How to Use Recognition Rewards to Improve Quality of Bureaucrats: Experimental Evidence from a Public Teacher Training Programme in Pakistan

Zahra Mansoor *

September 2023

Abstract

On-the-job training remains the primary vehicle for upgrading the quality of bureaucrats across many low-and middle-income countries. This paper reports experimental evidence on the impact of recognition rewards on teacher knowledge and skills in a mandatory teacher training program in Punjab, Pakistan. Public school head teachers attending the training were randomized into four different recognition rewards tied to knowledge acquisition as measured by training test scores, and a control group. The first recognition reward made peer-esteem from recognition salient, the second made potential career benefits from recognition salient, and the third and fourth treatments combined the first two treatments with a motivational framing to improve the design of the rewards. Results show the following: first, recognition rewards that make career benefits salient lead to higher training test scores. Second, adding a motivational framing makes teachers overconfident reducing their performance in the training. These findings suggest that recognition rewards can be a cost-effective way to improve the effectiveness of public teacher training programs, and civil service trainings more broadly. However, their framing is nontrivial and can have significant effects.

Keywords: Recognition; Peer-esteem; Career benefits; Framing effects; Public school teachers

JEL codes: C93; I28; J3; J45; J53; M52; O15

1 Introduction

Building a high quality workforce is a central goal of all organizations, including public organisations where quality and performance of workers has direct implications for service delivery, citizen well-being, and economic growth (Besley et al., 2022). This is arguably the most critical in the case of public school teachers given their central role in driving student learning outcomes in early life, as well as labour market outcomes in adult life (Araujo et al., 2016; Bruns and Luque, 2015; Chetty et al., 2014). However, in many low-and middle-income countries public school teachers continue to lack the knowledge and skills to teach students well (Bold et al., 2017; Bruns and Luque, 2015).[1] While many governments require teachers to have specific qualifications when they are recruited, the ability to hire high quality teachers is practically constrained given observed teacher characteristics have limited ability to explain teacher effectiveness (Rockoff, 2004; Bau and Das, 2020). Hence, once recruited, on-the-job training that is commonly known as in-service training or professional development remains the primary and most critical vehicle for upgrading skills and knowledge of public school teachers to improve their quality.

Given the critical role that on-the-job training is expected to play in upgrading teacher knowledge and skills, governments around the world spend a significant amount of funding on training programs (Fryer, 2016; Popova et al., 2016). Despite this investment, the evidence on these programs in low-and middle-income countries countries is sparse and shows mixed results, with very few studies focusing on at scale programs that are implemented by governments (see McEwan (2015) for an excellent review and Piper and Korda (2011), Kerwin and Thornton (2021), and Berlinski and Busso (2017) for specific examples).[2] A helpful review of evaluated training programs by Popova et al. (2021) highlights that, amongst other factors (such as the trainings having a specific subject focus, lesson enactment, and being face-to-face), linking teacher participation in trainings to financial or non-financial incentives such as status, career incentives, or salary can make them more effective. However, at scale trainings that are implemented by governments rarely include such incentives (Popova et al., 2021), highlighting a critical need to explore whether incentives could indeed improve the design and impact of trainings.

This paper addresses this gap by asking whether introducing non-financial incentives in at scale

[1] The authors show that across seven African countries, only about 7% of fourth grade teachers had the knowledge to teach language.

[2] One exception is the evaluation of an at scale professional development program in China by Loyalka et al. (2019) which shows null effects on teacher and student outcomes.

government teacher training programs can improve their impact? We specifically focus on non-financial incentives such as recognition rewards because financial incentives are costly and can often be distortionary by crowding out intrinsic motivations (Bénabou and Tirole, 2003). Non-financial incentives on the other hand are more cost-effective and can especially work in citizen-facing departments such as education where agents are known to put a lower weight on financial incentives (Besley and Ghatak, 2005). In addition, given on-the-job trainings are typically done on a regular basis, non-financial incentives are more financially feasible in such a setting.

Non-financial incentives such as recognition rewards could be designed by leveraging different margins of efforts. For example, such rewards can either leverage motivation for *peer-esteem* through recognition in front of peers or *career benefits* through signalling quality to supervisors (Frank, 1985; Besley and Ghatak, 2008; Dewatripont et al.;1999). In addition, the way these rewards are framed matters since the decision to exert effort may depend both on the reward as well as the *information* contained in the reward (Gneezy et al., 2011).[3] While recognition rewards have been shown to be effective across several lab and field settings (Ashraf et al., 2014a; Ashraf et al.; 2014b; Neckermann and Yang, 2017; Gauri et al., 2018; Cotofan; 2021), evidence on the mechanisms through which they operate is limited posing the question of whether leveraging certain margins of effort would be more effective than others.

We take these gaps in the evidence on training programs and recognition rewards into consideration, and provide experimental evidence on the impact of a recognition program in a public teacher training program by posing two research questions: first, can non-financial incentives like recognition rewards improve teacher knowledge and skills in teacher trainings? If yes, what is the impact of recognition that makes peer-esteem salient versus recognition that makes career benefits salient? Second, does framing recognition with a motivational framing improve its impact?[4]

We answer these questions in the context of a public teacher training program in Punjab, Pakistan. Punjab is Pakistan's most populous province - the public education department employs a workforce of approximately 450,000 teachers that are spread across 52,000 schools.[5] The Teacher

[3]This is particularly the case for tournament-based schemes like recognition that can often dampen employee morale and and beliefs in ability depending on how they are perceived (Connelly et al., 2014; Ashraf et al., 2014a; Mansoor, 2019).

[4]Note that while the content and the medium of content delivery in trainings is critical, we specifically focus on how to improve the effectiveness of trainings within the constraints of their existing design. Our focus is on understanding the role of incentives in training and not the impact of the training itself.

[5]Annual School Census Data 2017.

Training Academy in Punjab, called the Quaid-e-Azam Academy for Educational Development (QAED), is the primary vehicle through which all public school teachers upgrade their skills and knowledge.[6] The academy offers a range of trainings such as standard promotion-linked trainings, as well as those focused on content (e.g., literacy or numeracy, and other subject-based trainings), pedagogy, and leadership. Apart from promotion-linked trainings where promotions are based on performance in the training as measured by training test scores, none of the programmes include any incentives linked to performance in the training. This makes the focus of this study directly policy relevant.

The experiment was designed and implemented in collaboration with QAED by embedding different recognition schemes within a mandatory training on school leadership offered to head teachers. The academy randomly allocated 131 different training sessions (offered to 3,394 head teachers in 7 districts in Punjab) across 4 different recognition schemes and a control group. The recognition incentive was tied to training test scores which serves as a proxy for teacher knowledge and skill acquisition.[7] The design of the recognition arms was as follows: Treatment 1 (Peer arm) made peer-esteem salient – trainees were told that those who qualify will be provided certificates in a district-level ceremony which will be attended by their peers and colleagues in their district. Treatment 2 (Career arm) made career benefits salient - trainees were told that those who qualify will receive the certificates privately but at the same time their name will be added to an 'excellent teacher list' that would be shared with leadership in the training academy and their district office which could make them visible for future career opportunities. While formal career incentives such as promotions are purely ruled-based (i.e., based on seniority) in our setting, this treatment importantly leverages the availability of informal career incentives in the system (e.g., postings to preferred schools, transfers to lucrative lateral postings, or postings to vacant positions of higher ranks). Treatments 3 and 4 (Peer PLUS and Career PLUS) cross the first two treatments with a motivational framing targeted at boosting teacher morale and beliefs about ability with the aim to improve the the information contained in the reward, and hence its overall design.

[6]Teacher quality is widely understood as a significant constraint to improving learning outcomes. Descriptive evidence from a survey of 650 secondary school teachers highlights this showing that 40% of teachers scored less than 50% on grade 5 math problems (Mansoor, 2019).

[7]Note that the leadership training is offered to head teachers. Knowledge and skills of head teachers in management and leadership are particularly relevant and important given the growing evidence on the positive relationship between head teacher management practices and school outcomes (Leaver et al., 2019; Bloom et al., 2015.)

Our main treatment effects show two key results. First, we find that recognition can work when it makes career benefits salient. We find indicative evidence that the Career arm leads to a 0.33σ increase in training test scores as compared to the control group. In comparison, the Peer arm has a coefficient of 0.05 and is insignificant. Back of the envelope calculations show that the gain in training test scores in the Career arm is equivalent to learning gains attained in half a day of training. Given the total training was spread over four days, this indicates that the Career arm could improve cost effectiveness of the programme by about 12.5%. Second, we find that the positive effect of the Career arm backfires depending on how it is framed. The net impact of adding the motivational framing to the Career arm is to lower training test scores by 0.36σ. The net impact of adding the motivational framing to the Peer arm is also negative (although insignificant). Overall, the net impact of adding the motivational framing across both the Peer and Career arms is to reduce training test scores by 0.28σ.

Quantile treatment effects provide supporting evidence that the career benefits channel is stronger than the peer-esteem channel. We find that the Career arm has a positive coefficient in the range of 0.11-0.49 across the distribution. This is promising given treatment effects in tournament-based incentives such as recognition rewards often come from the upper tail of the distribution (see systematic review by Connelly et al., 2014). In the upper tail of the distribution, the Career arm coefficients are significant and also significantly different from the Peer arm. Kolmogorov–Smirnov tests of equality of distribution between the Peer and Career arm confirm that the distributions are significantly different from each other (p-value <0.01).

Next, we show that the treatment effects in the Career arm come from teachers who have stronger career concerns, such as through the permanent nature of their contract or an upcoming promotion. The impact of the Career arm on teachers who have a permanent contract is a 0.40σ increase in training test scores. In contrast, the impact of the Career arm on teachers who have a temporary contract is close to zero and insignificant (p-value difference of 0.01). Similarly, the impact of the Career arm on teachers whose promotion is due sooner (i.e., within the next 5 years) is a 0.42σ increase in training test scores, whereas the impact of the Career arm on teachers whose promotion is due after 5 years is positive but insignificant (p-value difference not significant). This provides supporting evidence that the effects in the Career arm are driven by career incentives.

Next, we turn towards explaining the treatment effects of the motivational framing. We show that the net impact of adding the motivational framing across the Peer and Career arms is to increase

our teacher motivation index by 0.08σ. This highlights that the framing in the PLUS treatments did bolster teacher morale and motivation as expected. Further analysis of mechanisms highlights that the motivational framing backfires because it increases teacher overconfidence by about 6 percentage points which results in reduced effort in the training. Mediation analysis (following guidelines as per Acharya et al., 2016) shows that overconfidence can explain up to 85% of the observed negative effects of the motivational framing on training test scores.

Taken together, these results highlight that it is possible to improve knowledge acquisition and hence teacher quality through introducing cost-effective and easy to implement recognition rewards in teacher trainings. However, at the same time the framing of these rewards is non-trivial and can have significant effects.

This paper provides, to the best of our knowledge, novel empirical evidence on the impact of recognition rewards in teacher trainings in a public sector setting. Our results show that cost effective recognition rewards that make career benefits salient can improve teacher trainings. Our findings are in-line with existing literature that highlights that linking teacher participation in trainings to incentives such as status, promotion, or salary can make teacher training more effective (Popova et al., 2021). Our results add to the literature on teacher trainings more specifically (Piper and Korda, 2011; Loyalka et al., 2019; Cilliers et al.,2020; Kerwin and Thornton, 2021), and to other sectors such as health and police more broadly where trainings are commonly used to upgrade skills and knowledge of employees (for example, see Bluestone et al. (2013) for health,Banerjee et al. (2021) for police departments, and Azulai et al. (2020) for civil service).

