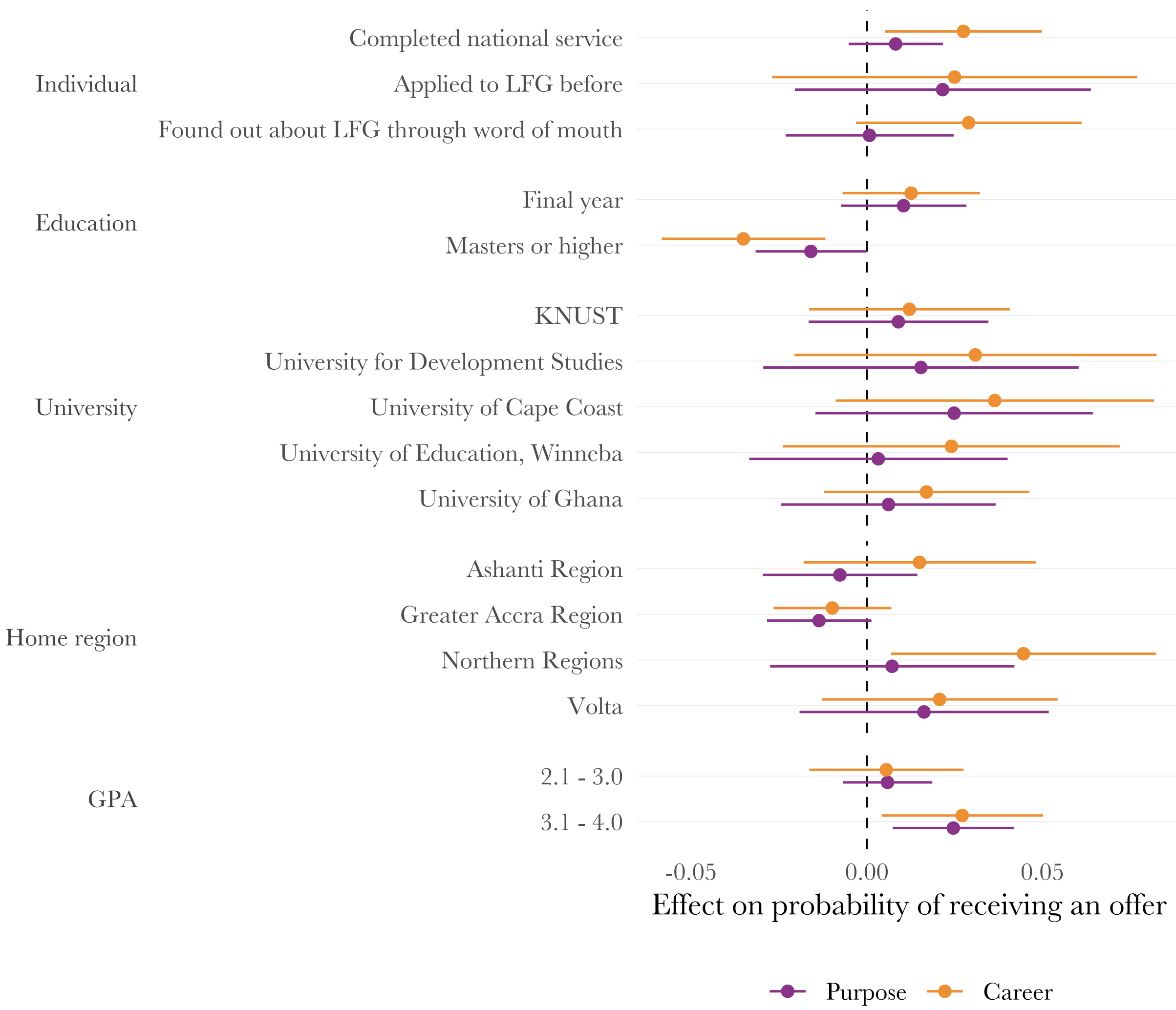
Notes: The figure shows the distribution of the time between starting and submitting an application, separately for men and women. The sample is the 308 candidates who submitted before our intervention began and therefore received no treatment email. We use this group because our experimental sample consists, by construction, of candidates who had not yet submitted two weeks before the deadline, and because the reminder emails may themselves have affected when candidates submitted. Bars show the percent of applicants of each gender in each three-day interval. The dashed and dotted vertical lines mark the median for men and women respectively.

Figure B.4: Who Applies Late?



Notes: The figure plots OLS coefficients from a regression of an indicator for taking longer than the median of five days to submit an application on the characteristics shown, estimated on the candidates who submitted before our intervention began. A positive coefficient means applying later. Horizontal lines are 95% confidence intervals from heteroscedasticity-robust standard errors, and coefficients whose interval excludes zero are highlighted. Omitted categories are a Bachelor's degree, KNUST, the Northern regions, a GPA below 2.1, and registration in the first month the portal was open.

Figure B.5: Are Observable Characteristics Equally Predictive of Offers in Both Arms?



Notes: The figure plots coefficients from OLS regressions of the unconditional offer indicator on the characteristics shown, estimated separately for each treatment arm on the full randomized sample (827 candidates in the *purpose* arm and 831 in the *career* arm with non-missing covariates). Both regressions additionally control for gender and randomization-round fixed effects; these coefficients are not displayed. Omitted categories are a Bachelor's degree, KNUST, the Northern regions, and a GPA below 2.1. Horizontal lines are 95% confidence intervals from heteroscedasticity-robust standard errors. Similar profiles across the two arms indicate that observable characteristics predict offers similarly regardless of treatment, so the offer results are unlikely to reflect differential screening by recruiters.