Milestone	Estimated Completion Date	Deliverable	Amount
1	Award date + 30 days = Oct 5, 2012	• Updated project implementation plan. • Complete list of indicators that will be tracked. The milestone is considered fulfilled when all deliverables have been submitted to the AOR for review/concurrence.	15% $14,584
2	Award date + 6 months = March 5, 2013	• Goal setting schemes are introduced at all health centers • Monthly data collection from health centers occurs as scheduled The milestone is considered fulfilled when all activities above are complete, which can be verified to AOR through a brief report summarizing the status, which includes a summary of the schemes.	25% $24,758
3	Award date + 10 months = July 5, 2012	• Endline surveys to collect subjective CHA is complete The milestone is considered fulfilled when all activities above are complete, which can be verified to AOR through a brief report summarizing the status.	25% $24,758
4	Award date + 2 years = Sept 5, 2014	• Endline report of preliminary findings • Estimated cost effectiveness of the interventions. • Dissemination approach/Outreach strategy developed that includes specific stakeholders and a timeline The milestone is considered complete when all activities above are complete and have been submitted to the AOR for review/concurrence.	35% $34,932.46
	90 days following close of award	Final project report that includes the following: • Summary of project implementation - which activities occurred as planned and where did the project experience challenges. • Next steps that will be required to achieve the intended scale up plans, including noting potential partners that are interested and/or commitments made by stakeholders such as the GoZ. • Final cost effectiveness calculations of the interventions.	
		Total Fixed Obligation	$99,032.46

Implementation Plan (Updated October 1, 2012)

Date Program Activity	**Research Activity**
Sept 2012	Finalize experimental protocols and systems • Test, troubleshoot and finalize helpline system and protocols • Test, troubleshoot and finalize audit system and protocols • Test, troubleshoot and finalize automated household visit sms system • Test, troubleshoot and finalize automated goal-setting sms system Compile data from incentives research
Oct 2012	Introduce goal-setting treatments at health centre level • Send ministry letters • Send SMSes to CHAs and supervisors • Call CHAs and supervisors to announce goal setting treatments Analyze recruitment and compensation data jointly and prepare final report for Government of Zambia and second academic paper on the impact of recruitment and compensation on CHW retention and performance
Nov 2012 – April 2013	Collect CHA performance data • Analyze incoming SMS receipts for household visits to monitor CHA performance • Analyze incoming monthly goal SMS messages • Collect monthly data from Ministry of Health (M&E Unit) to complement and verify CHA performance data collected via SMS receipts for household visits • Collect hard copy receipt books from CHAs to verify data collected via SMS receipts • Collect copies of community mobilization registers, acute care registers and household visit registers Data quality assurance • Conduct daily audits to confirm SMS receipts for household visits are genuine • Manage helpline to ensure CHAs and supervisors questions are promptly addressed Send feedback and reminders to CHAs and supervisors • Send out scheduled reminders for monthly goal submissions and targets • Send out monthly performance feedback to CHAs and supervisors
Jan – Feb 2013	Prepare for endline survey • Develop, translate and pretest questionnaires • Recruit and train surveyors • Plan logistics of survey
March-April 2013	Conduct endline survey
May-July 2013	Analyze recruitment, incentive and goal-setting data jointly Prepare final report for Government of Zambia Disseminate findings to local stakeholders Prepare second academic paper on the impact of recruitment and motivation on CHA retention and performance

List of Indicators that will be tracked

Indicator	Sources of information
Number and duration of household visits	1) Incoming SMS messages for household visit receipts 2) Hard copy household visit receipts
Goals set	Incoming SMS messages for goal-setting
Number and type of help request by CHA/supervisor	Helpline
Accuracy of household visit receipt data	1) Audit calls to clients 2) Hard copy household visit receipts 3) Household visit register
Number and type of community level meetings held	1) Monthly aggregated data from the M&E unit at the Ministry of Health 2) Community mobilization register
Health indicators at health post level (e.g. number of patients diagnosed and treated for malaria, number of households with a pit latrine)	1) Monthly aggregated data from the M&E unit at the Ministry of Health 2) Acute care register 3) Household visit register
Perceptions on CHA performance from clients in the CHA communities	Endline survey

DIV Milestone 2: Launch Of Goal-Setting Experiment And Summary Of Subsequent Activities.

The goal-setting experiment was successfully launched on 20 November 2012 after the research team completed the randomisation and set up of the rapidSMS system. The goal- setting experiment aims to understand whether and how different ways to set targets affect Community Health Assistant's (CHA's) motivation and performance. Two goal-setting schemes were launched;

- **Group 1: Bottom-up goals**. Based on their experience and work environment, every month, CHAs set their own individual specific goals on the number of average households to visit per day in a given month.

- **Group 2: Top-down goals**. Based on their expert knowledge of the local communities and the working environment, every month, CHA's Supervisors set individual specific goals on the number of average daily household visits each CHA should aim to visit in a given month.

As part of the launch of the goal-setting experiment, letters from the Ministry of Health with detailed instructions on how to set goals were sent to all supervisors. To complement the Ministry letters, the field team sent phone message instructions to both the CHAs and their Supervisors. In addition, phone calls were made to ensure that the goal-setting experiment was announced to all CHAs and supervisors.

The table below presents a summary of the goal experiment treatments;

T1 - CHA sets goal		**T2 - Supervisor sets goal**	
No. of CHAs	154	No. of CHAs	150
No. of Supervisors	75	No. of Supervisors	72
Total in T1	**229**	**Total in T2**	**222**
Overall Total			**451**

Since the successful launch of the goal-setting experiment, the field team has been implementing the following activities;

	Activity	Frequency
i	Sending reminders to CHAs and supervisors to set their monthly goals. For any given month, the goal setting window opens on the 20th of the preceding month and closes on the 7th of the goal setting month. Initial reminders are sent on the 20th, with follow up reminders to defaulters sent on the 1st, 3rd and 5th.	Monthly
ii	Reviewing of monthly goals; and investigating goal outliers.	Monthly
iii	Sending monthly performance feedback on; i) the total number of household visits, ii) the average number and duration of visits that each CHA reported.	Monthly
iv	Reviewing household visit logs and investigating any outliers.	Daily
v	Conducting household visit audits to ensure that household visit records as reported by CHAs are accurate.	Daily
vi	Attending to help request logs.	Daily
vii	Identifying programmatic changes in the study environment and communicating these to the Ministry of Health; and assessing any potential impact or threats of such changes to the study design.	Ad hoc

Since launching the goal-setting experiment, about 80% of the target CHAs and supervisors have at least set a monthly goal. The remaining 20% are unreachable due to poor or no phone network. The field team has been calling district focal persons to arrange to these CHAs/supervisors to call the researchers when in network. With regard to household visits, the CHAs are making an average of 30.6 household visits per CHA per month with an average duration of 34 minutes per visit per CHA.

Innovations for Poverty Action – Zambia
"Recruiting and Motivating Community Health Workers:
A National Field Experiment in Zambia"

USAID DIV Grant No. AID-OAA-G-12-00011

Milestone 3: Project Update Report

July 1, 2013

---

1. Summary of project activity between March and June 2013

The basic performance monitoring system for this project revolves around a mobile phone-based system in which Community Health Assistants report their household visits using structured SMS messages. During March-June 2013, CHAs have continued to report their household visits using the SMS system, and we have continued to monitor the system. In particular, we have continued to conduct audit calls to verify household visits for households for which a phone number was submitted in the on a sample taken from daily reported visits.

The experimental intervention we are investigating examines the role of autonomy in goal-setting and health worker performance. In one treatment group, the CHAs set their own goals for how many households they intend to visit per day during the next month. In the other treatment group, the CHA's supervisor sets the goal for the CHA. With DIV support, we have created an automated goal-setting system by which the CHAs and supervisors submit goals using mobile phones. The CHAs and supervisors in the *bottom-up goals* and *top-down goals* treatment groups, respectively, have continued setting monthly goals for household visits (see Section 2 below for more details).

While SMS reporting provides a real-time data source as well as a hedge against the risk of hard-copy reports being unretrivable, we also wish to collect the CHAs' hard-copy reports. Thus, as part of the performance monitoring of the goal-setting experiment, in collaboration with the Ministry of Health, we are retrieving the CHAs' hard-copy household visit records (i.e., receipts) that serve as official documentation of household visits. Collection from 160 communities across the most rural regions of Zambia is an arduous logistical task involving engagement and cooperation with all levels of the Ministry of Health (national headquarters, seven provincial offices, 47 district offices, and 160 CHA health posts). Once collected, these records will be double-entered, cleaned and integrated to make up for any visits that were conducted and not reported due to lack of network coverage.

Finally, in April-May 2013, we conducted an extensive follow-up survey of CHAs. In April, all CHAs were asked by the Minstry of Health to report to the training school in Ndola (the site of their year-long training) for a three-week continuing education course on new job responsibilities (e.g., first aid and giving injections). In conjunction with this training, we organized and administered a survey of the 298 out of 307 CHAs who reported for the training (see Section 4 below).

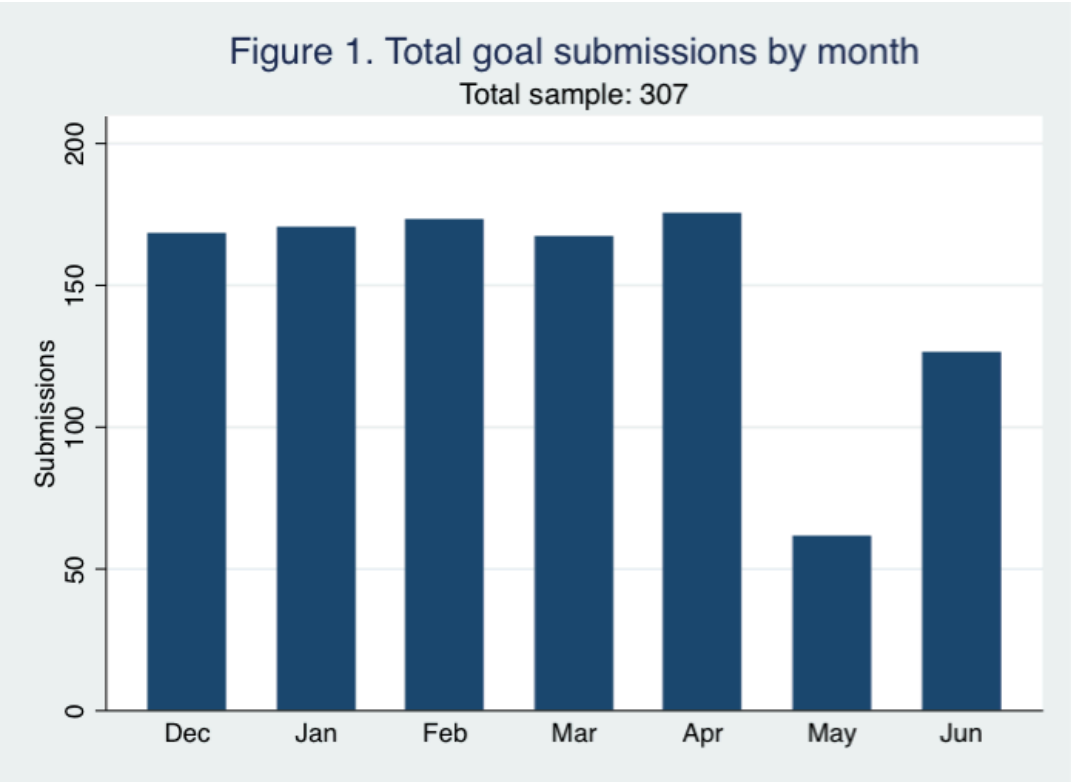
In addition to the main activities above, we have implemented the following ongoing activities:

	Activity	Frequency
i	Sending reminders to CHAs and supervisors to set their monthly goals. For any given month, the goal setting window opens on the 20th of the preceding month and closes on the 7th of the goal setting month. Initial reminders are sent on the 20th, with follow up reminders to defaulters sent on the 1st, 3rd and 5th.	Monthly
ii	Reviewing of monthly goals; and investigating goal outliers.	Monthly

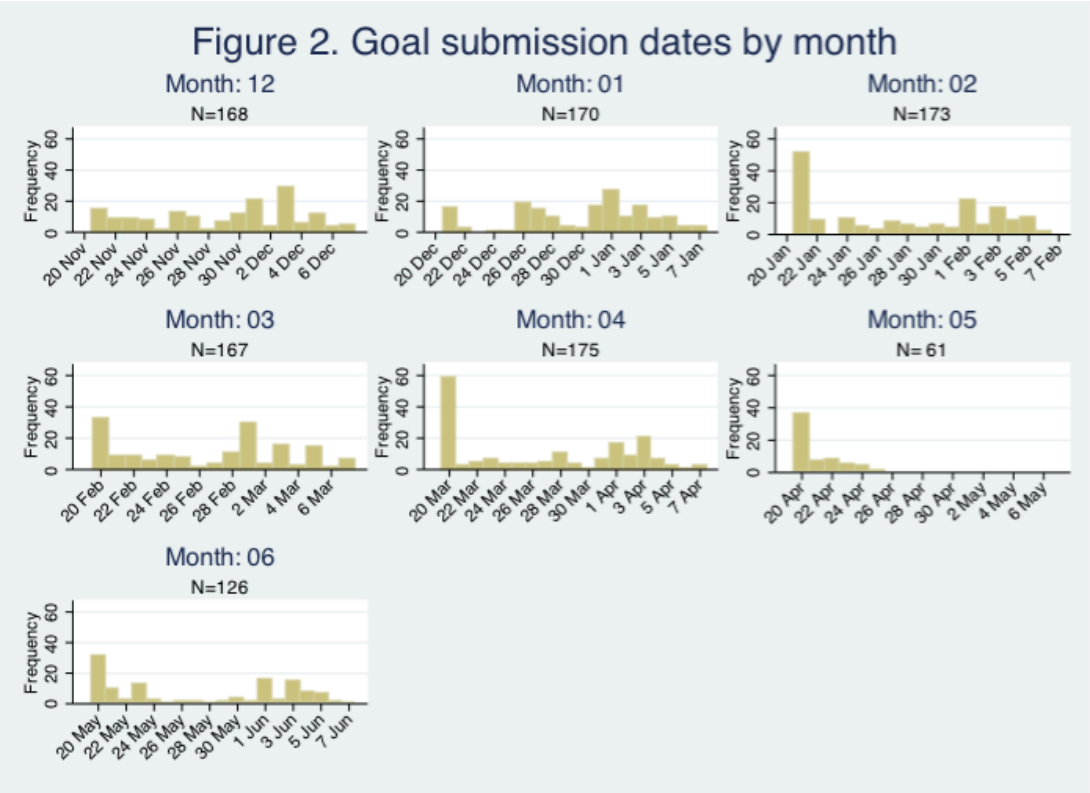
iii	Sending monthly performance feedback on: i) the total number of household visits, ii) the average number and duration of visits that each CHA reported.	Monthly
iv	Reviewing household visit logs and investigating any outliers.	Daily
v	Conducting household visit audits to ensure that household visit records as reported by CHAs are accurate.	Daily
vi	Attending to help request logs.	Daily
vii	Identifying programmatic changes in the study environment and communicating these to the Ministry of Health; and assessing any potential impact or threats of such changes to the study design.	As needed