Although our results highlight that recognition rewards can be effective, they also highlight that their framing is non-trivial and can have significant effects. In particular, when a recognition reward is combined with a motivational framing, it can backfire through an overconfidence channel. While the details of the incentive in a scheme (e.g., whether it is financial or non-financial and the precise rules for its qualification) are often carefully documented across studies, the precise way in which it is possibly administered (e.g., motivational, controlling, or neutral manner) is not always documented. This is non-trivial given the manner of administration could potentially vary across treatment units driving varying effects. While existing evidence on framing of rewards mainly focuses on what happens to employee performance when rewards are framed as losses or gains (Goldsmith and Dhar, 2013; Lagarde and Blaauw, 2021), our results contribute to this evidence by highlighting that whether a reward has a motivational framing or not could matter for its overall impact. This suggests caution in how such rewards are framed and administered

in the field.

Finally, this paper provides experimental evidence on the peer-esteem and career benefits channels of recognition rewards, and adds to the wider empirical evidence on non-financial incentives in the public sector (Ashraf et al., 2014a; Gauri et al., 2018; Khan et al., 2019; Ashraf et al.; 2014b; Cotofan, 2021). While existing evidence documents the impact of public versus private recognition (as in Ashraf et al., 2014a) or recognition conferred by the community versus employer (as in Gauri et al., 2018), this study unbundles the peer-esteem versus the career benefits channels of employer recognition. While this study explores these channels within the specific context of teacher training, the results highlight that these channels are distinct, which has implications for further research in the context of teachers in the classroom or other service delivery contexts.

This paper is organized as follows. Section 2 outlines the contextual setting and theory underlying the study. Section 3 describes the experimental design, randomization, and data sources. Section 4 presents the empirical strategy and main results, and Section 5 presents mechanisms for understanding the impact on our main outcome. Section 6 offers concluding remarks.

2 Setting

2.1 Punjab Education Sector

Punjab is Pakistan's largest province with 36 districts and a population of 110 million.[8] The public education system employs a workforce of approximately 450,000 teachers responsible for educating nearly 11 million children spread across 52,000 schools.[9] The School Education Department (SED) is the provincial public body responsible for policy implementation pertaining to primary and secondary education.[10]

Progress in learning outcomes has remained slow in Punjab. For example, the ASER (2019) report showed that nearly 40% of children in grade 5 have not reached grade 2 levels of learning

[8] Pakistan Population Census, 2017 (Pakistan Bureau of Statistics).

[9] Annual School Census Data 2017.

[10] Schools are further divided into primary (grades 1-5), elementary (grades 6-8), secondary (grades 9-10), and higher secondary (grades 11-12) schools.

in literacy and numeracy (this includes English, Math and the national language Urdu). The ASER (2021) report shows that these learning levels have likely declined post covid-19. Low levels of teacher quality and effort is widely perceived as one of the main reasons for low levels of student learning. This is also confirmed by descriptive evidence from a survey of 650 secondary school which showed that 40% of surveyed teachers scored less than 50% on grade 5 math problems (Mansoor, 2018).

2.2 Quaid-e-Azam Academy for Educational Development (QAED)

The Quaid-e-Azam Academy for Educational Development (QAED) is an attached department of the Punjab School Education Department that holds the mandate to provide on-the-job training to all public school teachers in Punjab. The academy offers a range of training programs such as standard promotion-linked trainings, as well as those focused on content (e.g., literacy or numeracy, and other subject-based trainings), pedagogy, and leadership. Apart from promotion-linked trainings, none of the programmes include any incentives linked to effort or performance in the training.

The recognition program in this study was embedded within such a non-promotion linked training at QAED called the 'Student Leadership Development Program' (SLDP). The training spanned over four days and was targeted at school head teachers across Punjab to improve management and leadership of head teachers.[11] The training was first provided to a selected pool of 634 master trainers, after which 500 master trainers were selected for cascading the trainings further down to the head teachers.

The training was organized and implemented at the district level at the relevant district training center. Given the high number of head teachers in each district (i.e., between 300-800 in one district), training sessions were organised across 4 rounds, where each round had 4 sessions operating simultaneously with about 30 teachers per session. The process of assigning trainees to these sessions was done by the QAED head quarters randomly such that each session had equal representation of rural and urban school head teachers. Each training session also included a training pre-test and post-test to measure knowledge acquisition from the training.

[11] The training was a specialized curriculum for providing skills in coaching, leadership, and school management. Training modules included the following: 1) The power of coaching, 2) Co-curricular activities , 3) Protecting children, 4) Student leadership, 5) Staff and distributed leadership, 6) Leave rules, and 7) Pupil voice.

2.3 Theory

Extrinsic utility from recognition: Despite existing evidence on recognition incentives (Ashraf et al., 2014a; Gauri et al., 2018; Ashraf et al.; 2014b; Cotofan, 2021), there is limited understanding of how agents weight possible sources of extrinsic motivation from recognition in a standard utility maximizing framework. We distinguish that agent extrinsic utility from recognition that is conferred by their employers could come from either peer-esteem or potential career benefits (Besley and Ghatak, 2008; Frank, 1985).

While effort in response to the peer-esteem channel could depend on the number of peers known, respect for peers, or perceived quality of peers, effort in response to the career benefits channel could depend on factors that increase the likelihood of accessing career benefits in the system such as having a permanent contract or an upcoming promotion. Note that while formal career incentives for public school teachers in our context are limited since promotions are linked to seniority, three types of informal career incentives might be relevant for how teachers can use the recognition reward to their advantage. First, teachers may want to be posted to better performing schools as opposed to poor performing schools. Second, once teachers become eligible for promotion they may want to be selected for promotion before other competing colleagues.[12] Third, teachers may have preferences to be posted laterally to positions with higher grade whilst having the same pay and grade. These informal career incentives mostly relate to transfers and postings which can be a sharp incentive as shown in Khan et al. (2019).

If the motivations for peer-esteem or career benefits from recognition are indeed distinct and agents put different weights over their extrinsic payoff from each channel, the impact of the recognition reward could vary depending on the margin of effort that is leveraged in the design of the reward.

Intrinsic utility from recognition: Frey (1997) and Deci et al. (1999) argue that rewards can be administered in a motivational or controlling way, where the former can harness intrinsic motivation but the latter can dampen it. For example, in situations where recognition rewards are framed in a way that bolsters individual self-esteem and confidence, intrinsic payoffs to effort should increase. However, if the information is perceived as controlling that dampens morale,

[12]In our context, employees who are eligible for promotion have to wait for their turn to get their promotion approved.

it would decrease intrinsic payoffs to effort.[13] Hence, when agents decide to exert effort, the decision depends on the interaction between the reward and the *information* contained within the reward (Gneezy et al., 2011). This implies that the impact of the recognition reward on equilibrium effort could vary depending on the framing of the rewards.

Design: Our experiment tests the strength of the peer-esteem channel versus the career benefits channel of recognition. Given the principal can shape the information in rewards through framing, we combine these incentives with a motivational framing to understand whether that improves the design of our recognition incentives.[14]

3 Experimental Design

3.1 Treatment Arms

The experimental design includes four different treatment arms of recognition rewards, and a control group. The design of the recognition reward is a standard tournament-based incentive tied to training test scores which serves as a proxy for teacher knowledge and skill acquisition. Within a training session, teachers who score the highest in the training post-test score or show the maximum improvement over the training pre-test score qualify for a prestigious certificate that is authenticated by the QAED head quarters. This design encourages teacher effort across the entire distribution of trainees' ability in the training session instead of only high ability teachers (as in Ashraf et al., 2014a).

The sequencing of activities over the four training days is as follows. On the first day, teachers take the training pre-test after which enumerators administer the relevant recognition incentive following a predetermined script.[15] This is followed by the scheduled training over the next four days. On the fourth and final training day, teachers take a training post-test at the end of the training. The winning teachers receive their recognition rewards nearly two months after the training. The details of each treatment arm are given below.

[13]Note that this makes intrinsic payoffs effort dependent, as compared to *warm glow* (as in Andreoni, 1990) where individuals intrinsic utility is non-effort dependent.

[14]We outline these details in a conceptual framework that is presented in Supplementary Online Appendix B.

[15]The detailed scripts are presented in the Supplementary Online Appendix C.

Control group: Teachers in this group are administered a neutral script by the enumerator which highlights the broad goals of the SLDP training. All other activities such as the training lectures, pre-test, and post-test operate as in all the other groups.

Peer Recognition (T1): Teachers in this treatment group are informed that if they meet the required qualification conditions, they would be eligible for receiving a prestigious recognition certificate in a public district ceremony which would be attended by their peers and colleagues in the district office. The script for T1 is exactly the same as the control group except for the additional information about the recognition incentive. All other activities such as the training lectures, pre-test, and post-test operate as in all the other groups. This treatment leverages the motivation for peer-esteem.

Career-based Recognition (T2): Teachers in this treatment group are informed that if they meet the required qualification conditions, they would be eligible for receiving a prestigious recognition certificate which would be given to them *privately*. In addition, they are also informed that their names would be included in an 'excellent teacher list' which would be shared with their district's leadership which could make them eligible for future career opportunities in the department. The script for T2 is exactly the same as T1 except for the difference in how career benefits as opposed to peer-esteem is made salient. All other activities such as the training lectures, pre-test, and post-test operate as in all the other groups. This treatment leverages the motivation for reaping potential career benefits through recognition.

Peer PLUS (T3) and Career PLUS (T4): Teachers in Peer PLUS and Career PLUS are administered the same script as T1 and T2 respectively. However, the administration is framed with a *motivational framing* to boost individual morale and beliefs in ability to do well in the training and the job more broadly. The motivational framing aims to improve the way information contained in the reward is perceived by the head teachers. More details on the framing are provided below.

Motivational Framing: The goal of this framing is to improve the information contained in the reward by bolstering teacher morale and beliefs in their capabilities.

The framing is structured as follows: teachers are first asked to reflect upon three key limitations and challenges in performing well in the training and their jobs more broadly. This is followed by the distribution of a one-pager with three inspirational stories of head teachers from Punjab that the trainees are asked to read. The stories are meant to serve as role models to bolster existing

levels of belief in one's capability and ability (as in Beaman et al., 2012 and Tanguy et al., 2014 for example). To create a final moment of reflection, trainees are asked to reflect on how they can address their own limitations (as identified in the first step) after reading the stories.[16]

3.2 Randomization

While the SLDP training was implemented across all 36 districts of Punjab, the recognition programme was rolled out in 7 districts spread across the north, south, and central regions of the province (See Figure 1). Training sessions in each district were assigned a *session number*. Stratifying by district, a total of 131 training sessions were randomly allocated to four different treatments and the control group. This yielded a sample of 3,394 head teachers across 131 training sessions in 7 districts of Punjab. Descriptive statistics in Table A.1 show that our sample is 57% female, with an average teacher age of 46 years with around 20 years of experience in the education department.

Figure 1 about here

3.3 Data and Balance Checks

3.3.1 Data

Teacher Training Test Score Data. Our primary outcome of interest is teacher training test scores. Both the pre and post-tests were developed by the SLDP staff at the QAED headquarters. The tests included a total set of 15 MCQ questions that were directly related to the taught content.[17] Our baseline pre-test score in Table A.1 shows that head teachers scored 34% on average with very few teachers subject to ceiling or floor effects.

[16] To design this framing, we draw on the seminal work of Bandura (1986), who defines the concept of *self-efficacy* as the "perception of one's capability to accomplish a given level of performance" as central to motivation and performance, and highlights the strength of role models and vicarious experiences in boosting self-efficacy.

[17] Given the training in each district had multiple rounds, the pre and post-test questions were different across rounds (although tested the same learning objectives) to reduce chances of gaming.

Teacher Surveys at endline. To understand mechanisms, we collect data on a selected set of measures of head teacher attitudes and perceptions that the may have been shaped by the treatments. These include teacher intrinsic motivation, self-efficacy, locus of control, and beliefs about performance on the training post-test.

Teacher Surveys at baseline. We also capture a range of variables in our baseline survey to study heterogeneous treatment effects. These include basic teacher characteristics such as age, gender, salary, and years of experience; non-cognitive traits and beliefs such as intrinsic motivation, pro-social motivation, self-efficacy, and locus of control; and training and work-related variables to validate the theory underlying our treatment arms such as number of peers known in training session, time till next expected promotion, contract type, and visibility to senior leadership.[18]

Enumerator and Master Trainer Data. We also collect data on enumerator characteristics such as age, years of experience, years of education, and communication skills to be able to control for enumerator effects in our estimation. In addition, we also collect on a range of master trainer characteristics such as age, years of experience, and number of trainings received. These are also used as controls in our analysis.

3.3.2 Balance Tests and Implementation

Table A.2 shows balance across treatment arms for four different categories of variables: basic teacher characteristics, job characteristics, training baseline test score, and teacher non-cognitive traits. We conduct tests of equality for each variable across all treatment groups. Our training baseline test score is balanced. Out of a total of 105 tests, 8 are different from zero at the 5% level. We account for this by adding controls in our analysis. We also conduct joint F tests across all groups. All p-values for the joint test are greater than or equal to 0.14.

Attrition was not a serious concern in our study given the trainings were mandatory for head teachers to attend. However, there is small attrition in our sample (3%) due to teachers being absent on the fourth day of the training when the post-test took place. Table A.3 shows that

[18] All measures are based on existing validated scales. We measure intrinsic motivation using validated scales as in Amabile et al. (1995), self-efficacy using Fackler and Malmberg (2016), and pro-social motivation through the Perry PSM index (Perry, 1996). We also conduct tests of internal reliability for these scales using Cronbah's alpha (Cronbach, 1951) that are presented in Table A.9 in the appendix. All measures (except the psm index) have a high internal consistency of α>0.60.

attrition is not related to any of our treatment groups and Table A.4 shows that the attrited and main sample are balanced across teacher characteristics at baseline.

Where spillovers are concerned, these are unlikely in our setting. The treatment is at the training session level and there is minimal interaction between sessions during the day as trainings are conducted from 8:00 am to 2:00 pm every day within specific training classrooms.

4 Empirical Strategy and Main Effects

4.1 Empirical Strategy

To identify the main treatment effects of our interventions on training test scores, we estimate the following:

$$y_{isd}Post = \beta_0 + \rho.y_{isd}Pre + \sum_{j=1}^{4} \beta_j T_{ij} + \gamma X_{isd} + \mu_d + \alpha_r + \epsilon_{isd} \quad (1)$$

Where y_{isd}^{Post} is the post-test score for teacher i in training session s, and district d; y_{isd}^{Pre} is the pre-test score that serves as our baseline measure for an ANCOVA estimation. The post-test and pre-test scores are normalized by the mean and standard deviation of the pre-test scores in the control group. Hence, the treatment effects are observed in standard deviations units. X_{isd} is a vector of teacher, master trainer, and enumerator controls that we include in our estimation for power. These are selected through the LASSO post double selection procedure following Belloni et al. (2014).[19] We also control for training round effects, α_r, by adding round dummies and include district fixed effects as captured by μ_d. Finally, errors are clustered at the training session level which is our unit of randomization (as suggested by Abadie et al., 2017). The β coefficients are the coefficients of interest.

We use Intention to Treat (ITT) to estimate our treatment effects. A small proportion of teachers

[19]Note that our results are also robust to an alternate selection of self-selected controls. These robustness checks are included in the Supplementary Online Appendix A.

(6%) refused to participate in the recognition scheme.[20] Table A.5 shows that non-consent is not significantly related to any of the treatment groups.

4.2 Main Effects

Treatment effects of Peer (T1) and Career (T2) on training test scores. Table 1 shows the treatment effects on standardized training test scores. We first focus on the treatment effects in the Peer and Career arms as shown in Column 2.[21] We find that teachers in the Career arm score 0.33σ higher training test scores as compared to the control group (significant at the 10% level). In comparison, the Peer arm has a coefficient of 0.05 and is insignificant. Kolmogorov–Smirnov tests of equality of distribution between the Peer and Career arm confirm that the distributions are significantly different from each other at the 1% level.[22] Back of the envelop calculations suggest that the knowledge acquisition in the Career arm is equivalent to about half a day of additional training. Given the training was spread across 4 days, this indicates that the Career arm could improve cost effectiveness by approximately 12.5%.

Given the Career arm made potential career benefits salient, these results point towards the value of informal career benefits in the system that the teachers could have accessed through the recognition certificate (such as getting transfers to preferred schools, getting selected for promotions faster once eligible, or getting appointed to higher grade positions on the same salary scale and grade if positions become vacant). Discussions with our main counterparts at QAED suggest that the strength of the peer-esteem channel may be weak for head teachers who have already risen through the ranks and established respect, reputation, and esteem amongst their peers and colleagues. Hence, such a channel may be more effective for primary and secondary school teachers who are younger and looking to establish their reputation amongst their peers. However, where the career benefits channel is concerned, head teachers tend to have strong informal career incentives in the system such as postings to their choice of school or other influential lateral appointments.

Treatment effects of Peer PLUS (T3) and Career PLUS (T4) on training test scores. Next, we focus on the treatment effects of the PLUS interventions with the motivational framing. Column

[20] This included 207 teachers which is roughly 6% of the sample.

[21] We focus on the estimations with controls given we had imbalance on a few teacher characteristics.

[22] The distributions are presented in Appendix B.6

2 in Table 1 shows that the net impact of adding the motivational framing in Peer PLUS and Career PLUS is negative – a reduction of 0.21σ in Peer PLUS (not significant) and a reduction of 0.36σ in Career PLUS (significant at the 5% level). Since the net impact of adding the framing moves in a negative direction for both arms, we pool the PLUS treatments (i.e. those receiving the framing) and the non-PLUS treatments (i.e. those not receiving the framing) in Columns 3 and 4. Column 4 shows that the net impact of adding the motivational framing to either of the arms is to lower training test scores by 0.28σ (significant at the 5% level).[23]

This negative effect implies that adding the motivational framing to our recognition incentives resulted in teachers reducing effort in the training, which runs counter to our hypothesis in subsection 2.3. While our findings are inconsistent with the positive effects of similar motivational interventions in other contexts such as job search and health-seeking behaviours (see Eden and Aviram, 1993; Haushofer, John, and Orkin, 2019), they are in line with arguments of skeptics who suggest that creating "positive illusions" of oneself can often have negative effects by leading to overconfidence (Baumeister, 1999; Swann, 1996). Overall, this suggests caution in how rewards are framed and administered.

Table 1 about here

5 Mechanisms

Our main treatment effects highlight two main sets of results: first, the Career arm appears to improve teacher knowledge as measured by training test scores. Second, the motivational framing has negative effects on training test scores suggesting that teachers may have reduced effort in the training. In this section, we follow these two sets of results one by one to understand mechanisms underlying these effects.

[23]Note that this interpretation of net impact assumes that the PLUS treatments are a linear combination of the recognition incentive and the motivational framing. However, if the two treatments interact together in non-linear ways our point estimates of the net impact of the framing would not be accurate. Irrespective of this assumption, our results highlight that the combination treatment with the framing does worse than the recognition incentive alone.

5.1 Mechanisms: Unpacking Treatment Effects in Career arm

Quantile Treatment Effects. We first study treatment effects in the Career arm across the distribution of training test scores to understand whether the effects are driven by the lower or upper tail of the distribution of training test scores. This is especially important in the context of tournament-based rewards (such as our recognition incentive) which often merely elicit effort from agents in the upper tail of the ability distribution (Connelly et al., 2014). In addition, this allows us to compare the distributional impacts of the Career arm versus the Peer arm.

We estimate quantile treatment effects for this analysis. Table 2 shows quantile treatment effects at quantile $\tau \in$ (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9). It shows that the Career arm has a positive coefficient in the range of 0.11 - 0.49 across the distribution of training test scores and the coefficient is always higher than the Peer arm. The coefficients of the Career arm are significant at the 5% level in the upper tail of the distribution, where they are also significantly different from the Peer arm (see Figure B.1 in the appendix). These trends confirm that the career benefits channel encourages effort across the distribution of training test scores, instead of merely high ability individuals. It also provides further evidence that the career benefits channel is likely to be different from the peer-esteem channel.

Table 2 about here

Heterogeneity by strength of career concerns. Next, we test whether the Career arm works as hypothesized in theory, i.e. whether agents respond to the recognition reward because they believe it could result in tangible career benefits in the future. To investigate this, we hypothesize that the Career arm should work better for teachers who have stronger career concerns.

We identify three categories of variables that capture strength of teacher career concerns: teachers who are permanent employees, have an upcoming promotion, or have higher visibility to senior leadership. It is intuitive that permanent teachers would value a recognition reward for career progression more so than temporary teachers. Nearly 17% of our sample includes head teachers who are working on a contractual basis and do not have the same career incentives as permanent teachers. We identify teachers who are on permanent contracts as *permanent* employees. Second, given teachers in our context have informal career incentives (e.g., getting posted to a school of their liking or to other lateral appointments), an upcoming promotion can make these incentives

more salient. In our sample, the median time to next promotion is 5 years. We identify teachers as more *promotion eligible* if they have an upcoming promotion within the next 5 years. Finally, higher visibility to senior leadership can increase opportunities to reap informal career incentives in the system. We identify teachers where frequency of visits by senior leadership (such as Secretary Education and District Education Officers) to the teachers' districts is at least once in 3 months as more *visible to leadership*.

Table 3 presents our results. Column 1 shows that the impact of the Career arm on teachers whose promotion is due sooner (i.e., within the next 5 years) is a 0.42σ increase in training test scores, whereas the impact of the Career arm on teachers whose promotion is due after 5 years is positive but insignificant (p-value difference not significant). Similarly, Column 2 shows that the impact of the Career arm on teachers who have frequent visits by the Secretary is a 0.52σ increase in training test scores, whereas the impact of the Career arm on teachers who receive less frequent visits is insignificant (p-value difference not significant). Finally, Column 3 shows that impact of the Career arm on teachers who have a permanent contract is a 0.40σ increase in training test scores. In contrast, the impact of the Career arm on teachers who have a temporary contract is close to zero and insignificant (p-value difference of 0.01). Overall, these results provide supporting evidence that when the Career arm works, the treatment effect comes from individuals who have stronger career concerns.[24]

Table 3 about here

5.2 Mechanisms: Unpacking Treatment effects in PLUS arms

Treatment effects on teacher motivation. The main assumption behind the design of the PLUS treatments is that the motivational framing should improve the design of the recognition rewards by boosting individual morale and beliefs about ability.