2. Monitoring of the Goal-setting experiment

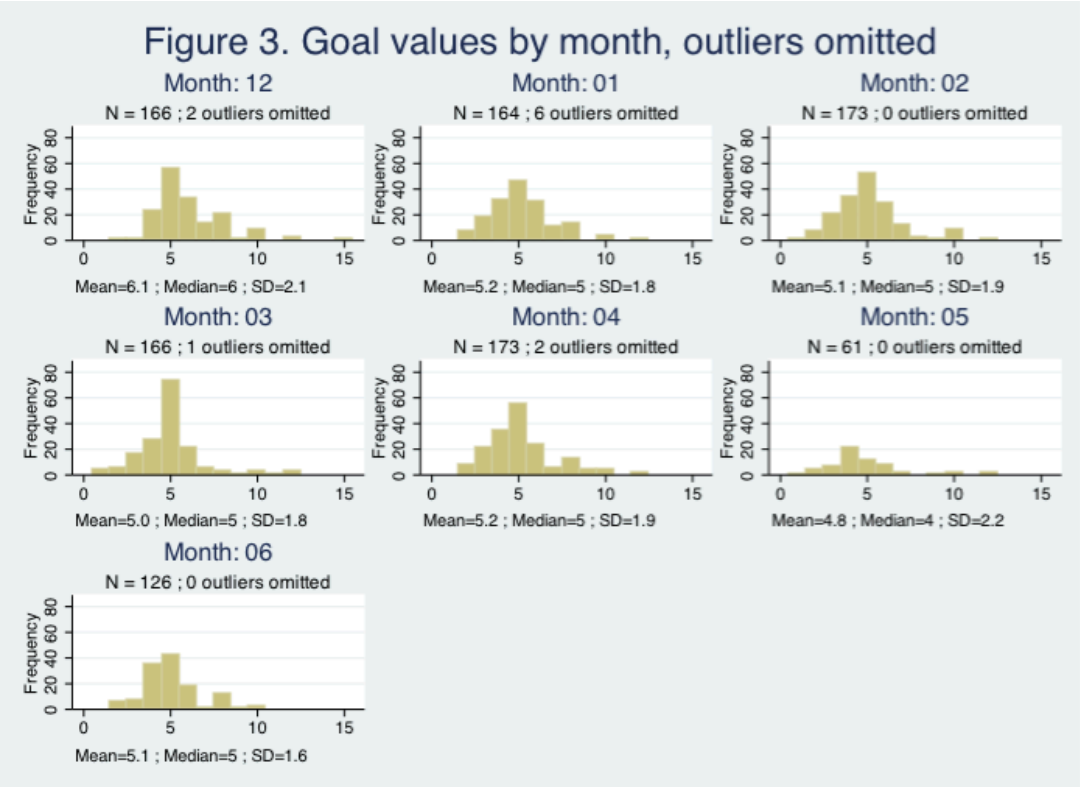
The goal-setting experiment was launched in December 2012 and has now been running for seven months. The number of CHAs choosing (in the bottom-up group) or receiving (in the top-down group) a goal each month is indicated in Figure 1 below. Thus far, on average, take-up has been 48.5% each month; that is, in an average month, 48.5% of CHAs have either chosen a goal or had a goal chosen for them. The drop in goal submissions is due to the training that occurred that month.



The goal-setting system allows us to monitor the timing of goal submissions (Figure 2). For a given month, a household visit goal can be submitted beginning on the 20th day of the preceding month, up to the 7th day of the goal month. In general, as the goal-setting experiment has proceeded, more CHAs/supervisors have begun to submit goals on the first day of the goal-setting window.



Finally, the system also tracks the actual value of the goal for how many households to visit per day. The distribution of goal values is shown in Figure 3. There is a concentration of values around 5 households per day, presumably because the Ministry of Health has set this as a guideline for all CHAs.



3. Status of Endline Survey

We had originally forecasted conducting a full endline survey in June 2013. However, due to network coverage issues, it took a considerable amount of time (six months) to officially announce the goal-

setting experiment to all 307 CHAs. Because of this delay, we have chosen to extend the length of the goal-setting experiment by six months, in order to capture twelve full months of household visit and goal-setting data. Thus, we intend to conduct the endline survey in November/December 2013. Instead of the endline survey, we conducted the aformentioned follow-up survey in April/May.

4. Ndola April 2013 Survey Summary

In April of 2013 the Ministry of Health and CHAI decided to gather all CHAs back in Ndola, where their year-long training took place, to conduct a continuing education training. We sent our project team to Ndola to administer a survey to each CHA, in order to collect mid-term data on CHAs' work patterns, time allocation across job responsibilities, job satisfaction, work challenges, and other work-related variables. In total, 298 CHAs reported to Ndola, all of whom we surveyed. Some summary statistics include: number of households in their catchment areas (average 1,500), main mode of transport (69% bicycle, 30% foot), average travel time to households (42 minutes by foot, 37 minutes by bicycle), health post has network reception at least some of the time (60%), average number of hours worked (42 per week), current job satisfaction (92% happy or very happy), has been paid (60%). Full results can be seen in the attached summary tables.

INNOVATIONS FOR POVERTY ACTION: NDOLA APRIL 2013 SURVEY SUMMARY

```
**********************************
**** SUMMARIZE ALL VARIABLES ****
**********************************
```

=1 if provided consent	Freq.	Percent	Cum.
Yes	298	100.00	100.00
Total	298	100.00	

Type of facility	Freq.	Percent	Cum.
Health Post	259	86.91	86.91
Rural Health Centre	38	12.75	99.66
Don't know	1	0.34	100.00
Total	298	100.00	

=1 if still at same HP as application HP	Freq.	Percent	Cum.
No	17	5.70	5.70
Yes	281	94.30	100.00
Total	298	100.00	

=1 if still living in same area as when applied	Freq.	Percent	Cum.
No	19	6.38	6.38
Yes	279	93.62	100.00
Total	298	100.00	

Variable	N	Mean	SD	Min	Max
Number of households in catchment area	255	1442.55	1697.60	60.00	10673.00

Main mode of transport	Freq.	Percent	Cum.
Foot	90	30.20	30.20
Bicycle	205	68.79	98.99
Motorcycle	2	0.67	99.66
Other	1	0.34	100.00
Total	298	100.00	

Main mode of transport (other)	Freq.	Percent	Cum.
Oxy cart	1	100.00	100.00
Total	1	100.00	

Travel time by mode of transport

-> transport = Foot

Variable	N	Mean	SD	Min	Max
Travel time by main mode of transport (minutes)	90	42.82	52.29	0.00	210.00

```
-------------------------------------------------------------------------------------------------------------------------------------
-> transport = Bicycle

                                          Variable |        N     Mean       SD      Min      Max
-----------------------------------------------------+--------------------------------------------
 Travel time by main mode of transport (minutes) |      205    37.00    36.59     1.00   230.00


-------------------------------------------------------------------------------------------------------------------------------------
-> transport = Motorcycle

                                          Variable |        N     Mean       SD      Min      Max
-----------------------------------------------------+--------------------------------------------
 Travel time by main mode of transport (minutes) |        2    10.00     7.07     5.00    15.00


-------------------------------------------------------------------------------------------------------------------------------------
-> transport = Other

                                          Variable |        N     Mean       SD      Min      Max
-----------------------------------------------------+--------------------------------------------
 Travel time by main mode of transport (minutes) |        1    15.00        .    15.00    15.00


HP STAFFING


      Number of |
      staff: CO |      Freq.     Percent        Cum.
----------------+-----------------------------------
              0 |        271       90.94       90.94
              1 |         26        8.72       99.66
              2 |          1        0.34      100.00
----------------+-----------------------------------
          Total |        298      100.00


      Number of |
   staff: Nurse |      Freq.     Percent        Cum.
----------------+-----------------------------------
              0 |        168       56.38       56.38
              1 |        125       41.95       98.32
              2 |          5        1.68      100.00
----------------+-----------------------------------
          Total |        298      100.00


      Number of |
 staff: Midwife |      Freq.     Percent        Cum.
----------------+-----------------------------------
              0 |        280       93.96       93.96
              1 |         16        5.37       99.33
              3 |          1        0.34       99.66
              4 |          1        0.34      100.00
----------------+-----------------------------------
          Total |        298      100.00


      Number of |
     staff: EHT |      Freq.     Percent        Cum.
----------------+-----------------------------------
              0 |        275       92.28       92.28
              1 |         22        7.38       99.66
              2 |          1        0.34      100.00
----------------+-----------------------------------
          Total |        298      100.00


      Number of |
 staff: Cleaner |      Freq.     Percent        Cum.
----------------+-----------------------------------
              0 |         82       27.52       27.52
              1 |        179       60.07       87.58
              2 |         30       10.07       97.65
              3 |          6        2.01       99.66
              4 |          1        0.34      100.00
----------------+-----------------------------------
          Total |        298      100.00


      Number of |
     staff: CHW |      Freq.     Percent        Cum.
----------------+-----------------------------------
```

0	34	11.41	11.41
1	47	15.77	27.18
2	53	17.79	44.97
3	36	12.08	57.05
4	26	8.72	65.77
5	14	4.70	70.47
6	25	8.39	78.86
7	9	3.02	81.88
8	10	3.36	85.23
9	3	1.01	86.24
10	9	3.02	89.26
11	2	0.67	89.93
12	7	2.35	92.28
13	1	0.34	92.62
14	4	1.34	93.96
15	4	1.34	95.30
16	1	0.34	95.64
17	2	0.67	96.31
19	1	0.34	96.64
21	1	0.34	96.98
30	3	1.01	97.99
40	2	0.67	98.66
49	1	0.34	98.99
50	1	0.34	99.33
55	1	0.34	99.66
56	1	0.34	100.00
Total	298	100.00	

Number of staff: CHA	Freq.	Percent	Cum.
0	20	6.71	6.71
1	271	90.94	97.65
2	7	2.35	100.00
Total	298	100.00	

Total staff at HP (excluding volunteer CHWs)	Freq.	Percent	Cum.
0	4	1.34	1.34
1	59	19.80	21.14
2	69	23.15	44.30
3	119	39.93	84.23
4	38	12.75	96.98
5	5	1.68	98.66
6	3	1.01	99.66
7	1	0.34	100.00
Total	298	100.00	

Network reception quality	Freq.	Percent	Cum.
All of the time	38	12.75	12.75
Most of the time	39	13.09	25.84
Some of the time	103	34.56	60.40
Rarely	59	19.80	80.20
Never	59	19.80	100.00
Total	298	100.00	

Variable	Obs	Mean	Std. Dev.	Min	Max
workHours	298	42.38926	8.856287	9	98

=1 if has supervisor	Freq.	Percent	Cum.
No	9	3.02	3.02
Yes	289	96.98	100.00
Total	298	100.00	

Frequency of communication with supervisor	Freq.	Percent	Cum.
5-7 days/wk	83	27.85	27.85
3-4 days/wk	30	10.07	37.92

1-2 days/wk	45	15.10	53.02
Monthly	57	19.13	72.15
Once per month	44	14.77	86.91
Less than monthly	17	5.70	92.62
Never	13	4.36	96.98
(Skipped)	9	3.02	100.00
Total	298	100.00	

Location of supervisor	Freq.	Percent	Cum.
Same facility as CHA	102	34.23	34.23
Different facility	186	62.42	96.64
Don't know	1	0.34	96.98
(Skipped)	9	3.02	100.00
Total	298	100.00	

Beans: Motivations

Variable	N	Mean	SD	Min	Max
Motivation: FutureCareer	298	7.79	5.76	0.00	37.00
Motivation: PaysWell	298	2.35	3.97	0.00	28.00
Motivation: InterestingJob	298	6.76	5.22	0.00	27.00
Motivation: UsefulSkills	298	10.88	6.12	0.00	50.00
Motivation: ServeCommunity	298	18.35	9.06	0.00	50.00
Motivation: StableIncome	298	1.93	3.25	0.00	25.00
Motivation: Status	298	1.95	3.28	0.00	21.00
Total beans: Job motivation	298	50.00	0.00	50.00	50.00

Beans: Overall activities

Variable	N	Mean	SD	Min	Max
Overall activities: HHVisits	298	15.79	5.80	0.00	31.00
Overall activities: Traveling	298	11.10	6.07	0.00	29.00
Overall activities: HP	298	7.85	5.22	0.00	30.00
Overall activities: Community	298	10.99	4.60	0.00	29.00
Overall activities: Supervisor	298	4.27	3.32	0.00	18.00
Total beans: Overall activities	298	50.00	0.06	49.00	50.00

Beans: HH visit tasks

Variable	N	Mean	SD	Min	Max
HH visit tasks: Greeting	298	4.64	3.30	0.00	23.00
HH visit tasks: Sick	298	5.16	3.04	0.00	16.00
HH visit tasks: HealthProfile	298	5.89	3.71	0.00	18.00
HH visit tasks: Questions	298	6.45	3.59	0.00	20.00
HH visit tasks: Counseling	298	10.36	4.44	0.00	24.00
HH visit tasks: Inspection	298	9.82	4.66	0.00	30.00
HH visit tasks: DocumentationCom	298	6.67	3.96	0.00	22.00
Total beans: HH visit tasks	298	48.99	7.03	0.00	50.00

Beans: HP tasks

Variable	N	Mean	SD	Min	Max
HP tasks: OPD	297	11.91	5.96	0.00	28.00
HP tasks: Meds	297	9.48	5.61	0.00	39.00
HP tasks: ANC	297	7.93	5.74	0.00	26.00
HP tasks: DocumentationHP	297	9.13	6.53	0.00	50.00
HP tasks: Cleaning	297	5.18	4.21	0.00	18.00
HP tasks: Deliveries	297	1.99	3.18	0.00	16.00
Total beans: HP tasks	298	45.47	14.38	0.00	50.00

Beans: Community tasks

Variable	N	Mean	SD	Min	Max
Community tasks: Campaigns	297	11.33	5.60	0.00	31.00
Community tasks: HealthTalks	297	17.73	5.78	0.00	50.00

Community tasks: School	297	10.12	4.70	0.00	24.00
Community tasks: NHC	297	10.23	4.80	0.00	30.00
Total beans: Community tasks	298	49.25	5.82	0.00	50.00

Beans: Job challenges

Variable	N	Mean	SD	Min	Max
Job challenges: TravelTime	298	15.09	8.35	0.00	50.00
Job challenges: FindClients	298	7.86	6.91	0.00	31.00
Job challenges: HHNoTime	298	1.22	2.88	0.00	15.00
Job challenges: EquipLow	298	11.24	8.95	0.00	50.00
Job challenges: StaffLow	298	6.81	6.84	0.00	33.00
Job challenges: SupervisionLow	298	5.35	6.60	0.00	35.00
Job challenges: TrainingLow	298	1.51	3.72	0.00	31.00
Job challenges: DemandsOther	298	0.92	3.04	0.00	27.00
Total beans: Job challenges	298	50.00	0.00	50.00	50.00