We investigate the impact of the PLUS treatments on different dimensions of teacher motivation such as intrinsic motivation (Deci and Ryan, 1985), self-efficacy (Bandura, 1986), and locus

[24]We also hypothesize that the Peer arm should work better if teachers know their peers well in the training session. This rests on the assumption that the *peer-esteem* from the Peer arm would be stronger if a teacher knows his/her peers. For completeness, we also test this theory and explore heterogeneous treatment effects by the number of peers each trainee knows in their group. We find no significant effects (See Columns 4 and 5).

of control (Rotter, 1966). We measure these using pre-existing validated scales and normalize them by the mean and standard deviation of the control group. To avoid challenges of multiple hypothesis testing, we develop an overall index of teacher motivation as an average of these three standardized measures. We also measure test specific teacher self-efficacy (i.e. teacher beliefs about their ability to perform well in the test) by asking teachers how much they believed they scored on the post-test on a scale of 1-100. To estimate our treatment effects on the motivation index and teacher beliefs about post-test performance, we run the same specification as (1) but with the teacher motivation index or beliefs about post-test performance as the outcome measure.

Table 4 presents our results. Column 1 shows that net impact of the framing in the Career PLUS arm is to increase teacher motivation 0.12σ. The net impact of the framing in the Peer PLUS arm is positive but insignificant. Column 2 shows that the net impact of the framing across both the PLUS arms is to increase teacher motivation by 0.08σ. In comparison, we do not find consistent evidence on teacher test-specific self-efficacy. These results confirm that while the motivational framing reduced training test scores, it did boost teacher motivation.[25]

Table 4 about here

5.3 Why did teachers reduce effort in the PLUS arms?

In this subsection we investigate the mechanisms underlying the negative impact of the motivational framing on training test scores. We hypothesize that while the framing improved teacher motivation, it could have simultaneously made teachers overconfident in their ability to do well in the training which could have led to a reduction in teacher effort (and ultimately training test scores). This explanation is consistent with skeptics who argue that interventions that aim to improve individual self-esteem or efficacy can at times *over-correct* beliefs about ability leading to dangers of overconfidence (Swann, 1996; Baumeister, 1999; Bénabou and Tirole, 2002).

Treatment effects on Overconfidence. In our endline survey, we ask teachers to report how much they believe they scored on the training post-test. This allows us to construct a direct measure

[25]Table A.6 shows these results by each dimension of teacher motivation - intrinsic motivation, self-efficacy, and locus of control - and confirms positive and significant effects of the PLUS arms on teacher self-efficacy, as well as positive coefficients on internal locus of control.

of teacher overconfidence as the difference between beliefs about performance and actual performance on the post-test. Typically, measures of overconfidence across economics and psychology are constructed by asking respondents a set of questions, along with their rate of confidence in the answers to each question. Overconfidence is then measured as the *positive bias*, when difference between average confidence level and the proportion of correct answers is greater than zero (Adams, 1957; Michailova, 2010). Since the fundamental idea in measuring overconfidence is to observe individual judgement compared to a gold standard of truth (Baumann et al., 1991), our measure of overconfidence is an example of a direct measure of overconfidence and similar to measures used by others such as Glaser et al. (2005).

To investigate treatment effects on overconfidence, we run the same specification as (1) but use teacher overconfidence as our outcome measure. Table 5 presents our results. Column 1 shows that the net impact of the motivational framing on overconfidence is positive and significant, making teachers 6.8% points more overconfident in Peer PLUS and 5.3% points more overconfident in Career PLUS. Column 2 shows that the net impact of adding the motivational framing across both the PLUS arms is to make teachers 6% points more overconfident. As an additional robustness check, we re-define our overconfidence measure as above and below median overconfidence to address potential outliers and repeat the estimation of our treatment effects in Columns 3 and 4. We find the same lines of results.

Table 5 about here

Mediation Analysis. We use mediation analysis to quantify the strength of the overconfidence channel in explaining the negative effects of the motivational framing.

We use the procedure of sequential g-estimation as laid out in Acharya et al. (2016) to identify the Average Controlled Direct Effect (ACDE) of the net impact of the motivational framing after accounting for the effects of overconfidence. This is based on several identifying assumptions that are detailed in our mediation analysis methodology in the Supplementary Online Appendix D). Table 6 presents our results, with the the original estimation in Column 1 (as in Table 1, Column 4) and the revised estimation based on the de-mediated outcome in Column 3. While the net impact of the motivational framing is -0.28σ (significant at the 5% level) in Column 1, the ACDE in Column 3 reduces to -0.04 (insignificant). This implies that overconfidence approximately explains up to 86% of the observed negative treatment effects of the PLUS arms.

Table 6 about here

6 Concluding Remarks

We present experimental evidence on the impact of recognition rewards on teacher knowledge and skill acquisition in a public teacher training programme in Punjab, Pakistan. The study shows that recognition can improve teacher knowledge if the rewards are linked to tangible career benefits, and that these effects are stronger for teachers who have stronger career concerns. At the same time, we show that the framing of these rewards is nontrivial. In particular, adding a motivational framing to the recognition treatment "over corrects" teacher beliefs about ability to do well in the training leading to overconfidence and reduced effort.

Our results have two key implications. First, they highlight that non-financial incentives can be a cost-effective way to improve teacher professional development programmes. The career-based recognition treatment used in this experiment had a negligible cost and was easy to embed within the existing training. A key feature of the design of the treatment was the identification of informal career incentives in the system and how they could be made salient. This opens up a discussion on how the public sector can leverage such motivators within the system and design "soft" non-financial incentives based on them to improve teacher trainings more specifically, as well as trainings of public sector employees and their performance in frontline service delivery contexts more broadly. Second, our results highlight the sensitivity of recognition rewards to framing effects. This indicates caution in how incentives are designed and framed, as well as in how evidence on recognition rewards is interpreted across different contexts.

Several additional questions remain open to inquiry. First, our experiment was only able to offer the recognition reward for a single time. Future work could look at the decay rate in the impact of such rewards, and circumstances under which the effects are sustained. Second, this paper only focused on the first step of measuring the impact of incentives on teacher professional development - teacher knowledge as measured by training test scores. However, measuring the impact of incentives could also involve measuring downstream outcomes such as the intended application of training content to on-the-job teacher practices and student outcomes. Future research on the extent to which incentives in trainings can encourage such downstream implementation, and whether certain types of incentives are more effective than others in achieving this would be

useful. Finally, our experiment shows that creating exogenous variation in motivation of public sector employees is possible (as also shown by Khan, 2020). This opens up the possibility of additional research on how to create and measure the impact of exogenous variation in intrinsic motivation on workplace identities, norms, culture, and performance.

Acknowledgements: This research was made possible through numerous insights from partners at the Punjab Education Department. In particular, Mr. Waseem Shirazi, Mr. Nadeem Hussain, and Mr. Adnan Bashier at the Quaid-e-Azam Academy for Educational Development offered many helpful reflections. I thank Julien Labonne, Stefan Dercon, Clare Leaver, Christopher Woodruff, Simon Quinn, Michael Callen, Adnan Qadir, Martin Williams, and participants at CSAE workshops (2019, 2020), Nordic Development Conference (2021), Applied Development Economics Conference (2021), and OxDEV (2021) for helpful comments. I gratefully acknowledge funding from Economic Development Institutions (EDI), Center for the Study of African Economies (CSAE), and the Blavatnik School of Government (BSG). Usman Zahid offered excellent research support. The study is pre-registered at the AEA RCT registry (0004482). IRB approval was obtained at the University of Oxford. All errors are my own.
∗ Lecturer in Public Policy, School for Policy Studies, University of Bristol

Declaration of Competing Interest: The author declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge. 2017. *When Should You Adjust Standard Errors for Clustering? [electronic resource].* Working paper series (National Bureau of Economic Research : Online) ; working paper no.24003. Cambridge, Mass.: National Bureau of Economic Research.

Acharya, Avidit et al. 2016. "Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects." 110 (3):512–529.

Adams, J K. 1957. "A confidence scale defined in terms of expected percentages." *The American journal of psychology* 70 (3):432.

Amabile, Teresa M., Karl G. Hil, Beth A. Hennessey, and Elizabeth M. Tighe. 1995. ""The Work Preference Inventory: Assessing intrinsic and extrinsic motivational orientations": Correction." *Journal of Personality and Social Psychology* 68 (4):580–580.

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *The Economic Journal* 100 (401):462. URL `http://search.proquest.com/docview/198090143/`.

Araujo, M. Caridad, Pedro Carneiro, YyannÃº Cruz-Aguayo, and Norbert Schady. 2016. "Teacher Quality and Learning Outcomes in Kindergarten." *The Quarterly Journal Of Economics* 131 (3):pp1415–1453.

ASER. 2019. "Annual Status of Education Report - ASER-PAKISTAN 2019." URL `https://palnetwork.org/wp-content/uploads/2020/02/Annual-Status-of-Education-Report-ASER-PAKISTAN-2019.pdf`.

———. 2021. "Annual Status of Education Report - ASER-PAKISTAN 2021." URL `http://aserpakistan.org/document/aser/2021/reports/national/ASER_report_National_2021.pdf`.

Ashraf, Nava et al. 2014a. "Awards unbundled: evidence from a natural field experiment." *Journal of economic behavior and organization* 100:44–63. URL `http://search.proquest.com/docview/1523807233/`.

———. 2014b. "No margin, no mission? A field experiment on incentives for public service delivery." *Journal of public economics* 120:1–17.

Azulai, Michel, Imran Rasul, Daniel Rogger, and Martin Williams. 2020. "Can Training Improve Organisational Culture? Experimental Evidence from Ghana's Civil Service." .

Bandura, Albert. 1986. *Social foundations of thought and action : a social cognitive theory.* Prentice-Hall series in social learning theory. Englewood Cliffs, N.J.: Prentice-Hall.

Banerjee, Abhijit, Raghabendra Chattopadhyay, Esther Duflo, Daniel Keniston, and Nina Singh. 2021. "Improving Police Performance in Rajasthan, India: Experimental Evidence on Incentives, Managerial Autonomy, and Training." *American economic journal. Economic policy* 13 (1):36–66.

Bau, Natalie and Jishnu Das. 2020. "Teacher Value Added in a Low-Income Country." *American economic journal. Economic policy* 12 (1):62–96.

Baumann, Andrea O et al. 1991. "Overconfidence among physicians and nurses: The 'micro-certainty, macro-uncertainty' phenomenon." *Social science medicine (1982)* 32 (2):167–174.

Baumeister, Roy F. 1999. *The self in social psychology.* Key readings in social psychology. Philadelphia, PA ; Hove: Psychology Press.

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova. 2012. "Female leadership raises aspirations and educational attainment for girls: a policy experiment in India." *Science (New York, N.Y.)* 335 (6068):582.

Belloni, Alexandre et al. 2014. "Pivotal Estimation Via Square-Root LASSO in Non-Parametric Regression." *The Annals of Statistics* 42 (2):757–788.

Berlinski, Samuel and Matias Busso. 2017. "Challenges in educational reform: An experiment on active learning in mathematics." *Economics letters* 156:172–175.

Besley, Timothy, Robin Burgess, Adnan Khan, and Guo Xu. 2022. "Bureaucracy and Development." *Annual review of economics* 14 (1):397–424.

Besley, Timothy and Maitreesh Ghatak. 2005. "Competition and Incentives with Motivated Agents." *American Economic Review* 95 (3):616–636.

———. 2008. "Status Incentives." *American Economic Review* 98 (2):206–211.

Bloom, Nick, Renata Lemos, Raffaella Sadun, and John Van Reenen. 2015. "Does management matter in schools?" *The Economic journal (London)* 125 (584):647–674.

Bluestone, Julia, Peter Johnson, Judith Fullerton, Catherine Carr, Jessica Alderman, and James BonTempo. 2013. "Effective in-service training design and delivery: evidence from an integrative literature review." *Human resources for health* 11 (1):51–51.

Bold, Tessa, Deon Filmer, Gayle Martin, Ezequiel Molina, Christophe Rockmore, Brian Stacy, Jakob Svensson, and Waly Wane. 2017. "What Do Teachers Know and Do? Does It Matter? : Evidence from Primary Schools in Africa."

Bruns, Barbara and Javier Luque. 2015. *Great teachers: How to raise student learning in Latin America and the Caribbean.* Latin American development forum series. Washington, DC: World Bank Group.

Bénabou, Roland and Jean Tirole. 2002. "Self-Confidence and Personal Motivation." *The Quarterly Journal of Economics* 117 (3):871–915.

———. 2003. "Intrinsic and Extrinsic Motivation." *Review of Economic Studies* 70 (3):489–520.

Chetty, Nadarajan et al. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." 104 (9).

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014. "Measuring the impacts of teachers, [part] II: teacher value-added and student outcomes in adulthood." *The American economic review* 104 (9):2633–2679.

Cilliers, Jacobus, Brahm Fleisch, Cas Prinsloo, and Stephen Taylor. 2020. "How to Improve Teaching Practice?: An Experimental Comparison of Centralized Training and In-Classroom Coaching." *The Journal of human resources* 55 (3):926–962.

Connelly, Brian L et al. 2014. "Tournament Theory: Thirty Years of Contests and Competitions." *Journal of management* 40 (1):16–47.

Cotofan, Maria. 2021. "Learning from praise: Evidence from a field experiment with teachers." *Journal of public economics* 204:104540.

Cronbach, Lee J. 1951. "Coefficient alpha and the internal structure of tests." *Psychometrika* 16 (3):297–334.

Deci, Edward et al. 1999. "A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation." *Psychological Bulletin* 125 (6):627–668. URL `http://search.proquest.com/docview/203472909/`.

Deci, Edward L. and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior [electronic resource]*. Perspectives in social psychology. New York.

Dewatripont, Mathias et al. 1999. "The economics of career concerns, Part I: comparing information structures." *Review of economic studies* 66(1) (226):183–198. URL `http://search.proquest.com/docview/38702024/`.

Eden, Dov and Arie Aviram. 1993. "Self-Efficacy Training to Speed Reemployment: Helping People to Help Themselves." *Journal of Applied Psychology* 78 (3):352. URL `http://search.proquest.com/docview/1290321664/`.

Fackler, Sina and Lars-Erik Malmberg. 2016. "Teachers' self-efficacy in 14 OECD countries: Teacher, student group, school and leadership effects." *Teaching and Teacher Education* 56.

Frank, Robert H. 1985. *Choosing the right pond : human behavior and the quest for status*. New York ; Oxford: Oxford University Press.

Frey, Bruno S. 1997. *Not just for the money : an economic theory of personal motivation*. Cheltenham: Edward Elgar.

Fryer, Roland G. 2016. "The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments." *NBER Working Paper Series* :22130.

Gauri, Varun, Julian Jamison, Nina Mazar, Owen Ozier, Shomikho Raha, and Karima Saleh. 2018. "Motivating Bureaucrats Through Social Recognition: Evidence from Simultaneous Field Experiments." URL `http://search.proquest.com/docview/2063163854/`.

Glaser, Markus et al. 2005. "Overconfidence of professionals and lay men : individual differences within and between tasks?"

Gneezy, Uri et al. 2011. "When and Why Incentives (Don't) Work to Modify Behavior." *Journal of Economic Perspectives* 25 (4):191–210.

Goldsmith, Kelly and Ravi Dhar. 2013. "Negativity Bias and Task Motivation: Testing the Effectiveness of Positively Versus Negatively Framed Incentives." *Journal of experimental psychology. Applied* 19.

Haushofer, J, A John, and K Orkin. 2019. "Can Simple Psychological Interventions Increase Preventive Health Investment?" .

Kerwin, Jason T and Rebecca L Thornton. 2021. "Making the Grade: The Sensitivity of Education Program Effectiveness to Input Choices and Outcome Measures." *The review of economics and statistics* 103 (2):251–264.

Khan, Adnan Q. et al. 2019. "Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings †." *American Economic Review* 109 (1):237–270.

Khan, Muhammad Yasir. 2020. "Mission Motivation and Public Sector Performance: Experimental Evidence from Pakistan." *Working paper* .

Lagarde, Mylène and Duane Blaauw. 2021. "Effects of incentive framing on performance and effort: evidence from a medically framed experiment." *Journal of the Economic Science Association* .

Leaver, Clare et al. 2019. *Measuring and Explaining Management in Schools: New Approaches using Public Data.*

Loyalka, Prashant, Anna Popova, Guirong Li, and Zhaolei Shi. 2019. "Does Teacher Training Actually Work? Evidence from a Large-Scale Randomized Evaluation of a National Teacher Training Program." *American economic journal. Applied economics* 11 (3):128–154.

Mansoor, Zahra. 2019. "Beyond the Paycheck: Intrinsic Motivation and Public Service Delivery." :22746URL `http://search.proquest.com/docview/1831255938/`.

McEwan, Patrick J. 2015. "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments." *Review of educational research* 85 (3):353–394.

Michailova, Julija. 2010. "Development of the overconfidence measurement instrument for the economic experiment." *IDEAS Working Paper Series from RePEc* URL `http://search.proquest.com/docview/1699082143/`.

Neckermann, Susanne and Xiaolan Yang. 2017. "Understanding the (unexpected) consequences of unexpected recognition." *Journal of economic behavior organization* 135:131–142.

Perry, James. 1996. "Measuring public service motivation: an assessment of construct reliability and validity." *Journal of public administration research and theory* 6 (1):5–24. URL `http://search.proquest.com/docview/39010044/`.

Piper, Benjamin and Medina Korda. 2011. *EGRA Plus: Liberia. Program Evaluation Report.* RTI International.

Popova, Anna, David Evans, Mary Breeding, and Violeta Arancibia. 2021. "Teacher Professional Development around the World: The Gap between Evidence and Practice." *The World Bank Research Observer* 37.

Popova, Anna et al. 2016. "Training Teachers on the Job : What Works and How to Measure It." 7834.

Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review* 94 (2):247–252.

Rotter, J.B. 1966. "Generalized Expectancies for Internal versus External Control of Reinforcement." *Psychological Monographs* 33:300–303.

Swann, William. 1996. *Self-Traps: The Elusive Quest for Higher Self-Esteem.* W H Freeman Co; First Printing edition (April 1, 1996).

Tanguy, B, Stefan Dercon, K Orkin, and AS Taffesse. 2014. "The future in mind: Aspirations and forward-looking behaviour in rural Ethiopia." .

Figures and Tables

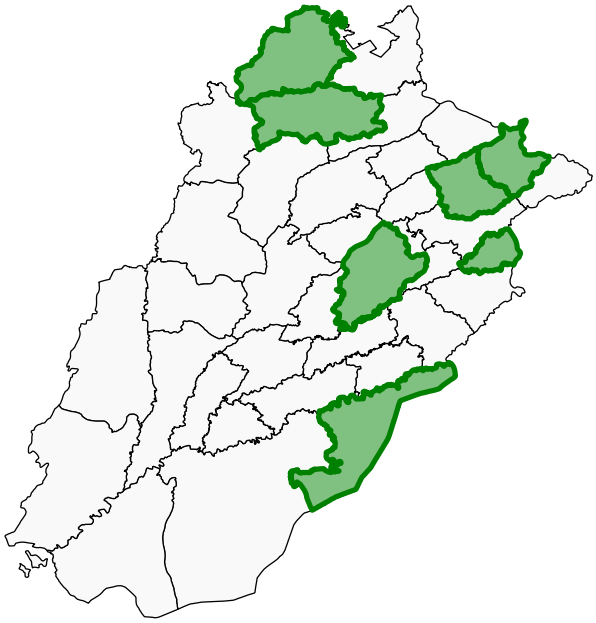


Figure 1: Punjab districts included in the recognition programme

Table 1: Treatment Effects on Training Test Scores

	(1)	(2)	(3)	(4)
Panel A				
Peer	0.017	0.048		
	(0.197)	(0.200)		
Career	0.308	0.329*		
	(0.198)	(0.196)		
Net Impact of Framing:				
Peer PLUS	-0.213	-0.210		
	(0.189)	(0.186)		
Net Impact of Framing:				
Career PLUS	-0.323*	-0.360**		
	(0.177)	(0.177)		
Peer and Career			0.164	0.189
			(0.175)	(0.175)
Net Impact of Framing:				
Pooled PLUS			-0.263*	-0.283**
			(0.136)	(0.135)
Panel B				
Peer PLUS*	-0.20	-0.16		
Career PLUS*	-0.02	-0.03		
PLUS*			-0.10	-0.09
Observations	3394	3392	3394	3392
PDS LASSO controls	No	Yes	No	Yes
District Fixed Effects	Yes	Yes	Yes	Yes

Notes: Errors are clustered at the training session level which is the unit of randomization. All regressions are an ANCOVA estimation with baseline values of the dependent variable and with district FE. Controls have been selected using the PDS Lasso procedure and include teacher, master trainer, and enumerator characteristics. Training scores are normalized by the mean and standard deviation of the pre test scores in the control group. The PLUS treatments with the asterisks present the overall impact of the treatments (incentive + the frame). Estimates are significant at the *10%, **5%, and ***1% level.

Table 2: Quantile Treatment Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	10th	20th	30th	40th	50th	60th	70th	80th	90th
Peer	0.041	0.017	0.008	-0.095	-0.056	0.014	0.041	0.053	-0.067
	(0.403)	(0.290)	(0.298)	(0.208)	(0.186)	(0.168)	(0.134)	(0.094)	(0.095)
Career	0.413	0.487	0.435	0.354*	0.309*	0.324**	0.312**	0.281***	0.105
	(0.530)	(0.319)	(0.308)	(0.214)	(0.182)	(0.162)	(0.127)	(0.101)	(0.076)
Net Impact of Framing:									
Peer PLUS	-0.229	-0.208	-0.120	-0.105	-0.080	-0.085	-0.103	-0.125	-0.111
	(0.293)	(0.270)	(0.243)	(0.204)	(0.198)	(0.149)	(0.121)	(0.096)	(0.086)
Net Impact of Framing:									
Career PLUS	-0.358	-0.397	-0.411*	-0.370**	-0.335**	-0.352**	-0.252**	-0.208**	-0.163**
	(0.382)	(0.303)	(0.239)	(0.181)	(0.157)	(0.137)	(0.115)	(0.095)	(0.064)
Observations	3392	3392	3392	3392	3392	3392	3392	3392	3392
PDS LASSO controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The regressions report quantile treatment effects. Errors are clustered at the training session level which is the unit of randomization. All regressions are an ANCOVA estimation with baseline values of the dependent variable and with district FE. Controls have been selected using the PDS Lasso procedure and include teacher, master trainer, and enumerator characteristics. Training scores are normalized by the mean and standard deviation of the pre test scores in the control group. Estimates are significant at the *10%, **5%, and ***1% level.