Frequency of nighttime work	Freq.	Percent	Cum.
5-7 days/wk	28	9.40	9.40
3-4 days/wk	53	17.79	27.18
1-2 days/wk	66	22.15	49.33
Monthly	39	13.09	62.42
Once per month	10	3.36	65.77
Less than monthly	14	4.70	70.47
Never	88	29.53	100.00
Total	298	100.00	

Current job satisfaction	Freq.	Percent	Cum.
Very happy	156	52.35	52.35
Happy	121	40.60	92.95
Neither happy nor unhappy	7	2.35	95.30
Unhappy	6	2.01	97.32
Very unhappy	3	1.01	98.32
Don't know	3	1.01	99.33
(Truly missing)	2	0.67	100.00
Total	298	100.00	

5-10 YEARS ENVISION

Variable	N	Mean	SD	Min	Max
5-10 years envisions being: CHA	289	0.19	0.40	0.00	1.00
5-10 years envisions being: Nurse	292	0.54	0.50	0.00	1.00
5-10 years envisions being: EHT	291	0.67	0.47	0.00	1.00
5-10 years envisions being: CO	293	0.53	0.50	0.00	1.00
5-10 years envisions being: Doctor	290	0.23	0.42	0.00	1.00
5-10 years envisions being: Teacher	294	0.01	0.08	0.00	1.00
5-10 years envisions being: Farmer	292	0.22	0.41	0.00	1.00

=1 if reports submitting any SMS visits	Freq.	Percent	Cum.
No	1	0.34	0.34
Yes	297	99.66	100.00
Total	298	100.00	

=1 if still has Samsung phone	Freq.	Percent	Cum.
No	16	5.37	5.37
Yes	282	94.63	100.00
Total	298	100.00	

=1 if Samsung phone still working	Freq.	Percent	Cum.
No	30	10.07	10.07
Yes	252	84.56	94.63
(Skipped)	16	5.37	100.00
Total	298	100.00	

=1 if still has Nokia phone	Freq.	Percent	Cum.
No	13	4.36	4.36
Yes	285	95.64	100.00
Total	298	100.00	

=1 if Nokia phone still working	Freq.	Percent	Cum.
No	55	18.46	18.46
Yes	228	76.51	94.97
Don't know	2	0.67	95.64
(Skipped)	13	4.36	100.00
Total	298	100.00	

=1 if has been paid	Freq.	Percent	Cum.
No	122	40.94	40.94
Yes	176	59.06	100.00
Total	298	100.00	

=1 if remembers pay date	Freq.	Percent	Cum.
No	7	2.35	2.35
Yes	169	56.71	59.06
(Skipped)	122	40.94	100.00
Total	298	100.00	

Date of payment	Freq.	Percent	Cum.
10jan2013	1	0.34	0.34
18jan2013	1	0.34	0.67
20jan2013	1	0.34	1.01
22jan2013	1	0.34	1.34
24jan2013	1	0.34	1.68
01feb2013	1	0.34	2.01
02feb2013	3	1.01	3.02
04feb2013	3	1.01	4.03
05feb2013	2	0.67	4.70
07feb2013	1	0.34	5.03
09feb2013	1	0.34	5.37
10feb2013	3	1.01	6.38
11feb2013	2	0.67	7.05
13feb2013	1	0.34	7.38
14feb2013	2	0.67	8.05
15feb2013	2	0.67	8.72
16feb2013	1	0.34	9.06
17feb2013	1	0.34	9.40
20feb2013	5	1.68	11.07
21feb2013	3	1.01	12.08
22feb2013	3	1.01	13.09
23feb2013	5	1.68	14.77
24feb2013	4	1.34	16.11
25feb2013	5	1.68	17.79
26feb2013	10	3.36	21.14
27feb2013	4	1.34	22.48
28feb2013	11	3.69	26.17
01mar2013	4	1.34	27.52
02mar2013	3	1.01	28.52
03mar2013	3	1.01	29.53
04mar2013	5	1.68	31.21
05mar2013	4	1.34	32.55
06mar2013	4	1.34	33.89
07mar2013	1	0.34	34.23
08mar2013	3	1.01	35.23
10mar2013	1	0.34	35.57
12mar2013	1	0.34	35.91
13mar2013	1	0.34	36.24
14mar2013	2	0.67	36.91

15mar2013	2	0.67	37.58
16mar2013	2	0.67	38.26
17mar2013	1	0.34	38.59
19mar2013	1	0.34	38.93
20mar2013	1	0.34	39.26
21mar2013	1	0.34	39.60
22mar2013	4	1.34	40.94
23mar2013	2	0.67	41.61
25mar2013	4	1.34	42.95
26mar2013	1	0.34	43.29
27mar2013	2	0.67	43.96
28mar2013	1	0.34	44.30
30mar2013	1	0.34	44.63
06apr2013	1	0.34	44.97
07apr2013	1	0.34	45.30
08apr2013	1	0.34	45.64
12apr2013	1	0.34	45.97
14apr2013	1	0.34	46.31
15apr2013	4	1.34	47.65
16apr2013	1	0.34	47.99
17apr2013	3	1.01	48.99
19apr2013	2	0.67	49.66
20apr2013	1	0.34	50.00
22apr2013	3	1.01	51.01
23apr2013	1	0.34	51.34
24apr2013	5	1.68	53.02
25apr2013	1	0.34	53.36
26apr2013	4	1.34	54.70
27apr2013	2	0.67	55.37
28apr2013	3	1.01	56.38
29apr2013	1	0.34	56.71
(Skipped)	129	43.29	100.00
Total	298	100.00	

Recruiting and Motivating Community Health Workers: A National Field Experiment in Zambia

USAID DIV Grant No. AID-OAA-G-12-00011

Milestone 4: Project Update Report

Nava Ashraf * Oriana Bandiera Scott S. Lee

In collaboration with Innovations for Poverty Action

April 8, 2015

1 Introduction

As described in our DIV project proposal, this project encompasses three distinct experiments: (1) a **recruitment** experiment designed to test the selection effect of advertising career incentives, (2) a **non-monetary awards** experiment designed to separately test the causal mechanisms that underlie nonmonetary awards, and (3) a **goal-setting and autonomy** experiment designed to test the performance effects of different goal-setting policies.

In this project update, we report on each of these experiments in turn. Regarding the **recruitment** experiment, from November 2014 to February 2015, we carried out a nationwide household survey in all 47 districts served by the first cohort of Community Health Assistants (CHAs). In addition, we have obtained administrative data on health facility utilization through the Government of Zambia's Health Management Information System (HMIS). These two data sources allow us to estimate the impact of recruitment with career incentives on health care utilization and health outcomes. In Section 2, we show that, relative to communities served by CHAs who were recruited under a "status quo" policy of no career incentives, communities served by CHAs recruited with career incentives exhibit higher rates of facility utilization and superior household health practices and outcomes.

*Corresponding author: Ashraf: HBS and NBER, nashraf@hbs.edu. Bandiera: Department of Economics and STICERD, LSE, o.bandiera@lse.ac.uk; Lee: Harvard Medical School and Harvard Business School, ssl@mail.harvard.edu.

In Section 3, we discuss our ongoing efforts for dissemination and scale-up of the career-oriented recruitment strategy.

The **non-monetary awards** experiment has been published, and we present a summary of the findings in Section 4. In brief, we find that both employer recognition and social visibility increase the performance of health trainees, but social comparison dampens performance. These results highlight the importance of anticipating the distributional consequences of incentives in settings such as health care, where the performance of every worker matters for social welfare.

Finally, the **goal-setting and autonomy** experiment is now complete, and we are currently performing data analysis. We present preliminary findings in Section 5. In brief, we find no average treatment effect of the two goal-setting treatments.

2 Impact of recruitment with career incentives on facility utilization and health outcomes

The CHA program leads to a substantial increase in the number of health staff: in the communities where CHAs are deployed, the number of health staff associated with the health post increases on average from 1.5 to 3.5. Given the size of the program relative to existing personnel counts and the magnitude of the treatment effect on visits and community mobilization meetings, we expect treatment to affect aggregate outcomes. To shed light on this we present data from the Ministry's HMIS administrative records on the number of individuals seen at government facilities as well as household survey data on health practices and outcomes in the study areas.

2.1 Impact on facility utilization

The Ministry of Health's HMIS administrative records are compiled by facilities' senior staff and transmitted to MoH via an electronic platform. Two level of facilities serve these communities: health centers and health posts.[1] The main remit of the CHA job is mother and child health, and CHAs are supposed to encourage women to give birth at the closest health center and to bring in children for regular visits and immunizations at the closest facility (health center or health post). The importance of institutional deliveries in this context cannot be understated: Zambia's maternal mortality rates are very high and health centers have the equipment and medical supplies that can prevent these deaths. Regular children's visits ensure that conditions such as diarrhea are treated before they become dangerous. Immunizations protect children from serious and potentially fatal illnesses.

[1] Health facilities in Zambia are structured according to a population-based hierarchy. Health posts are the first-level health facility for most rural communities and provide basic medical care (no inpatient or surgical services). Health centers, which typically serve a population encompassing four to five health posts, provide both outpatient and inpatient services, including labor and delivery and minor surgical procedures. District hospitals in turn encompass several health center catchment areas and are primarily focused on inpatient care.

To test whether the observed performance gap is associated with a change along these margins, we obtain information on institutional deliveries, children's visits, and immunizations for the period January 2011-June 2014 and estimate the following difference-in-difference specification:

$$y_{hdpt} = \alpha + \beta C_{hd} + \gamma A_t + \delta C_{hd} * A_t + Z_h\theta + E_d\phi + \rho_p + \xi_{hdpt}$$

where y_{hdpt} is the outcome in health facility h in district d and province p at quarter t.[2] h represents the lowest level of government facility to which the CHAs can refer their patients. This is the health post if it is operational in HMIS, and if not the closest health center. The only exception is childbirths that are always measured at the health center level, as that is where they are supposed to take place. C_{hd}=1 if facility h is located in a district where CHAs were recruited via career incentives. We have data for 14 quarters, equally divided before and after the CHAs' arrival, and A_t=1 after the CHAs' arrival (4th quarter of 2012). To minimize composition bias and to test for robustness to facility fixed effect models we restrict the sample to the facilities for which we have at least three observations before and after the CHAs' arrival.[3] Z_h is a vector of area characteristics, which includes the number of staff at the health post, cell network coverage, and the distribution of households between farms and villages. We control for the stratification variables, district-level high school graduation rate E_d, and provinces indicators ρ_p throughout. Standard errors clustered at the level of randomization—the district.

The parameter of interest is δ, the difference in differences between facilities in treatment and control districts before and after the CHA's arrival. Under the parallel trend assumption δ captures the effect of career incentives for CHAs on these outputs.

Table 1 shows that indeed, career incentives improved clinic utilization outputs. In particular, the number of women giving birth at the health center increases by 31% relative to the mean in control areas at baseline. Regarding child health, the number of children under age five visited increases by 24%, the number of children under five weighed increases by 23%, and the number of children under 12 months of age receiving polio vaccination increases by 20%. The effects on postnatal visits for women, BCG, and measles vaccinations are also positive and in the 8-15% magnitude range, but are not precisely estimated. Reassuringly, there are no significant differences between treatment and control areas in any of these outcomes before the CHAs' arrival: all the estimated β coefficients are small and not significantly different from zero.

[2]HMIS data should be transmitted to MoH monthly, but in practice (due to poor connectivity), reports are missing for some months and the information added to the following month. We aggregate the data at the quarterly level to smooth out monthly fluctuations due to this.

[3]This restriction keeps 77% of the health posts and 70% of the health centers in the sample.

2.2 Impact on household health practices and outcomes

To provide evidence on the effect of treatment on health practices and outcomes we survey households in 47 randomly chosen communities located in each of the 47 districts where the CHAs operate. We randomly choose 16 households in each community, surveying 738 in total.[4] As the main focus of the CHA job is mother and child health we only survey households that contain a child under age five years. The survey contains modules on health and sanitation knowledge, health practices, incidence of illnesses and anthropometrics for the youngest child. Knowledge, practices and illnesses are self-reported, deworming and immunization data are drawn from the child health card and anthropometrics are measured by trained enumerators. We interview the main carer of the child, which is their mother in 90% of the cases and either a grandparent or a sibling in the remaining 10%. All questions are drawn from the DHS Zambia questionnaire, with the exception of the health knowledge module which we designed based on the CHA curriculum, and mid-upper arm circumference, which the DHS does not measure.

Table 2 reports the estimates of:

$$y_{idp} = \alpha + \beta C_{id} + D_i\gamma + \delta E_d + \rho_p + \epsilon_{idp} \tag{2.1}$$

where y_{idp} is the outcome of child (or respondent) i in district d and province p, C_{id} equals 1 if child (or respondent) i lives in a district that is assigned to the career incentives treatment. D_i is a vector of child, respondent and household characteristics that include child age and gender, household size and number of assets, and the education level of the respondent. As above, we control for the stratification variables, district-level high school graduation rate E_d and provinces indicators ρ_p throughout and cluster standard errors at the district level.

Column 1 shows that the average respondent answers 75% of the knowledge questions correctly and this is does not differ by treatment status. In contrast, treatment affects all the health practices we collect information on. In particular, Columns 2 and 3 show that children under 2[5] living in treatment areas are 5 percentage points more likely to be breastfed, and their stools are 12 percentage points more likely to be safely disposed; these effects represent a 8% and 20% increase from the control group mean, respectively. Columns 4 and 5 show that treatment also increases the incidence of deworming treatments by 15% and the likelihood that the child is on track with the immunization schedule by 4.7 percentage points, which is 81% of the control group mean

[4] A complete sample would have been 752 households. The difference of 14 households is due to several factors. In some communities, safety concerns related to local political tensions forced the survey team to leave the community before completing surveying. In other communities, especially low-density communities where travel times between households could exceed one hour, the survey team was unable to find 16 eligible households within the allotted survey time. One household interview was lost due to malfunction of the mobile device on which the interview was recorded. The minimum number of households surveyed in a community was 13.

[5] WHO recommends breastfeeding until the age of two years.

(5.8%).[6] Importantly, the treatment affects the incidence of immunizations for children who are young enough to exposed to CHAs when their immunization period started (as shown in Column 5) but not for those that were too old to start the cycle when the CHAs started working. This echoes the findings in Table1 that show no difference in immunization rates between treatment and control areas before the CHAs started working.

Columns 6-8 measure treatment effects on the incidence of three main illness symptoms: fever, diarrhea and cough. These are fairly common as 47%, 26% and 45% of children in control areas had experienced them in the past two weeks. We find that treatment reduces the incidence of cough symptoms by 7 percentage points while leaving the others unchanged. Finally, Columns 9-12 show treatment effects on anthropometric measurements. We report weight-for-age z-scores and mid-upper arm circumference (MUAC). The combination of these two allows us to measure both chronic and acute malnutrition.[7] Following WHO's guidelines we use the -2SD and -3SD thresholds for weight-for-age z-scores to measure moderate and severe underweight, respectively, and 12.5cm and 11.5cm for MUAC to measure moderate and severe wasting, respectively (Food and Nutrition Technical Assistance Project, 2011). According to these measures, 21% of the children in control areas are underweight and 5% severely so. The incidence of wasting is much lower with 3.6% of the children exhibiting some wasting and 1.4% severe wasting. These data, which match the corresponding DHS figures for rural Zambia (Government of Zambia, 2014), suggest that these areas are characterized by high rates of chronic malnutrition but low rates of acute malnutrition.