Table 3: Heterogeneous Treatment Effects - by Moderators

	Training Test Scores				
	(1)	(2)	(3)	(4)	(5)
Heterogeneous effects by:	Time till next promotion	Secretary visibility	Nature Contract	Peer known in class	Proportion peers known
Below Median (temp contract) x Peer	0.065	-0.004	0.001	0.115	0.191
	(0.224)	(0.204)	(0.245)	(0.204)	(0.257)
Above Median (perm contract) x Peer	0.068	0.308	0.066	-0.037	-0.188
	(0.202)	(0.271)	(0.202)	(0.222)	(0.288)
Below Median (temp contract) x Career	0.421**	0.293	0.005	0.304	0.146
	(0.213)	(0.206)	(0.228)	(0.211)	(0.307)
Above Median (perm contract) x Career	0.262	0.523**	0.397**	0.354*	0.327
	(0.222)	(0.220)	(0.196)	(0.205)	(0.246)
Below Median (temp contract) x Peer PLUS	-0.260	-0.243	-0.244	-0.288	-0.074
	(0.253)	(0.230)	(0.255)	(0.234)	(0.319)
Above Median (perm contract) x Peer PLUS	-0.051	0.259	-0.142	-0.047	-0.357
	(0.230)	(0.267)	(0.234)	(0.254)	(0.291)
Below Median (temp contract) x Career PLUS	0.030	-0.032	-0.016	-0.009	0.181
	(0.214)	(0.189)	(0.202)	(0.200)	(0.228)
Above Median (perm contract) x Career PLUS	-0.012	-0.010	-0.032	-0.073	-0.421
	(0.199)	(0.262)	(0.195)	(0.202)	(0.291)
Observations	2181	3392	3392	3392	3392
PDS LASSO controls	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes

Notes: Errors are clustered at the training session level which is the unit of randomization. All regressions include district FE. Each column represents heterogeneous treatment effects by a different moderator. Controls have been selected using the PDS Lasso procedure and include teacher, master trainer, and enumerator characteristics. Estimates are significant at t he *10%, **5%, and ***1% level.

Table 4: Treatment Effects on Motivation Index and Beliefs about Post Test Performance

	Motivation Index		Beliefs about Test Performance (out of 100)	
	(1)	(2)	(3)	(4)
Panel A				
Peer	0.051		-0.589	
	(0.034)		(1.100)	
Career	0.003		1.533	
	(0.034)		(1.220)	
Net Impact of Framing:				
Peer PLUS	0.042		2.376**	
	(0.037)		(1.006)	
Net Impact of Framing:				
Career PLUS	0.117***		-1.109	
	(0.040)		(1.397)	
Peer and Career		0.026		0.515
		(0.029)		(1.028)
Net Impact of Framing:				
Pooled PLUS		0.081***		0.542
		(0.029)		(0.961)
Panel B				
Peer PLUS*	0.09**		1.79	
Career PLUS*	0.12***		0.42	
PLUS*		0.11***		1.06
Observations	3373	3373	3072	3072
PDS LASSO controls	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes

Notes: Errors clustered at the training session level which is the unit of randomization. All regressions include district FE. Controls have been selected using the PDS Lasso procedure and include teacher, master trainer, and enumerator characteristics. The motivation index at baseline and endline is normalised by the mean and standard deviation of the control group at baseline. The PLUS treatments with the asterisks present the overall impact of the treatments (Incentive + the frame). Estimates are significant at the *10%, **5%, and ***1% level.

Table 5: Treatment Effects on Overconfidence

	Overconfidence (beliefs about performance - actual performance)		Overconfidence (=1 if above median overconfidence)	
	(1)	(2)	(3)	(4)
Panel A				
Peer	0.710		0.028	
	(2.878)		(0.063)	
Career	-2.581		-0.076	
	(2.891)		(0.067)	
Net Impact of Framing:				
Peer PLUS	6.819**		0.129**	
	(2.829)		(0.058)	
Net Impact of Framing:				
Career PLUS	5.329**		0.122**	
	(2.239)		(0.058)	
Peer and Career		-0.899		-0.023
		(2.601)		(0.058)
Net Impact of Framing:				
Pooled PLUS		5.982***		0.124***
		(1.901)		(0.044)
Panel B				
Peer PLUS*	7.53**		0.16**	
Career PLUS*	2.75		0.05	
PLUS*		5.08*		0.10*
Observations	3072	3072	3072	3072
PDS LASSO controls	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes

Notes: Errors clustered at the training session level which is the unit of randomization. The dependent variable is overconfidence. In the first two columns, its is constructed as a continuous variable that is the difference between teacher beliefs of how well they scored on the test and actual test score at endline. In the last two columns, we construct a dummy variable of above median overconfidence based on the continuous variable. All regressions include district FE. Controls have been selected using the PDS Lasso procedure and include teacher, master trainer, and enumerator characteristics. The PLUS treatments with the asterisks present the overall impact of the treatments (Incentive + Controls have been selected using the PDS Lasso procedure and include teacher, master trainer, and enumerator characteristics. the frame). Estimates are significant at the *10%, **5%, and ***1% level.

Table 6: Mediation Analysis: Average Controlled Direct Effects

	(1)	(2)	(3)
	Post Test Scores	Post Test Scores (with mediator)	Post Test Scores (de-mediated)
Peer and Career	0.189	0.177	0.177
	(0.175)	(0.180)	(0.173)
Net Impact of Framing:			
Pooled PLUS	-0.283**	-0.040	-0.037
	(0.135)	(0.154)	(0.126)
Observations	3392	3212	3212
PDS LASSO controls	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes

Notes: Errors are clustered at the training session level which is the unit of randomization. In column 1, the dependent variable is the training post test score. In column 3, the dependent variable is the de-mediated post-test score based on the sequential g-estimation procedure (as in Acharya et al. (2016)). This is calculated by: 1) regressing the main outcome on treatment, pre-treatment controls, the mediator, interaction between the mediator and treatment, and interaction between the mediator and all other pre-treatment variables; 2) calculating the de-mediated post-test scores which is the predicted outcome excluding all coefficients that include the mediator fixed at a specific value. Controls have been selected using the PDS Lasso procedure and include teacher, master trainer, and enumerator characteristics. Estimates are significant at the *10%, **5%, and ***1% level.

Appendix

Appendix A: Tables

A. 1: Descriptive Statistics

	(1)	(2)	(3)	(4)	(5)
	Mean	Sd	p0.25	p0.50	p0.75
Basic teacher characteristics					
Age	45.54	10.31	37	49	54
1[Male]	0.43				
Salary	77604.47	31779.54	51000	71000	97328
Years of experience	19.99	10.94	10	22	30
Years of education	15.72	0.83	16	16	16
1[Married]	0.90				
Total teachers in a session	27.38	6.48	23	26	31
Basic job characteristics					
Job Grade	15.53	2.58	15	16	17
Time till next promotion (in yrs)	6.06	4.83	2	5	10
HT's school's enrollment capacity	467.05	480.86	189	317	555
Baseline Performance					
Pre Test Scores (out of 100)	0.34	0.14	0.27	0.33	0.40
Pre Test Scores (normalised)	-0.07	1.03	-0.57	-0.07	0.44
Non-cognitive traits					
Personality traits & Self-efficacy					
BFI Index	0.01	0.55	-0.32	0.02	0.35
Openness	0.01	1.00	-0.63	0.02	0.68
Extraversion	0.01	1.00	-0.71	-0.13	1.04
Conscientiousness	0.01	1.10	-0.07	-0.07	0.27
Agreeableness	-0.00	1.00	-0.77	0.00	0.76
Neuroticism	-0.00	0.99	-0.93	0.31	0.93
Self-Efficacy	-0.01	0.99	-0.68	-0.12	0.92
Primary Motivational Orientation					
Extrinsic Motivation	0.25				
Intrinsic Motivation	0.41				
Pro-social Motivation	0.31				
Other intrinsic measures					
PSM Index	0.00	0.38	-0.26	-0.01	0.25
Donation in hypothetical game (total PKR 10,000)	4052	2876	2000	4000	5000
Observations	3394				

Notes: Pretest scores, overall personality index, each individual personality trait, and self-efficacy are normalized against the control group.

A. 2: Randomization Balance - All Treatments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Control	Peer	Career	Peer +	Career +	C-Peer	C-Career	C-Peer +	C-Career +	Peer- Career	Peer-Peer +	Career-Car +
Age	44.79	46.70	46.65	46.16	45.39	0.03**	0.03**	0.15	0.53	0.96	0.58	0.18
	(0.90)	(0.88)	(0.83)	(0.97)	(0.97)							
1[Male]	0.42	0.57	0.50	0.56	0.45	0.01***	0.18	0.03**	0.67	0.26	0.94	0.39
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)							
Salary	69009	74779	77471	74336	70669	0.06	0.02**	0.11	0.61	0.45	0.89	0.07
	(3578)	(3408)	(3671)	(3404)	(3404)							
Years of Education	15.66	15.72	15.73	15.72	15.73	0.30	0.17	0.29	0.19	0.74	0.95	0.99
	(0.06)	(0.06)	(0.06)	(0.06	(0.06)							
1[Married]	0.91	0.95	0.93	0.93	0.94	0.02**	0.47	0.57	0.14	0.17	0.16	0.52
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)							
Basic job characteristics												
Time till next promotion (in yrs)	6.05	6.05	5.93	6.36	5.97	1.00	0.75	0.44	0.81	0.75	0.44	0.91
	(0.44)	(0.48)	(0.48)	(0.43)	(0.43)							
HT's school's enrollment capacity	237	267	330	256	246	0.32	0.04**	0.64	0.79	0.15	0.77	0.06
	(26.65)	(26.55)	(35.78)	(31.67)	(31.67)							
1[Urban]	0.11	0.15	0.17	0.07	0.08	0.35	0.19	0.30	0.39	0.68	0.08	0.06
	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)							
Baseline Performance												
Pre Test	-0.04	0.02	-0.06	-0.09	-0.15	0.70	0.84	0.64	0.33	0.58	0.45	0.40
	(0.14)	(0.15)	(0.13)	(0.12)	(0.14)							
Non-Cognitive Traits												
Overall BFI Index	0.09	0.05	0.06	0.08	0.03	0.21	0.34	0.69	0.07	0.78	0.46	0.39
	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)							
Self-efficacy Index	-0.04	-0.03	-0.03	-0.04	-0.02	0.87	0.95	0.99	0.80	0.91	0.86	0.82
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)							
Intrinsic Motivation	0.51	0.48	0.52	0.50	0.51	0.35	0.73	0.90	0.97	0.22	0.53	0.75
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)							
Extrinsic Motivation	0.20	0.21	0.17	0.20	0.21	0.61	0.26	0.80	0.51	0.06	0.80	0.03**
	(0.03)	(0.03)	(0.03)	(0.03	(0.03)							
Pro-social Motivation	0.27	0.29	0.29	0.28	0.26	0.54	0.52	0.90	0.56	0.97	0.67	0.23
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)							
PSM Index	0.08	0.08	0.08	0.07	0.06	0.98	0.96	0.79	0.48	0.93	0.75	0.46
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)							
Joint F-Test						0.24	0.70	0.87	0.54	0.94	0.78	0.14
Observations	716	649	687	635	707							

Notes: The first five columns report the mean and standard errors of the four recognition treatments and the control group. The last eight columns show equality of means between the control group and the treatment group, and between each treatment, for each variable of interest. Estimates are significant at the **5%, and ***1% level.

A. 3: Attrited Sample and Treatments

	(1) Attrited (=1 if sample attrited)
Peer	0.006
	(0.019)
Career	-0.004
	(0.013)
Peer PLUS	-0.009
	(0.014)
Career PLUS	-0.008
	(0.013)
Observations	3493
Controls	No
District FE	Yes

Notes: Errors clustered at the training session level which is the unit of randomization. Estimates are significant at the *10%, **5%, and ***1% level.