The findings in columns 9-10 show that children in treatment areas are 5 percentage points less likely to be underweight (25% of the control group mean) and 3 percentage points less likely to be severely underweight (60% of the control group mean). In line with this, columns 11 and 12 show a large percentage reduction in wasting, but given the limited occurrence of this in our sample the effects are not precisely estimated.

Taken together, the findings in this and the previous section show that the selection effect of career incentives is strong enough to generate discernible differences in facility utilization, household behaviors, and child health outcomes.

[6] A child is defined to be on track if she has completed all immunizations required for her age. At age 3 months, this includes BCG, OPV 0-2, PCV 1-2, DPT-HepB-Hib 1-2, and rotavirus 1-2. At 4 months, this includes, additionally, OPV 3, PCV 3, and DPT-HepB-Hib 3. At 9 months, this includes OPV 4 if OPV 0 was not given, and measles 1. The immunization series is complete at age 18 months with measles 2. Finally, we consider a child to be on track for vitamin A supplementation if she has ever been supplemented.

[7] We elected not to measure height for two reasons. First, compared to weight, height measurement is more invasive, requiring, for children under two, laying the child down on a height board and having two enumerators hold the child while collecting the measurement. During survey piloting, many respondents (and the children themselves) balked at this procedure. Second, accurate height measurement is made difficult by high measurement error relative to standard effect sizes. For example, 1 millimeter is 12 percent of the increase in height-for-age typically observed in dedicated child nutrition programs (Dewey and Adu-Afarwuah, 2008).

3 Scaling up recruitment with career incentives

3.1 Cost-effectiveness of recruiting with career incentives

Our research shows that CHAs recruited with career incentives conduct more work (i.e., home visits, community meetings, patient consultations), and this performance difference translates into positive health impacts. With the Government of Zambia planning to scale-up the CHA program to 5,000 workers by 2018, we estimate that the adoption of our career-oriented recruitment strategy will translate into 315,000 additional household visits and 70,000 more community meetings over the next three years. Extrapolating globally, the lessons learned from this study could deliver life-saving health education through 85 million additional household visits and 19 million community health meetings—with little-to-no additional cost to governments.

In order to add greater confidence to these projections, we hope to carry out a full endline survey of all 162 CHA communities one year from now, at which point the CHAs will have been deployed for three years. This three-year follow-up period is critical, since, after two years of service, CHAs become eligible to access further training opportunities as well as to resign from their posts without penalty. In other words, since career incentives do not become relevant until after two years of service, tracking retention and performance over three years is critical for assessing the equilibrium impact of a career recruitment strategy.

3.2 Dissemination plan

The first cohort of 307 CHAs are active in 162 communities across 47 districts in 7 provinces of Zambia. Scaling of the CHA program began in 2012, with three additional waves of recruitment now having been undertaken. Currently, a total of 1,124 additional CHAs have undertaken training, 592 of which have been deployed and are now bringing critical health services to rural communities. As a direct result of our research, 100 percent of third- and fourth-wave CHAs were recruited using the career incentives strategy.

In March 2015, we officially presented the findings of the recruitment experiment in a series of high-level dissemination meetings with the President's Cabinet Office, the Ministry of Health, and the University of Zambia. Based on these meetings, senior officials within the Government of Zambia are now considering expanding the use of career incentives for the recruitment of other civil servants, such as bureaucrats and teachers. We are continuing to engage in close dialogue with senior government officials, in addition to continuing to track performance outcomes.

4 Non-monetary awards experiment

The results of the second in our series of three experiments have been published as Ashraf et al. (2014). In the context of the one-year training of the first cohort of CHAs, we designed a field ex-

periment to unbundle several mechanisms by which non-monetary awards may affect performance: by conferring employer recognition, by enhancing social visibility, and by facilitating social comparison. We find that employer recognition and social visibility increase performance, while social comparison reduces it, especially for low-ability trainees. These effects appear when treatments are announced and persist through training.

These results highlight the importance of anticipating the distributional consequences of incentives in settings such as health care, where the performance of every worker matters for social welfare. As such, it would not have been advisable to scale up non-monetary awards in the CHA context. We reported our findings to the Ministry of Health and recommended against using awards as an incentive tool in the deployment phase of the CHA program.

5 Preliminary analysis of goal-setting experiment

With DIV support, we created an SMS-based goal-setting system that allowed CHAs or their supervisors (depending on treatment assignment) to set goals for how many daily household visits to conduct each month. The theoretical motivation was to assess how providing health workers with autonomy and discretion over their performance (i.e., setting their own performance goals) affects their performance, relative to the more common scenario in which supervisors set performance targets.

The goal-setting experiment was launched in December 2012 and ran for 12 months. The number of CHAs choosing or receiving a goal each month is indicated in Figure 1. Two findings are of note. First, over the course of the experiment, there has been a steady decline in participation in the goal-setting system.[8] We believe this is largely due to technological factors. The goal-setting system accepts the submission of goals only from phone numbers that were registered at baseline as belonging to each CHA/supervisor. When CHAs/supervisors change phone numbers, the goal-setting system no longer accepts goal submissions without registration of the new phone number. However, due to the physical remoteness of the CHAs/supervisors, in most cases we were unable to ascertain these new phone numbers for registration.

Second, in all but one month, participation in the supervisor-sets-goals condition is lower than in the CHA-sets-goals condition. Again, this is due in part to technological factors. When a supervisor is transferred and a new supervisor is assigned, because the new supervisor has a new phone number, the goal-setting system is no longer able to accept goals from the new supervisor. Since supervisors can be transferred but CHAs cannot, this introduces a composition bias in the comparison of the two treatments. We are currently in the process of collecting data on the timing of supervisor transfers so that we can better control for the change in composition of supervisors.

[8] The steep drop-off in May 2013 was largely attributable to a two-week training that all CHAs were required to attend that month.

In the meantime, in the analysis that follows, we show results for both the first 12 months of the experiment, as well as the first four months only, when participation was relatively high in both treatment groups.

Table 3 reports average treatment effects. Columns 1-2 report estimates for cross-sectional regressions of the average value of daily home visit goals on treatment assignment. Conditional on setting a goal, CHAs set 12-15% lower home visit goals than supervisors. However, using a fixed-effects model with panel data, we find no significant difference between treatments on actual home visits (Columns 3-4). Although the experimental intervention only facilitating goal-setting for home visits, we also check for treatment effects on the CHAs' two other main tasks—community meetings and outpatient consultations—and find no significant effect on either.

Further analysis is ongoing. In particular, we are investigating distributional effects. For example, a zero average treatment effect may mask a positive effect for some CHAs (e.g., more capable or career-oriented CHAs) that is masked by a counterbalancing negative effect for other CHAs. Once the analysis is complete, we will assess cost-effectiveness and potential dissemination.

References

Ashraf, Nava, Oriana Bandiera, and Scott S. Lee, "Awards unbundled: Evidence from a natural field experiment," *Journal of Economic Behavior and Organization*, 2014, *100*, 44–63.

Dewey, Kathryn G and Seth Adu-Afarwuah, "Systematic review of the efficacy and effectiveness of complementary feeding interventions in developing countries," *Maternal & Child Nutrition*, 2008, pp. 24–85.

Food and Nutrition Technical Assistance Project, "Anthropometry: Assessing Children Under 5 Pocket Reference," 2011.

Government of Zambia, "Zambia Demographic and Health Survey 2013-14: Preliminary Report," Technical Report 2014.

Table 1: The effect of career incentives on health facility utilization.

Dependent variable: total over each quarter 2011:1-2014:2	institutional deliveries	postnatal (0-6 weeks) visits	children under 5 visited	children under 5 weighted	children under 1 receiving BCG vaccinations	children under 1 receiving polio vaccinations	children under 1 receiving measles vaccinations
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Career incentives	-1.796	-13.81	-80.19	-86.81	10.48	-1.335	0.747
	(10.29)	(9.534)	(142.0)	(133.2)	(12.07)	(9.075)	(10.05)
After	3.479	15.08**	55.40	102.7	-1.611	-1.643	-1.517
	(4.425)	(5.191)	(63.22)	(63.91)	(4.566)	(3.717)	(3.591)
Career incentives*After	14.68**	8.253	318.1**	284.3**	7.158	14.98**	11.47
	(6.322)	(9.562)	(98.05)	(110.2)	(8.906)	(4.803)	(7.255)
Area characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of dependent variable in control in year 1	46.7	49.9	1312.8	1261.5	89.8	73.9	73.6
Adjusted R-squared	0.331	0.203	0.246	0.246	0.146	0.143	0.108
Number of facilities	89	118	123	123	121	120	121
Number of observations	1269	1528	1618	1610	1518	1530	1535

Notes: OLS Estimates, standard errors clustered at the district level. Data source is the Health Management and Information System (HMIS) available monthly from January 2011 until June 2014. Health center and health post staff are required to submit monthly reports that summarize their activities at the health post/community level. These are aggregated at the quarter level in the regressions. The variable in Column (1) is defined at the health center level because health centers are equipped for child births and health posts are not. The variables in columns (2)-(7) are defined at the health post level if this reports data, at the health center otherwise. After=1 after September 2012 (from 2012:4 onwards), when CHAs started working. All regressions include the stratification variables (province dummies and share of high school graduates in the district). Area characteristics include: number of staff in the health post, geographical distribution of households in the catchment area, and an indicator variable that equals 1 if the CHA reports to have good cell network coverage most of the time or all the time.

Table 2: The effect of career incentives on household health practices and outcomes.

	Information	Health practices				Incidence of Illness			Anthropometrics			
Dependent variable	% of correct answers in medical knowledge test	=1 if child under 2 yr old is breastfed	=1 if child's stool are safely disposed	number of deworming treatments	=1 if child exposed to CHA is on track with immunization schedule	= if child has experienced fever in the last two weeks	= if child has experienced diahrrea in the last two weeks	= if child has experienced cough in the last two weeks	=1 if weight for age z score <2 SD (moderately or severely undernourished)	=1 if weight for age z score <3 SD (severely undernourished)	=1 if MUAC<12.5 (moderately or severely wasted)	=1 if MUAC<11.5 (severely wasted)
Career incentives	0.002	0.051**	0.121***	0.225*	0.047**	-0.003	0.037	-0.070**	-0.053*	-0.028*	-0.023	-0.014
	(0.010)	(0.023)	(0.039)	(0.129)	(0.020)	(0.037)	(0.027)	(0.033)	(0.030)	(0.015)	(0.015)	(0.014)
household controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
child controls	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Mean of dep var in control	.740	.641	.595	1.44	.058	.469	.255	.450	.210	.051	.036	.014
Adjusted R-squared	0.057	0.561	0.161	0.263	0.024	0.077	0.017	0.021	-0.006	0.003	0.018	0.017
N	738	613	736	659	462	731	731	731	582	582	581	581

Notes: OLS estimates, standard errors clustered at the district level. The medical knowledge test contains 14 questions on topics that CHAs are supposed to cover; these questions were drafted by the researchers in consultation with CHA program officials and the CHA curriculum Breastfeeding and stool disposal are self-reported. In line with UNICEF (2014), we define stools to be safely disposed if flushed in toilet/latrine. Deworming, immunisation data and schedule as reported in the child health card. A child is defined to be on track if they have completed all immunisations required for their age in months. The immunisation sample is restricted to children who were 3 months or younger (including unborn) when the CHAs started working. Thresholds for weight-for-age and MUAC are taken from WHO guidelines; following these, data are restricted to children between 6-59 months. Household controls include size, education level of the respondent, number of assets. Child controls include age and gender. All regressions include the stratification variables.

Figure 1: Goal submissions by treatment and month.

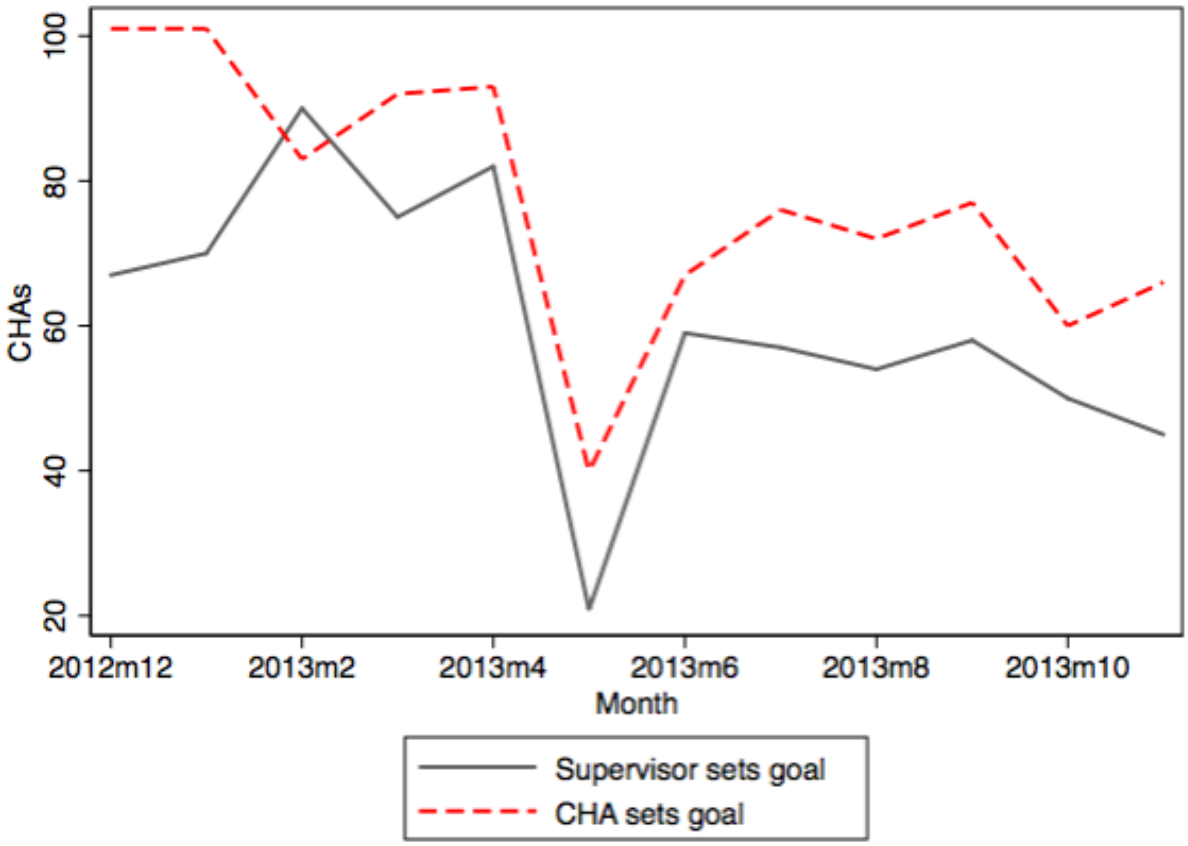


Table 3: Average treatment effects of goal-setting experiment.

	Home visit goals		Home visits		Community meetings		Outpatient consultations	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-intervention period			16.13***	4.847**	1.271***	0.819**	84.43***	33.00**
			(3.29)	(2.287)	(0.275)	(0.320)	(17.07)	(13.70)
CHA-sets-goal x Post-intervention period	-0.697**	-0.865***	1.994	1.439	-0.213	-0.567	25.81	6.464
	(0.314)	(0.293)	(3.625)	(2.458)	(0.281)	(0.434)	(24.35)	(15.51)
Mean in supervisor-sets-goal	5.81	5.64	30.52	18.89	2.37	2.12	97.18	66.23
Unit	CHA	CHA	CHA-month	CHA-month	HP-month	HP-month	HP-month	HP-month
Pre-intervention months	4	4	4	4	4	4	4	4
Post-intervention months	4	12	4	12	4	12	4	12
Baseline controls	Y	Y	--	--	--	--	--	--
CHA/HP fixed effects	--	--	Y	Y	Y	Y	Y	Y
Month fixed effects	--	--	Y	Y	Y	Y	Y	Y
Adjusted R^2	0.085	0.163	0.403	0.410	0.540	0.090	0.465	0.430
Observations	239	260	2456	4912	1296	2592	1296	2592

Notes: Table reports OLS estimates for cross-sectional (Columns 1-2) and fixed-effects panel regressions (Columns 3-8). Column headings are dependent variables. Standard errors clustered at the supervisor level are in parentheses. Baseline controls in Columns 1-2 include the stratification variables (cellular network coverage, number of staff at health post, population settlement pattern, recruitment experiment treatment assignment), district-level educational attainment, province, and whether the CHA or supervisor was reached by phone to announce the initial launch of the goal-setting experiment. The mean in the supervisor-sets-goal condition is reported for the post-intervention period. In Columns 3-8, the odd-numbered columns restrict the sample to the first four months of the experiment, whereas the even-numbered columns include a 12-month panel in the post-intervention period. Estimates are significant at the *10%, **5%, and ***1% level.

DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Awards Unbundled: Evidence from a Natural Field Experiment

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

Citation	Ashraf, Nava, Oriana Bandiera, and Scott S. Lee. "Awards Unbundled: Evidence from a Natural Field Experiment." Journal of Economic Behavior & Organization 100 (April 2014): 44–63.
Published Version	DOI: 10.1016/j.jebo.2014.01.001
Accessed	May 8, 2015 2:15:31 PM EDT
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:13348076
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP

(Article begins on next page)

Awards Unbundled: Evidence from a Natural Field Experiment

Nava Ashraf, Oriana Bandiera, and Scott Lee[1]

Abstract

Organizations often use non-monetary awards to incentivize performance. Awards may affect behavior through several mechanisms: by conferring employer recognition, by enhancing social visibility, and by facilitating social comparison. In a nationwide health worker training program in Zambia, we design a field experiment to unbundle these mechanisms. We find that employer recognition and social visibility increase performance while social comparison reduces it, especially for low-ability trainees. These effects appear when treatments are announced and persist through training. The findings are consistent with a model of optimal expectations in which low-ability individuals exert low effort in order to avoid information about their relative ability.

JEL codes: D84, D83, J33, M52, O15.

Keywords: awards, social comparison, optimal expectations, incentives.

[1]Ashraf: HBS and NBER, nashraf@hbs.edu; Bandiera: Department of Economics and STICERD, LSE, o.bandiera@lse.ac.uk; Lee: Harvard Medical School and HBS, slee@hbs.edu. This research is the fruit of a collaboration with the Ministry of Health, Zambia, to whose staff we owe a tremendous debt, especially Mutinta Musonda and Miriam Libetwa. We also thank Alice Nyirenda and the teachers of the CHA Training School for their patience and valuable feedback. Madeleen Husselman, Mardieh Dennis, Manpreet Singh, Katherine Otto, Johann Blauth, and, especially, Kristin Johnson provided skilled research assistance. We thank Charles Angelucci, Ed Glaeser, Francesca Gino, Daisuke Hirata, Kelsey Jack, Bryce Millett Steinberg, Emily Oster and Michele Rigolizzo, and participants at the Stanford Institute for Theoretical Economics, Psychology and Economics session, for very helpful input. Financial support by the International Growth Centre is gratefully acknowledged.

1. Introduction

> What are the advantages which we propose by that great purpose of human life which we call bettering our condition? To be observed, to be attended to, to be taken notice of with sympathy, complacency, and approbation, are all the advantages which we can propose to derive from it.
>
> —Adam Smith, "Of the Origin of Ambition, and of the Distinction of Ranks," The Theory of Moral Sentiments (1759)

The innate human desire for approbation might make status awards a cost-effective tool to incentivize good performance (Besley and Ghatak, 2008; Moldovanu et al., 2007). Awards can motivate employees to exert effort in order to gain recognition and visibility, both of which are free for the employer to bestow but valuable to the employee. However, given that awards derive their value from their scarcity, they inevitably facilitate social comparisons, which might be *demotivating* to employees.[2]

Our goal in this paper is to "unbundle" awards—that is, to provide evidence on the mechanisms that underlie their effectiveness and potential harm. We conduct a natural field experiment to separately identify channels through which awards can affect behavior, unbundling the effect of social comparison through the disclosure of rank information, from the effect of employer recognition and social visibility.

We study the effect of awards in the context of a nationwide training program for health workers in Zambia. Our agents are 314 health workers recruited from 162 rural communities and brought to professional school for a one-year training program aimed at teaching community-based health care. After training, trainees will be employed by the Ministry of Health and deployed to their communities of origin, where they will become the first point of contact for health services. Incentivizing learning is key in this context because trainees have

[2]Lazear (1989) describes the tradeoff in relative performance evaluation: it could motivate employees to work harder, but could also create an excessively competitive work environment and decrease employee morale. Major et al. (1991)'s review of the literature in social psychology provides evidence on the demotivating effect of social comparisons. A related literature in management emphasizes the importance of concealing relative performance information to improve employee motivation (Milkovich and Newman, 1996).

no previous medical training; thus, the skills they learn will determine their effectiveness in the field.[3]

During the training program, trainees take courses on several topics, on which they are tested at baseline (at the beginning of the year) and at the end of each course. The field experiment randomly allocates trainees to two broad classes of treatments (in addition to control): those that only provide information on trainees' relative performance, and those that also offer awards. After each exam, trainees in the control group receive a letter from the school reporting their absolute score and their value added, measured as improvement over their baseline score for the given course. Trainees in the "private social comparison" treatment (T1) receive the same letter with added information on their rank in the class distribution of both absolute score and value added. Trainees in the "public social comparison" treatment (T2) receive the same letter as in the previous treatment as well as the names of the top four performers in the class (top two by absolute score and top two by value added).

The third and fourth treatments add awards to rank information. Awards are given to the trainees with the top two scores and those with the top two most improved scores (from baseline). The latter ensures that weaker trainees have a chance to win and are therefore motivated by the award. In the "employer recognition award" treatment (T3), the top four performers receive a congratulatory letter from the Ministry of Health. In the "social visibility award" treatment (T4), one of the top four performers is randomly selected to be featured in an interview, which is printed along with the candidate's photo in a newsletter distributed back to their community of origin. Under a linearity assumption the difference between each of the award treatments and the "public social comparison" treatment isolates the effect of awards from the effect of the social comparisons they inevitably create.

Our setting has three key features that make it ideal for the purpose of this experiment. First, since trainees take four courses during the study period and treatments are announced at the beginning of the first course, we can assess whether they change their behavior in anticipation of receiving rank information and awards or only after these have been provided. Second, during training, the performance of the health workers is measured by an institution (the school) that is different from their employer (the Ministry), and the health workers are

[3] A number of field experiments have evaluated the effect of financial incentives on student learning; the evidence of their efficacy is mixed (Fryer, 2011; Angrist and Lavy, 2009; Kremer et al., 2009; Leuven et al., 2010).

physically removed from their home communities. This allows us to separate the effect of information on relative performance (provided by the school) from that of the employer's recognition and from visibility to one's social circle (the home community). In most settings, the employer measures and provides information on performance, such that the provision of information necessarily entails some recognition. The fact that trainees are distant from their communities is similarly useful, as no treatment other than the social visibility award can be used to enhance visibility within their social circles. In most settings in which agents are co-located with their social network, any treatment that reveals their rank in the distribution could potentially be used to enhance visibility.

Third, performance in this setting is uni-dimensional (trainees are solely meant to attend classes and study the topics on the syllabus), and thus not subject to a multitasking problem in the face of additional incentives. Moreover, performance can be measured objectively and precisely by test scores. Value added in test scores is a good measure of learning, as is often the case when knowledge at baseline is very limited (Muralidharan and Sundararaman, 2011). Critically, we can show that in this context exam performance is correlated with future performance in the field (Ashraf et al., 2013b).

The analysis reveals that social comparison and awards have opposite effects on performance. Compared to trainees in the control group, the "private social comparison" treatment significantly reduces test scores by 0.31 standard deviations, and the "public social comparison" treatment reduces it by 0.38 standard deviations. Importantly, the two social comparison treatments reduce performance as soon as they are announced—i.e. before trainees get the first letter with their rank information. A likely explanation is that individuals value the *belief* that they have high relative ability, and the anticipatory utility this provides. They may thus prefer to exert low effort in order to decrease the informativeness of the ranking signal. This is akin to refusing to take a medical test for a disease, so as to justify holding an optimistic belief about one's health status (Oster et al., 2013), and is consistent with the literature on belief utility and information avoidance (Bénabou and Tirole (2002); Köszegi (2002)). In our context, the negative effect due to information avoidance seems to dominate

the potential positive effect of competition among trainees (Charness and Grosskopf, 2001; Freeman and Gelber, 2010).[4]

Adding awards to rank information significantly improves performance. Compared to trainees in the "public social comparison" treatment, mean scores of trainees in the "employer recognition award" and "social visibility" treatments are 0.38 and 0.44 standard deviations higher, respectively. Recognition from one's employer can increase performance if agents have career concerns (Dewatripont et al., 1999) or preferences for reciprocity (Fehr and Schmidt, 1999). The net effect of either type of award is nil: because the positive effects of employer recognition and social visibility are nullified by the negative effect of providing information on relative ranks, trainees in the two award treatments perform as well as trainees in control.[5]

Quantile treatment estimates show that both the negative effect of social comparison and the positive effects of recognition and visibility are stronger on the left tail of the conditional productivity distribution, and both are zero at the top two deciles. In line with this, we also find that these negative and positive effects are stronger for low-ability trainees, and zero for high-ability trainees, where ability is measured by baseline test score. The fact that the negative effect of ranking is stronger for the weakest trainees is intuitive, as these are more likely to receive a negative signal about their skills.[6] That the positive effects of recognition and visibility are also stronger for the weakest trainees may be due to the fact that, since awards are given to trainees with the highest value added (the "most improved"), those who start at the bottom have a better chance to win. That the effects are zero for the top two deciles is consistent with the fact that scores are capped, such that the highest performing trainees have little room for improvement. Evidence from settings with no cap on performance suggests that, in contrast, awards are most effective at the top of the distribution (Nalbantian and Schotter, 1997; Müller and Schotter, 2010; Bandiera et al., 2013; Leuven et al., 2010).

[4]The empirical evidence on the ex-post effect of providing rank information is markedly mixed. Tran and Zeckhauser (2012), Azmat and Iriberri (2010), and Dur et al. (2013) show positive effects of rank information on performance, while Bandiera et al. (2013) and Barankay (2012) find that this information reduces productivity.

[5]This echoes the findings of Bandiera et al. (2013) who show the impact of receiving information on relative rankings can offset the positive impact of monetary prizes. In their setting, however, the effect is driven by changes in team composition.

[6]Tran and Zeckhauser (ibid.) find that private rank disclosure motivates high-ability more so than low-ability trainees, whereas Azmat and Iriberri (ibid.) find uniformly positive effects across the distribution.

An important implication of these findings is that, due to the negative effect of social comparison, awards might increase the dispersion of performance by weakening the weakest. In our setting, this may matter just as much as—and perhaps more than—mean effects. Health workers wield considerable power to influence the utility of their patients, especially for the worse (e.g., missed diagnoses, incorrectly dosed medications, wrong-site amputations). Both because of the potential for harm and the government's mandate to ensure equity of services across populations, the distribution of performance during training in this field experiment is crucial. The findings thus suggest caution in using mechanisms that facilitate interpersonal comparisons in contexts in which worsening performance at the bottom of the distribution is costly. This is particularly germane in the policy domains of public service delivery, such as in health and education, where the use of awards is increasingly common (Mathauer and Imhoff, 2006; Ashraf et al., 2013a), but where distributional effects on agents' performance could have severe welfare consequences for those they are serving.

The paper is organized as follows. Section 2 describes the context and the experiment. Section 3 presents the empirical analysis. Section 4 interprets the findings in light of optimal expectations theory (Brunnermeier and Parker, 2005; Oster et al., 2013). Section 5 concludes.

2. Experimental Design

2.1. Context

In 2010, the Government of Zambia (GOZ) launched a national effort to create a new civil service cadre called the Community Health Assistant (CHA). In the program's first year, GOZ sought to recruit, train, and deploy approximately 300 Community Health Assistants across seven of Zambia's nine provinces.[7] Within these seven provinces, based on population density, GOZ chose the 48 most rural of the 58 constituent districts, and across these, GOZ identified 165 underserved communities, each with an average population of 3,500 individuals (Government of Zambia, 2011). From each community, the intention was to recruit two Community Health Assistants. The recruitment and selection process occurred

[7] This is the first generation of community health workers trained by the Government of Zambia. Although this paper does not evaluate the efficacy of community health workers, they have been shown to improve health outcomes in randomized controlled trials in other countries (Baqui et al., 2008; Bhandari et al., 2011; Spencer et al., 2011).

at the community level, with on-the-ground implementation coordinated by district health officials.[8]

In total, 314 individuals accepted GOZ's training offer and moved to Ndola, Zambia's second-largest city, to join a newly established training school. The training program lasted one year and was structured in a modular format (see Figure A.1).[9] At the beginning of the training, all trainees took a baseline exam which covered all the material that would be subsequently taught during the training year. After each course, trainees took an exam covering the material taught in that course.[10] During this time, the trainees engaged only in attending classes and studying for their exams. The training school was divided into five classrooms, each accommodating roughly 60 trainees. The school was led by a principal and staffed by ten full-time teachers. The trainees were not formally paid during the training year, but their tuition and room and board were covered by the Ministry. In addition, the participants were aware that wages upon completing training would be the same for all CHAs, and that opportunities for promotion would be available. The program is thus effectively training "on the job" and career concerns were likely at play.