A. 4: Balance across Attrited and Main Sample

	(1) Attrited Sample	(2) Main Sample	(3) P-value difference
Basic teacher characteristics			
Age	46.02	45.87	0.89
	(1.22)	(0.69)	
Gender (=1 if male)	0.45	0.50	0.38
	(0.06)	(0.04)	
Salary	73545	73186	0.39
	(4472)	(2885)	
Years of Experience	21.19	20.11	0.95
	(1.43)	(0.90)	
Years of Education	15.77	15.72	0.42
	(0.10)	(0.05)	
Married (=1 if married)	0.91	0.93	0.39
	(0.04)	(0.01)	
Basic job characteristics			
Time till next promotion (in yrs)	5.26	6.10	0.26
	(0.88)	(0.38)	
HT's school's enrollment capacity	206	269	0.07
	(38.20)	(16.36)	
School Location of HT (=1 if urban)	0.18	0.12	0.21
	(0.06)	(0.03)	
Baseline Performance			
Pre Test Scores (normalised)	-0.06	-0.20	0.26
	(0.18)	(0.14)	
Non-Cognitive Traits			
Overall BFI Index	0.00	0.06	0.31
	(0.07)	(0.03)	
Self-Efficacy Index	0.10	-0.04	0.15
	(0.12)	(0.06)	
Intrinsic Motivation	0.47	0.50	0.49
	(0.06)	(0.03)	
Extrinsic Motivation	0.24	0.19	0.35
	(0.06)	(0.02)	
Pro-social Motivation	0.28	0.28	0.99
	(0.06)	(0.02)	
PSM Index	0.01	0.07	0.09
	(0.06)	(0.02)	
Joint F			0.11
Observations	100	3394	

Notes: Errors are clustered at the training session level which is the unit of randomization. The first two columns present the means for the attrited and the main sample, whereas the third column presents the p-value difference for each variable of interest. Estimates are significant at the **5%, and ***1% level.

A. 5: Non-Consenting Sample and Treatments

	(1)
	Non-Consent (=1 if did not consent)
Peer	0.004
	(0.05)
Career	0.040
	(0.05)
Peer PLUS	0.008
	(0.05)
Career PLUS	0.030
	(0.05)
Observations	3394
Controls	No
District Dummies	Yes

Notes: Errors clustered at the training session level which is the unit of randomization. Estimates are significant at the *10%, **5%, and ***1% level.

A. 6: Treatment Effects on Motivation and Self Beliefs

	Intrinsic Motivation	External Locus	Self Efficacy
Peer	0.062	-0.064	0.031
	(0.043)	(0.062)	(0.047)
Career	-0.002	-0.022	0.001
	(0.044)	(0.064)	(0.058)
Net Impact of Framing:			
Peer PLUS	-0.026	-0.071	0.089*
	(0.048)	(0.066)	(0.052)
Net Impact of Framing:			
Career PLUS	0.092*	-0.114	0.134**
	(0.050)	(0.070)	(0.066)
Peer PLUS*	0.04	-0.13*	0.12**
Career PLUS*	0.09*	-0.14*	0.14**
Observations	3337	3306	3364
PDS LASSO controls	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes

Notes: Errors clustered at the training session level which is the unit of randomization. All regressions are an ANCOVA estimation with baseline values of the dependent variable and district FE. Controls include trainee-level teacher controls, master trainer controls, and enumerator controls that have been selected through the PDS lasso procedure. All dependent variables are normalized by the mean and standard deviation of the control group. Estimates are significant at the *10%, **5%, and ***1% level.

A. 7: Correlation between Actual and Predicted Overconfidence

	(1)	(2)	(3)
	Actual Overconfidence	Predicted Overconfidence	Post Test Scores
Actual Overconfidence	1.000		
Predicted Overconfidence	0.303	1.000	
Post Test Scores	-0.735	-0.366	1.000

Notes: Predicted overconfidence is estimated by predicting actual overconfidence using baseline variables (restricted to control group) using LASSO. Actual overconfidence is constructed as a continuous variable that is the difference between teacher beliefs of how well they scored on the test and actual post-test score at endline.

A. 8: Treatment Effects on Predicted Overconfidence

	(1)	(2)	(3)	(4)
Peer	0.758	0.539		
	(0.676)	(0.514)		
Career	0.623	0.430		
	(0.646)	(0.565)		
Net Impact of Framing: Peer PLUS	1.999***	1.579***		
	(0.674)	(0.554)		
Net Impact of Framing: Career PLUS	1.439**	0.910		
	(0.662)	(0.588)		
Peer and Career			-0.055	0.695
			(0.941)	(0.560)
Net Impact of Framing: Pooled PLUS			1.403*	1.694***
			(0.752)	(0.492)
Observations	3259	3259	3259	3259
PDS LASSO controls	No	Yes	No	Yes
District Fixed Effects	Yes	Yes	Yes	Yes

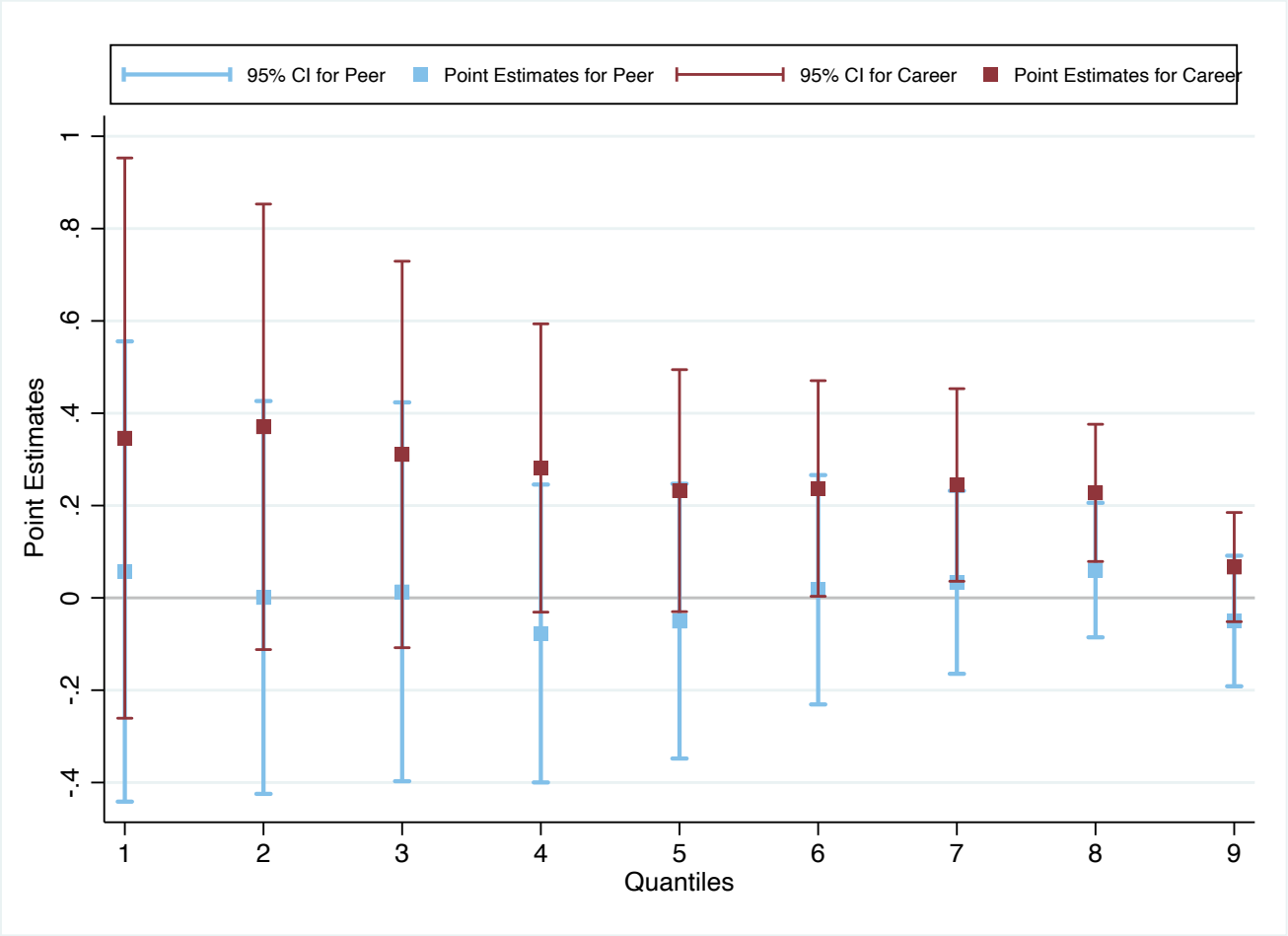
Notes: Errors clustered at the training session level which is the unit of randomization. The dependent variable is predicted overconfidence. All regressions include district FE. Controls have been selected using the PDS lasso procedure. Estimates are significant at the *10%, **5%, and ***1% level.

A. 9: Data Validity (Cronbah's Alpha)

	(1) Cronbah Alpha
Self-efficacy	0.73
Intrinsic motivation	0.66
Locus of control	0.75
PSM Index	0.52

Notes: We present cronbah's alpha for each of these scales as a measure of internal consistency.