Once deployed to the field, the CHAs were to routinely visit households and provide a variety of home-based services: basic medical care to any sick persons, health education and counseling, and referrals to nearby health facilities as needed. Two key features of the job illuminate how critical the training period was to subsequent performance. First, CHAs were expected to provide a very broad scope of health services to all age groups. Second, they were to do so with a great deal of autonomy. In contrast to nurses, whose job it classically is to implement a physician's orders, and who typically are not trained to diagnose and treat illnesses, the CHA is more like a physician, making decisions autonomously without direct supervision. The human capital required to perform such varied activities is substantial, and the one-year training was consequently critical.

[8]See Ashraf et al. (2013b) for details on the selection process of the Community Health Assistants.

[9]The training curriculum was designed in the months leading up to the training launch through a consultative process led by the Zambian Ministry of Health, with input from health educators, clinicians, and public health and development practitioners.

[10]Each course also had a "practical" component for which trainees visited field sites. Performance in this component was not tested.

2.2. Performance measurement

Since the trainees' only task during training was to attend classes and study the material taught therein, we measure performance by exam test scores. All exams were based on multiple choice questions created by external medical advisors based on the content of the official training textbook.[11] Grading was done electronically by the research team. After each course exam, each student's completed exam was returned to him or her, along with an answer key.

Several measures were taken to prevent cheating or gaming. First, all exams were administered under timed, proctored conditions. Second, each exam had four versions, in which the order of the answer choices for each question was randomly varied. The exam versions were distributed within a classroom in sequential fashion, such that no two neighboring trainees had the same exam version. Even if they were alike, the exam version was indicated discreetly in the lower corner of the exam, such that a student attempting to cheat by copying a neighbor's answers would have difficulty determining whether the neighbor's exam version was the same as hers.

2.3. Experimental design

We worked with the school administrators to randomly allocate trainees to five groups (one control and four treatments) of approximately 60 individuals each, stratified by average baseline exam score and other potential determinants of performance.[12] To minimize contamination across treatments, all trainees in a given treatment group were assigned to the same classroom, and classrooms were kept together for the entire duration of the experiment that lasted nine consecutive months. For each course, each classroom was co-taught by two teachers. The teachers rotated and were assigned to different classrooms after each course, using a schedule that was determined by the principal and by the researchers with the aim of ensuring even coverage of teachers across classrooms. Teachers and trainees in all groups used the same textbook that was developed by GOZ for the CHA training.

[11] Multiple choice questions are a standard question type on exams in Zambian secondary and tertiary education.

[12] We used the "T-min-max" method to balance the classrooms on gender, baseline exam score, any previous health experience, employment status and district-level recruitment strategy (different districts advertised the CHA position with different emphases on social vs. private benefits as described in Ashraf et al. (2013b)). For a discussion of this randomization method, see Bruhn and Mckenzie (2009).

The experimental treatments are as follows (see Figure 1 for a schematic diagram). After each exam, trainees in the control group receive a letter from the school reporting their absolute score and their value added over their baseline sub-score for the relevant course content after each exam (see Appendix Figure A.2). Trainees in the "private social comparison" treatment (T1) receive the same letter from the school with added information on their rank in the distribution of both absolute score and value added (see Appendix Figure A.3). Trainees in the "public social comparison" treatment (T2) receive the same letter from the school with added information on their rank and the names of the top four performers in the class—that is, the top two by absolute score and the top two by value added (see Appendix Figure A.4).[13] We include the "most improved" category based on theoretical and empirical evidence that multiple prizes at different points in the distribution are more motivating across the distribution than a single prize (Moldovanu and Sela, 2001; Freeman and Gelber, 2010; Dur et al., 2013). Importantly, trainees did not know that they could have won "value added" prizes when they took the baseline test, so there is no scope for gaming by obtaining a low score at baseline.

Trainees in the "employer recognition award" treatment (T3) receive the same letter as those in the "public social comparison" treatment, and, in addition, the top four performers receive a congratulatory letter from the Ministry of Health, handwritten and signed by the director of the department that runs the CHA program (see Appendix Figure A.5 for an example). Trainees in the "social visibility award" treatment (T4) receive the same letter as those in the "public social comparison" treatment, and, in addition, one of the top four performers is randomly selected to be featured in an interview that is printed together with the candidate's photo in a newsletter that is distributed back to their community of origin (see Appendix Figure A.6 for an example).

The timeline of the experiment is as follows. Trainees took four sequential courses during the experimental period (covering 9.5 out of the 12 months of training) and sat exams at the end of each course. Courses varied in duration from two weeks (course 2) to four months (course 1). At the beginning of the first course, trainees were told the content of the letters

[13]The design decision to make comparisons only at the very top public was done in consultation with senior teachers who were concerned about severe demotivating effects of publicizing the lowest performers. Although ranking throughout the entire distribution is publicly displayed in many professional schools in Zambia, there is no consensus on whether this is helpful or harmful.

that they would receive after each exam, and a reminder was delivered towards the middle of the same course. In an assessment given to all trainees after the reminder announcement, 79% of CHAs responded to the question, "How clear do you find the information presented in the letters?," with "very clear" and an additional 14% with "somewhat clear." In addition, trainees in the two awards treatments were shown sample employer recognition letters and community newsletters, respectively.

Table 1 shows balance across the five groups on a number of variables including baseline exam score, English exam score, gender, age, health experience and employment status. Only 3 out of 24 (12%) pairwise differences are different from zero at the 10% level, as expected by chance.

During the course of the training year, 6 of 314 (1.9%) trainees dropped out of the program. These trainees were distributed across four of the five treatment conditions (one in the control group, one in T2, three in T3, and one in T4) and have been excluded from our analysis. Due to the very low rate of attrition, any differences between the attrited and the non-attrited, or between treatment conditions as a result of the attrition, are indistinguishable from random error.

3. Analysis

3.1. Methodology

During the experimental study period, trainees attend four courses and take an exam at the end of each. Trainees in all treatments take the same courses and complete the same exams at the same time.

To evaluate treatment effects we estimate the following model using panel data at the trainee-course level:

$$score_{ic} = \alpha + \sum_{t=1}^{4} \beta_t T_i^t + \gamma B_{ic} + X_i\delta + I_{ic}\eta + \varepsilon_{ic} \tag{1}$$

where $score_{ic}$ is trainee i's test score in course c normalized by the mean and standard deviation of test scores for the same course in the control group; treatment effects are thus measured in standard deviation units. B_{ic} is trainee i's score in the baseline exam content for course c that was administered at the start of the training program. The difference between the "post" exam score and the baseline score is used to measure the value added that is

reported in the treatment letters and to rank trainees for the "most improved" awards. X_i are individual characteristics that include all stratification variables as well as age and trainee i's test score for an English language test. I_{ic} are teacher-specific traits such as teacher ratings (as reported anonymously by the trainees) and expertise in the subject matter of course c. Since all trainees take the same courses at the same time, different trainees have different teachers for the same course. The teacher rotation schedule was determined by the principal and by the researchers with the aim of ensuring even coverage of teachers across classes.

Standard errors are clustered at the trainee level, as trainee-specific unobservables in the error term create correlation within trainee. Since courses are of different durations and trainees have more time to exert learning effort the longer the course is, we weight observations by course duration.

The parameters β_t measure the causal effect of treatment t vis-à-vis the control group under the identifying assumption of no contamination across treatments. Contamination can occur if the response to treatment j is affected by the knowledge that treatment k exists. For instance, trainees might respond differently to being given information on relative rankings if they know that other trainees are also getting employer recognition awards while they are not. To minimize the risk of contamination, we allocate trainees in different treatments to different classes and keep classes together for the duration of the experiment. Trainees were told that other classrooms may receive different types of letters with their their exam scores, as this was the pilot year of the government program and different classrooms were trying different things. Trainees were not told that this was a research experiment, thus mitigating potential experimenter demand effects. Reassuringly, no student ever complained or raised the issue of different treatments for the entire duration of the experiment. Despite these precautions, trainees in the non-award treatments could have come across the award recognition letter and community newsletter given to trainees in other treatments. This, however, could have occurred only after the awards were distributed—that is, after the first test. To provide evidence on the practical relevance of contamination, we estimate equation (1) using scores from the first exam only, which was taken before trainees could have seen letters given to their colleagues in other treatments.

To separate the different mechanisms through which awards can affect performance, treatments are designed to be cumulative so that the "public social comparisons" treatment (T2) also contains information about relative ranks (T1), and the two awards treatments (T3

and T4) contain the same information as T2. Under the assumption that the effect of each component does not interact with the effect of the others, the net effect of each additional component can be identified by the appropriate linear combination of β_t estimators. For instance, the net effect of employer recognition is given by $\beta_3 - \beta_2$. We report all relevant linear combinations at the foot of each table.

3.2. Average treatment effects

Table 2 reports the estimates of equation (1). Test scores are normalized by the mean and standard deviation of the control group for each test so that treatment effects are measured in standard deviation units. Columns (1) and (2) estimate average treatment effects with and without teacher characteristics (average teacher rating and whether at least one of the teachers had specific expertise in the subject matter).

Three findings are of note. First, trainees in treatments 1 (private social comparison) and 2 (public social comparison) perform significantly worse than trainees in the control group. Estimates from the baseline specification in Column (1) show that giving private information about relative rankings lowers performance by 0.31 standard deviations, while rank information combined with a public list of the top four performers lowers performance by 0.38 standard deviations. By contrast, trainees in treatments 3 and 4, where top performers receive awards either in the form of a letter from the Ministry of Health or a profile in the organization's newsletter, performed the same as trainees in the control group.

Importantly, since trainees can do very little other than studying during the program, we can rule out that the drop in exam performance is associated with an increase of effort devoted to other tasks. In addition, in Ashraf et al. (2013b) we find that low performance on test scores during training is highly predictive of low performance in the field in CHA work. This strongly suggests that, even if there were other tasks, there is no compensating effort, either on other tasks during training or in later job performance that can make up for negative performance during the training period.

Second, we can identify the net effect of each additional treatment component by differencing out common elements under the independence assumption discussed above. This exercise reveals that the difference between T2 and T1 is small and not statistically significant, suggesting that making the list of top performers public does not motivate individuals to work harder. This might be driven by the possibility that other trainees are not the natural peer group to whom these individuals compare themselves. Alternatively, given that in

T1, individual ranks were printed on a letter, top performers in T1 could have easily made themselves known even if the list was not public.

Similarly, we can identify the effect of employer recognition and social visibility by differencing out the effect of providing rank information. Column (1) of Table 2 shows that both mechanisms have a strong positive effect on performance. The difference between T3 and T2 is 0.38 standard deviations, whereas the difference between T4 and T2 is 0.44 standard deviations. Both effects are precisely estimated.

Third, the estimated treatment effects are not sensitive to the inclusion of teacher controls (i.e., average teacher rating in students' evaluations and whether at least one of the teachers had specific expertise in the subject matter). This allays the concern that differences among treatments could be due to correlated unobservables at the class-course level. Since including teacher controls reduces the sample as these were not collected for course 2, we use the specification without teacher controls in what follows.

Taken together, the findings indicate that the three mechanisms described at the outset of this paper are relevant in this setting, but their signs and magnitudes differ: social comparisons (whether private or public) weaken motivation whereas employer recognition and social visibility strengthen it.[14] Thus, depending on the size of these opposing effects in a given context, the net effect of relative performance-based awards is uncertain.

3.3. Timing

Since treatments are announced at the beginning of course 1, we are able to identify whether their effect differs before and after the first round of letters containing rank information and awards are handed out. This test can help us shed light on why individuals change their behavior in response to rank information. For instance, if individuals are uncertain about the production function of test scores and use rank information to update on the marginal return of their effort, we would expect them to respond differentially before and after the information is provided. In contrast, if individuals know the production function and intrinsically care about their performance rank, we would expect their responses to be the same before and after.

[14]It could be argued that social visibility (in our case, visibility in one's home community) is a more powerful motivator in this setting; indeed, while the estimated effects are similar, the underlying treatment strength is different, as all top four performers receive the employer recognition award, whereas only one of them is randomly selected for the social visibility award.

To test whether responses vary before and after rank information is first disclosed, we estimate equation (1) augmented by the interactions of all treatment indicators with an indicator variable that takes value 0 for the first test (i.e., before treatment letters were distributed) and 1 for the following tests. Columns (3) and (4) of Table 2 report the estimated treatment effects on performance in the first test and performance in the following tests, respectively. Column (5) reports the p-value of the null hypothesis that these are equal.

The findings show that social comparisons, whether private or public, lead to a similar reduction in performance in the first and subsequent tests. Thus, this effect cannot be driven by "demotivation" in the traditional sense, or updating on the marginal return to effort, since trainees' performance dropped *in anticipation* of receiving rank information. Likewise, the findings are at odds with the assumption that trainees have competitive preferences (Charness and Grosskopf, 2001; Freeman and Gelber, 2010); otherwise, we should have observed an *increase* in effort in anticipation of receiving rank information. In Section 4, we describe a model of optimal expectations that is consistent with this pattern, in which trainees choose low effort (i.e., self-sabotage) to avoid rank information that delivers a signal about their relative ability.

In addition, columns (3) and (4) cast doubt on the relevance of the concern that responses to treatments 1 and 2 might be contaminated by the awareness of the other two treatments. Indeed, we show that providing rank information with or without a public list of top performers reduces performance by the same amount even before trainees in these treatments were likely to have become aware of the other treatments.[15]

The findings also show that both the employer recognition award and the social visibility award are effective at increasing performance conditional on rank information both before and after the first exam. The effect of the recognition award becomes weaker after the first exam, possibly because individuals revise their chance of winning downwards once the first round of rank information is revealed. We do not, however, find a similar pattern with respect to the social visibility award. We will return to this below when we allow responses to differ by baseline ability levels.

[15]It was of course possible for trainees in the award treatments to tell their colleagues in other treatments before the first exam, and before the awards were distributed. However, these claims should have been more credible and hence contamination stronger after the awards letters became potentially visible to all. The fact that treatment effects are stable throughout courses casts doubt on this.

3.4. Distributional effects

Awards are likely to affect individuals at different points in the performance distribution differently, as the incentive power of awards should be stronger for those who have a greater chance of winning the award—in our case, more able trainees and those who have more potential to improve. Similarly, the effect of social comparison is likely to depend on whether, given their knowledge of their own ability and expectation about others', individuals expect to be ranked high or low.

To provide evidence on these distributional issues we estimate quantile treatment effects at each decile. Figure 2 shows this graphically, and Appendix Table A.1 reports the regression coefficients by decile. The estimated treatment effect at each decile is the difference in conditional test score between two statistical trainees—one in the treatment group and one in the control group—both positioned at the same decile of the distribution of test scores within her group. Figure 2 shows that both the negative effect of relative rankings and the positive effects of recognition and visibility are stronger on the left tail of the conditional test score distribution, and they gradually diminish to zero at the top two deciles. Standard errors reported in Appendix Table A.1 show that all effects are statistically different from zero until the seventh decile. Taken together, these findings indicate that information on relative ranks, with or without a public component, increases the dispersion of performance by reducing performance on the left tail.

To provide further evidence on this issue, Table 3 allows treatments to have heterogeneous effects by trainees' ability, measured by their score in the baseline exam. The estimates show that both the negative effect of relative rankings and the positive effects of recognition and visibility are stronger for low-ability trainees and zero for high-ability trainees. The fact that the negative effect of relative rankings is stronger for the weakest trainees is intuitive, as these are more likely to receive a negative signal about their skills. That the positive effects of recognition and visibility are also stronger for the weakest trainees is presumably due to the fact that, since awards are given to trainees with the highest value added (the "most improved"), those who start at the bottom have a better chance to win. That the

effects are zero for the top two deciles is consistent with the fact that scores are capped, such that the very best trainees have little room for improvement.[16]

Finally, Appendix Table A.2 allows treatments to have heterogeneous effects by trainees' ability and exam timing. This confirms that most effects are driven by trainees at the bottom of the ability distribution and that, as discussed above, rank information leads to a similar reduction in performance in the first and subsequent tests. The results also show some evidence that the positive effect of awards becomes weaker after the first exam for trainees in the bottom tercile while it becomes stronger for those in the middle tercile. These findings should, however, be taken with caution as samples are small and tests have low power.

4. Interpretation

Our results suggest that the prospect of receiving information about one's rank in the distribution makes trainees exert lower effort, whereas the possibility of receiving employer recognition or improving one's visibility in the community makes them exert higher effort. In this section, we attempt to interpret these findings in a unified utility maximization framework.

The fact that individuals put in more effort when it is announced that doing so might earn them an award is intuitive. The main challenge is to explain why individuals, especially those with low ability, reduce effort in the rank information treatments even before being told their rank. A growing literature models the possibility that information lowers utility leading to information avoidance, at the cost of taking worse decisions (Oster et al., 2013; Brunnermeier and Parker, 2005; Köszegi, 2006; Stone, 2004). In these models, agents choose their beliefs optimally to maximize their lifetime utility, including an interim period of anticipatory utility arising from the belief. This implies that individuals may take actions to avoid a precise signal that has the potential to threaten their belief, for instance by choosing not to take a medical test.

[16] Across the entire sample, mean absolute post-test score in the highest decile is 90.0% and in the second-highest decile is 86.5%. The single highest absolute score across all exams was 96.1%. As our team, in consultation with medical training experts, wrote the exams to ensure quality and precision, we purposely included extremely difficult questions making scores of 100% very difficult. We thus take the evidence on top decile scores as support for the existence of capping, although we cannot rule out the possibility that other effects (such as a lack of desire to be singled out among peers, due to being taxed by expectations of assistance) may have also been occurring at the top decile.

In our setting, all agents know their individual baseline test scores and are told their individual exam scores in all treatment groups (including control). This, combined with the fact that individuals have a relatively good sense of their own individual ability since they all enter as adults, allows us to assume all that all individuals know their own ability. What individuals do not know is the ability of others—a reasonable assumption in our setting given that trainees entered the training school from an extremely diverse set of geographic and skill backgrounds. Realistically, individuals have a noisy expectation over their peers' ability.

In this setting, rank information (as provided in all treatment groups but not in control) increases the precision of the estimate of each individual's expectation over his peers' average ability. This would not affect behavior in standard models where agents only care about the expected value, but it might matter if agents prefer a fuzzy signal which can support optimistic beliefs about their relative ability.

We assume that agents are risk-neutral, and individual i with ability level a_i chooses effort e_i to maximize his expected utility:

$$\beta_i^t t(a_i, e_i) + \beta_i^{sc}(a_i - a_{-i}^*)(1 + T_i^{sc}\sigma(e_i)) + \beta_i^a T_i^A p(a_i, e_{i,} b_i)A - d(e_i) \tag{2}$$

The first term captures the effect of effort on learning proxied by test scores, which might provide utility either directly as individuals care about learning or through future wages. $\beta_i^t > 0$ is the weight individual i puts on learning, and learning is a function of effort and ability with $t_e > 0, t_a > 0$, $t_{ee} < 0$. We assume that individuals know their own ability, as, again, individuals know their baseline scores and their absolute test scores in control and treatment alike.

The second term captures the utility deriving from social comparisons to which individual i gives weight $\beta_i^{sc} > 0$. Social comparisons enter additively as in Kandel and Lazear (1992), and we assume that individuals care about being of higher ability than their peers rather than having higher test scores *per se*. We assume that individuals enjoy social comparisons when their own ability is higher than the average of their peers a_{-i}^*, and the effect is larger the larger is the ability gap. Conversely, individuals suffer from social comparisons when their own ability is lower than the average of their peers, and the effect is larger the larger is the ability gap.

Finally, as discussed above, we assume that when individuals are given information on test score ranks, they can obtain more precise information about their relative ability. Thus, $T_i^{sc} = 1$ if individual i is in one of the four treatment groups that provide rank information on test scores and 0 in control. In particular, we assume that whether test score ranks give information on ability ranks depends positively on effort through the "signal" function $\sigma(e_i)$ with $\sigma_e > 0$. Intuitively, since test scores are a function of ability and effort, receiving a low rank when exerting low effort could still allow an individual to retain the option of believing that he is of truly high relative ability, but receiving a low rank with high effort could not.[17] The "signal" function $\sigma(e_i)$ is a reduced form representation of "choosing beliefs" in the spirit of Yariv (2002), Köszegi (2006), Oster et al. (2013) and Brunnermeier and Parker (2005).[18]

The third term captures the utility deriving from award A, which in our setting is either employer recognition or social visibility. Thus, $T_i^A = 1$ if individual i is in one of the two treatment groups that provide awards, and 0 otherwise. Note that $T_i^A = 1 \Rightarrow T_i^{SC} = 1$ but not vice versa. The probability of winning the award $p(a_i, e_i, b_i)$ depends on effort, ability, and the baseline score b_i, with $p_e > 0$, $p_{ee} < 0, p_a > 0$, and $p_b < 0$. The latter captures that "most improved" awards are more easily obtainable by those who did poorly in the baseline test.

Finally, $d(e)$ is the disutility of effort, with $d_e > 0$, $d_{ee} > 0$ as is standard.

Maximizing equation (2) with respect to e_i yields:

$$\begin{cases} e_i^C & s.t. \quad \beta_i^t t_e(a_i, e_i^c) - d_e(e_i^c) = 0 & if \quad T_i^{SC} = T_i^A = 0 \\ e_i^{SC} & s.t. \quad \beta_i^t t_e(a_i, e_i^{SC}) + \beta_i^{sc}\sigma_e(e_i^{SC})(a_i - a_{-i}^*) - d_e(e_i^{SC}) = 0 & if \quad T_i^{SC} = 1, T_i^A = 0 \\ e_i^A & s.t. \quad \beta_i^t t_e(a_i, e_i^A) + \beta_i^{sc}\sigma_e(e_i^A)(a_i - a_{-i}^*) + \beta_i^a p_e(a_i, e_{i,}^A b_i)A - d_e(e_i^A) = 0 & if \quad T_i^{SC} = T_i^A = 1 \end{cases}$$

[17] This holds regardless of whether i expects others to exert high or low effort. If he chooses high effort and expects others to do so as well, a low rank in test scores implies a low rank in ability. If he expects others to choose low effort, a low rank in test scores implies a low rank in ability *a fortiori*.

[18] Similarly, Benabou and Tirole (2002) consider a model where agents can manipulate their interim belief through the choice of information structure. They show that less information can be preferable as it can weaken the time inconsistency problem and induce more effort in the future. We do not assume time inconsistency in our setting, which would exacerbate our effect as it would further discount the future impacts of taking distorted actions for interim belief utility.

The first order conditions for effort in the three cases inform the comparison of performance in each treatment vs. the control group and across treatments. These comparisons map into the empirical findings as follows.

First, $e_i^C > e_i^{SC}$ *iff* $a_i < a_{-i}^*$; namely, providing rank information reduces effort for individuals whose ability is lower than their expectation of their peers' ability. Intuitively, all other things equal, effort provides them with something undesirable (a precise private signal of their ranking in the class). This is related to the psychology literature on "self-handicapping," which can take the form of withdrawing effort in performance settings where there is potential for self-image-damaging feedback (Jones and Berglas, 1978; Berglas, 1985).[19] In line with this, a qualitative survey we administered before implementing the treatments reveals that 43% of the trainees in the bottom quartile of baseline scores did not want to know their relative rank in the class, while only 24% of those in the top quartile said the same.

These findings are similar to the behavioral model in Oster et al. (2013), in which people at risk for Huntington's disease prefer not to be undergo a test to learn whether they have the genetic mutation that causes the disease because the anticipatory utility of believing they might not get sick outweighs the costs of potentially distorted actions.

Second $e_i^A > e_i^{SC}$ for all i; namely, providing awards in addition to rank information can only increase effort as long as awards are valuable, since $\beta_i^a p_e(a_i, e_i^A, b_i)A > 0$. The strength of the effect depends on ability and baseline scores, both of which determine the marginal return to effort $p_e(a_i, e_{i,}^A b_i)$.

Third, $e_i^A > e_i^C$ for all i such that $a_i \geq a_{-i}^*$, as both the second (rank information) and third (award) term in the first order condition are weakly positive, thus increasing the marginal return to effort. In contrast, $e_i^A \lesseqgtr e_i^C$ for all i such that $a_i < a_{-i}^*$ since the second (rank information) term is negative and the third (award) term is positive. Thus, providing awards might reduce effort compared to the control group that receives no awards or rank information if the response to rank information is stronger than the response to the award itself.

[19] This allows the trainee to manipulate the attribution of failure to himself (Kelley, 1971). If failure occurs, the extent to which the outcome is attributed to his lack of ability is discounted because of the equally likely reason for failure: decreased effort. However, if the trainee obtains a high grade on the exam, the attribution to ability is strengthened because the decreased learning effort made it even more difficult to do well.

5. Conclusion

Our results suggest that awards can have a negative effect on performance as they facilitate social comparison, even though they have a positive effect through employer recognition and enhanced social visibility. In our context, the negative effect of rank information on learning was large enough to undo the positive effects of awards. Since learning is directly related to future labor productivity, and test scores are significantly predictive of future performance, the distortion in effort is practically significant, and detrimental to the objective of increasing overall learning and later on-the-job productivity.

While the relative magnitudes of these effects are likely to be context-specific, the possible negative effects of rank information and social comparisons are important to consider in the overall productivity effects and design of performance awards. In particular, we show that individuals may even reduce effort in anticipation of learning rank information. This is consistent with a model in which individuals have preferences over their self-perceived ability ranking, and thus engage in information avoidance to be able to retain a positive view of themselves. In effect, in this model, agents self-sabotage in order to avoid the signal contained in ranking information so as not to have to update their beliefs about their own relative ability. As the risk of a low-rank signal is greater in lower ability ranges, this self-handicapping is worse at the bottom of the distribution.

From an employer or policy maker perspective, the cost of an incentive structure that differentially affects the lower tail depends on the nature of the production function. It is particularly costly when there are complementarities in production or when the performance of the lower tail is critical to the principal's goal, as in our setting where the government wants to ensure equitable provision of health services to remote rural areas. In domains such as innovation in science and finance, the effect on productivity of the upper tail of the distribution might be most important. But in domains such as health services delivery where the potential for harm is high, it is critical to employ incentives that are not detrimental to the lower tail of the distribution. This depends in part, as well, on the ease of exit and entry; if the goal is to induce the lower tail to withdraw effort, and potentially exit, then providing rank information could be a highly effective means to do so. It is left to future research to unbundle these effects across cultures and professional sectors.

Angrist, J. and Lavy, V. (2009). The effects of high stakes high school achievement awards: Evidence from a randomized trial. *The American Economic Review*, 99(4):1384–1414.

Ashraf, N., Bandiera, O., and Jack, K. (2013a). No margin, no mission? A field experiment on incentives for public service delivery. Manuscript.

Ashraf, N., Bandiera, O., and Lee, S. (2013b). Mission incentives: Selection and compensation of community health workers in Zambia. Manuscript.

Azmat, G. and Iriberri, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, 94(7):435–452.

Bandiera, O., Barankay, I., and Rasul, I. (2013). Team incentives: Evidence from a firm level experiment. Forthcoming, Journal of the European Economic Association.

Baqui, A., Kumar, V., Mohanty, S., Kumar, A., Misra, R. P., Santosham, M., and Awasthi, S. (2008). Effect of community-based behaviour change management on neonatal mortality in Shivgarh, Uttar Pradesh, India: A cluster-randomised controlled trial. *The Lancet*, 372(9644):1151–1162.

Barankay, I. (2012). Rank incentives: Evidence from a randomized workplace experiment. Working Paper, Wharton School, University of Pennsylvania.

Bénabou, R. and Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics*, 117(3):871–915.

Berglas, S. (1985). Self-handicapping and self-handicappers: A cognitive/attributional model of interpersonal self-protective behavior. *Perspectives in Personality*, 1:235–270.

Besley, T. and Ghatak, M. (2008). Status incentives. *The American Economic Review*, 98(2):206–211.

Bhandari, N., Paul, V. K., Sachdev, H. S., Mavalankar, D., Ramachandran, P., Sankar, M. J., and Sreenivas, V. (2011). Reproductive health, child health and nutrition in India: Meeting the challenge. *The Lancet*, 377:332–349.

Bruhn, M. and Mckenzie, D. (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics*, 1(4):200–232.

Brunnermeier, M. and Parker, J. A. (2005). Optimal expectations. *The American Economic Review*, 95(4):1092–1118.

Charness, G. and Grosskopf, B. (2001). Relative payoffs and happiness: An experimental study. *Journal of Economic Behavior and Organization*, 45:301–328.

Dewatripont, M., Jewitt, I., and Tirole, J. (1999). The economics of career concerns, part i: Comparing information structures. *Review of Economic Studies*, 66:183–198.

Dur, R., Bradler, C., Neckermann, S., and Non, A. (2013). Employee recognition and performance: A field experiment. ZEW Discussion Paper No. 13-017.

Fehr, E. and Schmidt, K. (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, 114(3):817–868.

Freeman, R. and Gelber, A. (2010). Prize structure and information in tournaments: Experimental evidence. *American Economic Journal: Applied Economics*, 2(1):149–64.

Fryer, R. (2011). Financial incentives and student achievement: Evidence from randomized trials. *The Quarterly Journal of Economics*, 126:1755–1798.

Government of Zambia (2011). National human resources for health strategic plan 2011–2015. Technical report, Ministry of Health.

Jones, E. and Berglas, S. (1978). Control of attributions about the self through self-handicapping strategies: The appeal of alcohol and the role of underachievement. *Personality and Social Psychology Bulletin*, 4(2):200–206.

Kandel, E. and Lazear, E. (1992). Peer pressure and partnerships. *Journal of Political Economy*, 100(4):801–817.

Kelley, H. (1971). *Attribution in Social Interaction*. General Learning Press, Morristown, New Jersey.

Köszegi, B. (2002). Emotional agency. *The Quarterly Journal of Economics*, 121(1):121–155.

Köszegi, B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European Economic Association*, 4(4):673–707.

Kremer, M., Miguel, E., and Thornton, R. (2009). Incentives to learn. *The Review of Economics and Statistics*, 91(3):437–456.

Lazear, E. (1989). Pay equality and industrial politics. *Journal of Political Economy*, 97(3):561–580.