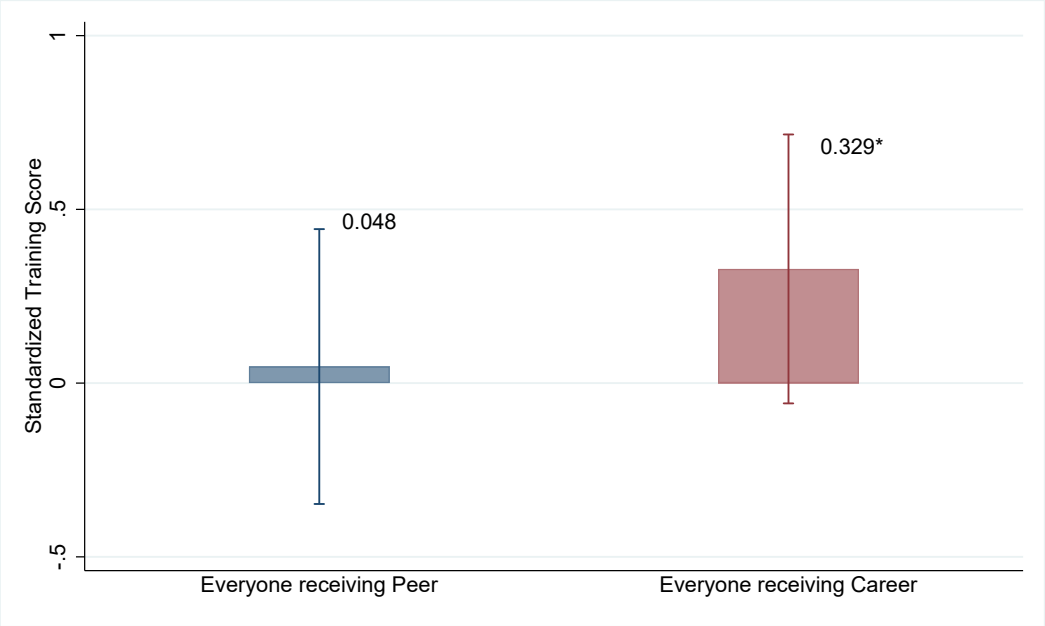
Appendix B: Figures



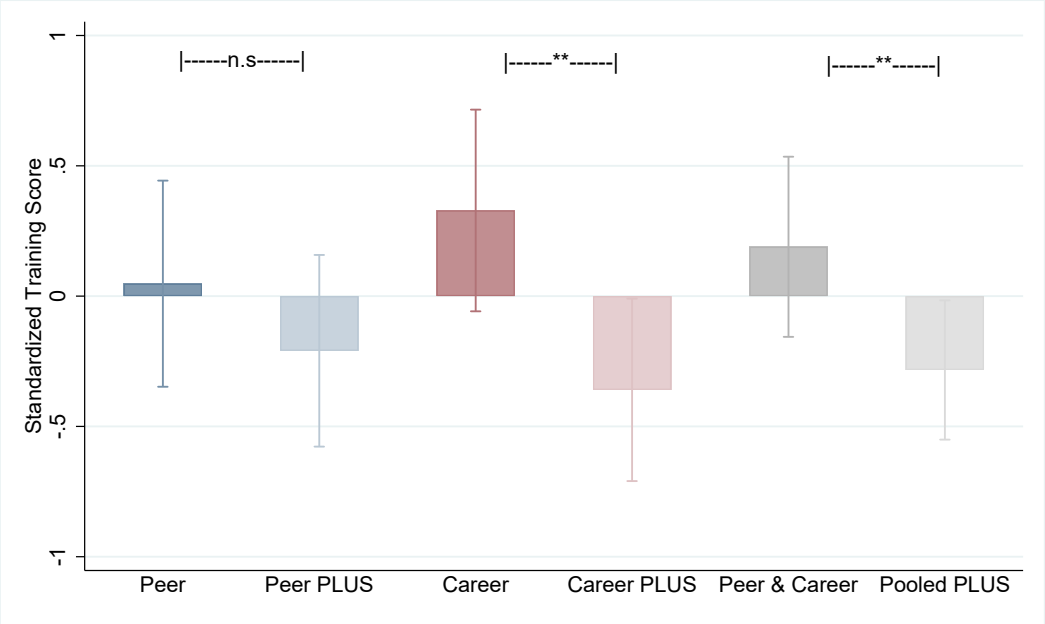
B. 1: Quantile Treatment Effects

B. 2: Recognition Certificate

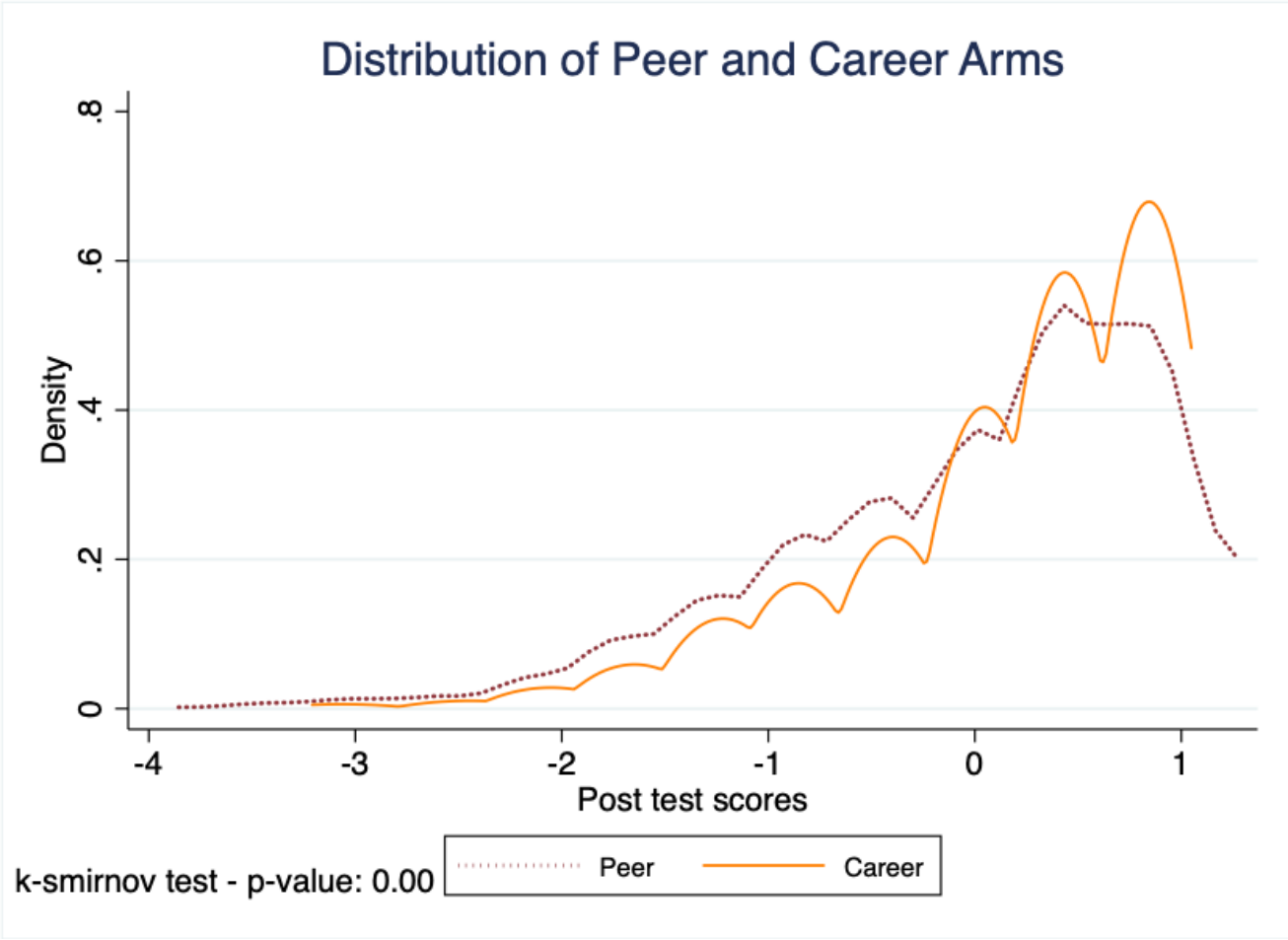
B. 3: Motivational Framing Hand-out



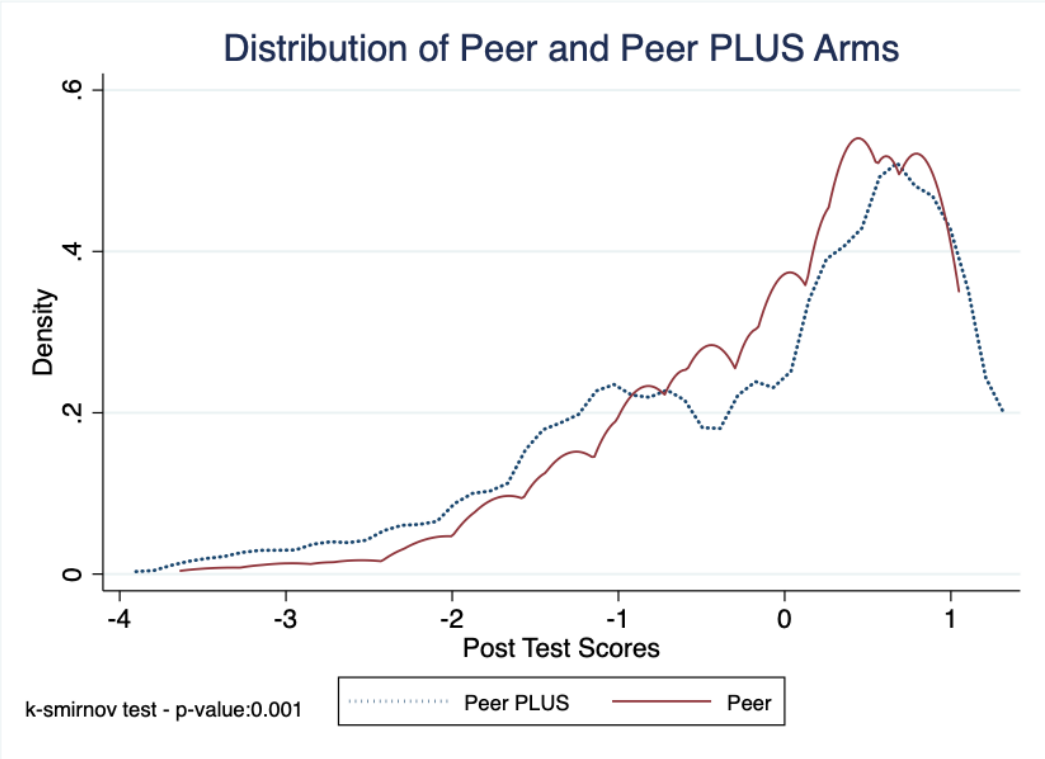
B. 4: Treatment effects (Peer and Career)



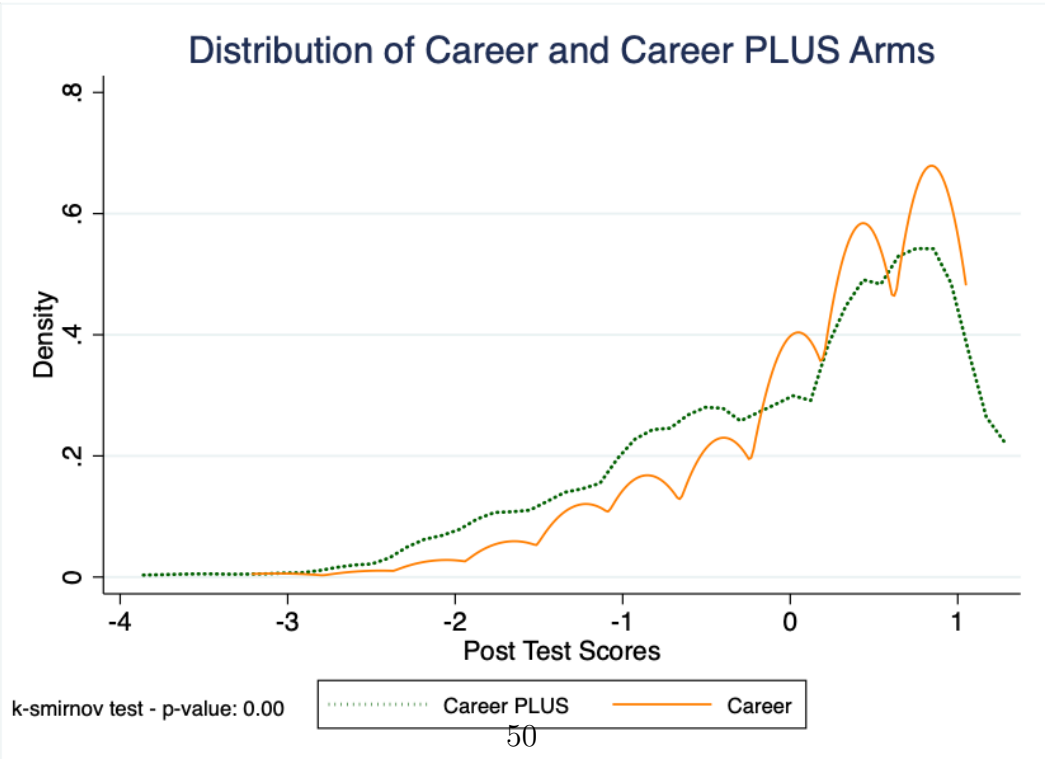
B. 5: Treatment effects (PLUS arms)



B. 6: K smirnov-test: Peer and Career Distribution



(a) Peer and Peer PLUS



(b) Career and Career PLUS

B. 7: K smirnov-test: Peer/Career and PLUS Counterparts