Leuven, E., Oosterbeek, H., and van der Klaauw, B. (2010). The effect of financial rewards on students' achievement: Evidence from a randomized experiment. *Journal of the European Economic Association*, 8(6):1243–1265.

Major, B., Testa, M., and Bylsma, W. H. (1991). *Social Comparison: Contemporary Theory and Research*. Erlbaum, Hillsdale, NJ.

Mathauer, I. and Imhoff, I. (2006). Health worker motivation in Africa: The role of non-financial incentives and human resource management tools. *Human Resources for Health*, 4(1):24.

Milkovich, G. and Newman, J. (1996). *Compensation*. Irwin, Homewood, IL, 4 edition.

Moldovanu, B. and Sela, A. (2001). The optimal allocation of prizes in contests. *The American Economic Review*, 91:542–558.

Moldovanu, B., Sela, A., and Shi, X. (2007). Contests for status. *Journal of Political Economy*, 115:338–363.

Müller, W. and Schotter, A. (2010). Workaholics and dropouts in organizations. *Journal of the European Economic Association*, 8(4):717–743.

Muralidharan, K. and Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy*, 119(1):39–77.

Nalbantian, H. and Schotter, A. (1997). Productivity under group incentives: An experimental study. *The American Economic Review*, 87(3):314–341.

Oster, E., Shoulson, I., and Dorsey, E. R. (2013). Optimal expectations and limited medical testing: Evidence from huntington disease. *American Economic Review*, 103(2).

Smith, A. (1759). *The Theory of Moral Sentiments*. Liberty Fund, Indianapolis.

Spencer, M., Rosland, A.-M., Kieffer, E. C., Sinco, B., Valerio, M., Palmisano, G., Anderson, M., Guzman, J. R., and Heisler, M. (2011). Effectiveness of a community health worker intervention among African American and Latino adults with type 2 diabetes: A randomized controlled trial. *American Journal of Public Health*, 101(12):2253–2260.

Stone, R. (2004). Self-handicapping, tournaments, and ego utility. *Unpublished manuscript*.

Tran, A. and Zeckhauser, R. (2012). Rank as an inherent incentive: Evidence from a field experiment. *Journal of Public Economics*, 96:645–650.

Yariv, L. (2002). I'll see it when I believe it–A simple model of cognitive consistency. Cowles Foundation Discussion Paper No. 1352.

Figure 1: Schematic of Experimental Treatment Conditions.

	Components				
	After each course exam, each trainee is given a private letter that indicates his or her **absolute score** on the exam as well as his or her **value added** with respect to his or her course sub-score in the baseline exam.	The letter reports the trainee's **rank** within the classroom on absolute score and value added, respectively.	The two trainees with the highest exam score and the two trainees with the highest value added are **publicly named in the letters** given to all trainees within the classroom.	The four top-performing trainees are given a **personalized letter from the program director at the Ministry of Health** congratulating them on their achievement.	One of the four top-performing trainees is randomly chosen to be **profiled in a newsletter** that is sent to all of the trainees' communities of origin.
C: Control	✔				
T1: Private Social Comparison	✔	✔			
T2: T1 + Public Social Comparison	✔	✔	✔		
T3: T2 + Employer Recognition Award	✔	✔	✔	✔	
T4: T2 + Social Visibility Award	✔	✔	✔		✔

Table 1: Balance Across Treatments

	C: Control	T1: Private Social Comparison	Test of equality: C=T1	T2: T1 + Public Social Comparison	Test of equality: C=T2	T3: T2 + Employer Recognition Award	Test of equality: C=T3	T4: T2 + Social Visibility Award	Test of equality: C=T4	Joint test of equality: C=T1=T2=T3=T4
Baseline test score (0-100)	0.610	0.600	$p = 0.541$	0.598	$p = 0.463$	0.609	$p = 0.936$	0.613	$p = 0.864$	$p = 0.865$
	(0.088)	(0.090)		(0.094)		(0.087)		(0.084)		
English test score (0-100)	39.97	44.56	$p = 0.048$	42.81	$p = 0.219$	40.1	$p = 0.944$	43.26	$p = 0.155$	$p = 0.193$
	(11.89)	(11.07)		(15.70)		(10.70)		(13.62)		
Number of O-Levels	5.97	5.93	$p = 0.898$	5.89	$p = 0.807$	5.91	$p = 0.847$	5.86	$p = 0.720$	$p = 0.993$
	(1.36)	(1.56)		(1.57)		(1.68)		(1.67)		
Age (years)	26.24	26.02	$p = 0.843$	27.01	$p = 0.466$	27.33	$p = 0.313$	28.47	$p = 0.040$	$p = 0.169$
	(5.73)	(5.60)		(5.95)		(6.05)		(6.60)		
Gender (=1 if female)	0.459	0.583	$p = 0.174$	0.484	$p = 0.784$	0.484	$p = 0.784$	0.516	$p = 0.529$	$p = 0.695$
	(0.502)	(0.497)		(0.504)		(0.504)		(0.504)		
Unemployed at time of application (=1 if yes)	0.705	0.639	$p = 0.460$	0.565	$p = 0.112$	0.565	$p = 0.112$	0.565	$p = 0.112$	$p = 0.393$
	(0.460)	(0.484)		(0.500)		(0.500)		(0.500)		
Work experience in the health sector (=1 if yes)	0.361	0.344	$p = 0.855$	0.484	$p = 0.168$	0.452	$p = 0.309$	0.516	$p = 0.082$	$p = 0.221$
	(0.484)	(0.479)		(0.504)		(0.502)		(0.504)		
District used community recruitment (=1 if yes)	0.475	0.525	$p = 0.589$	0.500	$p = 0.786$	0.419	$p = 0.537$	0.484	$p = 0.926$	$p = 0.828$
	(0.503)	(0.504)		(0.504)		(0.497)		(0.504)		
Number of trainees	61	61		62		62		62		

Notes: Table reports means and standard deviations (in parentheses) for the five experimental conditions, as well as tests of equality for each control-treatment pair. Baseline test score is the score (0-100) on a baseline test administered at the beginning of the training program to measure trainees pre-existing knowledge of the materials taught during the training program. English test score is the score (0-100) on a test administered at the beginning of the training program to measure trainees' knowledge of English. Unemployed at time of application is a self-reported variable that equals 1 if the trainee had no full- or part-time employment when they applied for the CHA position. Work experience is a self-reported variable that equals 1 if the trainee has held a paid or unpaid position in the health sector in the past. District used community recruitment equals 1 if trainees come from a district where the main mission of the CHA job was advertised as "serving the community;" it equals 0 if trainees come from a district where the main mission of the CHA job was advertised as "fostering your career."

Table 2: Average Treatment Effects

	(1)	(2)	(3)	(4)	(5)
			Heterogeneous effects by exam period		
	Baseline	Teacher controls	First exam	Following exams	Test of equality
T1: Private Social Comparison	-0.308*	-0.351**	-0.287*	-0.322*	$p =$ 0.844
	(0.161)	(0.171)	(0.168)	(0.194)	
T2: T1 + Public Social Comparison	-0.379***	-0.409***	-0.456***	-0.324**	$p =$ 0.378
	(0.145)	(0.155)	(0.172)	(0.156)	
T3: T2 + Employer Recognition Award	0.005	-0.044	0.128	-0.080	$p =$ 0.123
	(0.133)	(0.174)	(0.165)	(0.136)	
T4: T2 + Social Visibility Award	0.064	0.122	-0.140	0.112	$p =$ 0.087
	(0.135)	(0.148)	(0.165)	(0.143)	
Net effect of Public Social Comparison (T2-T1)	-0.071	-0.058	-0.170	-0.003	$p =$ 0.381
	(0.166)	(0.164)	(0.174)	(0.201)	
Net effect of Employer Recognition Award (T3-T2)	0.384***	0.365**	0.585***	0.244*	$p =$ 0.026
	(0.141)	(0.161)	(0.175)	(0.148)	
Net effect of Social Visibility Award (T4-T2)	0 .443***	0.531***	0.316*	0.436***	$p =$ 0.458
	(0.141)	(0.151)	(0.170)	(0.153)	
Trainee controls	yes	yes		yes	
Teacher controls	no	yes		no	
Number of clusters (trainees)	307	307		307	
Number of observations (trainee-courses)	1149	850		1149	
Adjusted R-squared	0.213	0.212		0.224	

Notes: Table reports OLS estimates, weighted by course duration, with standard errors clustered at the trainee level in parentheses. Dependent variable is normalized exam score, normalized by the mean and standard deviation of the control group for each exam. Trainee controls include: score in the baseline test for each of the four courses, English test score, gender, age, previous experience in the health sector, employment status at the time of application, district recruitment strategy. Teacher controls include: average teacher rating and whether at least one of the teachers had specific expertise in the subject matter. Both variables are defined at the course-treatment level. Columns (2) and (3) report coefficients estimated in the same regression where we include all treatments interacted with an indicator variable that takes value 0 in the first period and 1 thereafter.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table 3: Heterogeneous Treatment Effects by Baseline Test Score

	(1) First tercile of baseline test score	(2) Second tercile of baseline test score	(3) Third tercile of baseline test score
T1: Private Social Comparison	-0.672**	-0.233	0.010
	(0.298)	(0.271)	(0.213)
T2: T1 + Public Social Comparison	-0.812***	-0.179	-0.113
	(0.260)	(0.199)	(0.203)
T3: T2+ Employer Recognition Award	-0.280	0.157	-0.005
	(0.216)	(0.250)	(0.159)
T4: T2+Social Visibility Award	-0.102	0.204	-0.031
	(0.309)	(0.223)	(0.177)
Net effect of Public Social Comparison (T2-T1)	-0.140	0.054	-0.153
	(0.334)	(0.243)	(0.214)
Net effect of Employer Recognition Award (T3-T2)	0.532**	0.336	0.107
	(0.255)	(0.214)	(0.166)
Net effect of Social Visibility Award (T4-T2)	0 .710**	0.383**	0.082
	(0.342)	(0.183)	(0.173)
Trainee controls	yes	yes	yes
Number of clusters (trainees)	92	107	107
Number of observations (trainee-courses)	350	401	398
Adjusted R-squared	0.186	0.079	0.157

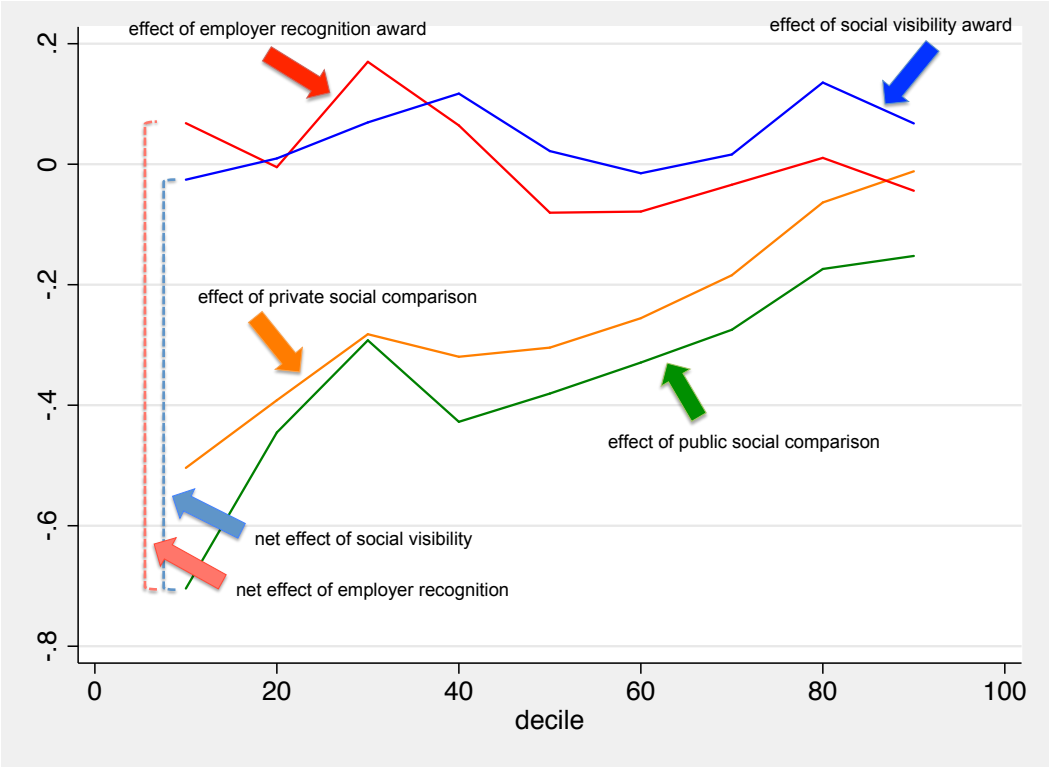
Notes: Table reports OLS estimates, weighted by course duration, with standard errors clustered at the trainee level in parentheses. Dependent variable is normalized exam score, normalized by the mean and standard deviation of the control group for each exam. Individual controls include: score in the baseline test for each of the four courses, English test score, gender, age, previous experience in the health sector, employment status at the time of application, district recruitment strategy.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Figure 2: Quantile Treatment Effects



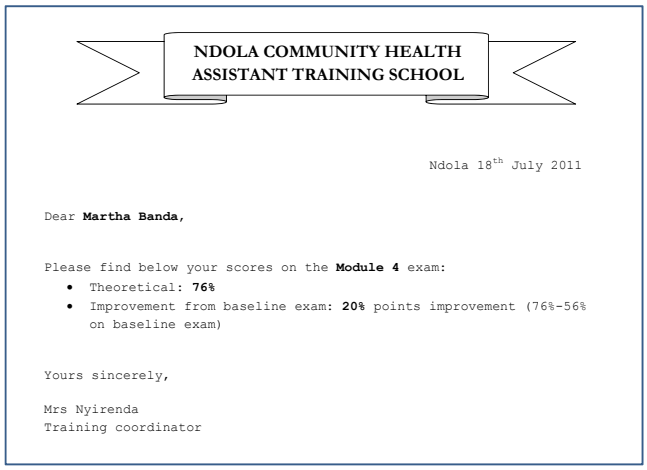
Notes: Each line connects treatment effects of each treatment estimated at each decile. Point estimates and standard errors are reported in Table A1. Individual controls include: score in the baseline test for each of the four courses, English test score, gender, age, previous experience in the health sector, employment status at the time of application, district recruitment strategy.

Appendix

Figure A.1: Outline of One-year Community Health Assistant Training Curriculum

Course	Module	Teaching Days	Topics
0	1: Health care system in Zambia	4	• Organization and functions of the health care system in Zambia • Roles and responsibilities of a Community Health Assistant • Code of conduct • Gender equality • Community mobilization and networking
	2: Behavioral sciences	19	• Introduction to psychology • Mental health and common psychiatric conditions • Introduction to sociology • Family and community
	3: Health promotion	9	• Introduction to health promotion • Communication skills
	4: Environmental health	31	• General principles of infection prevention • Water supply • Excreta disposal • Solid waste management • Food hygiene and safety • Housing and health • Insect and rodent control
1	5: Epidemiology	34	• Infectious disease epidemiology • Epidemic investigation and management • Epidemiological surveillance • Data collection
	6: Reproductive and child health	35	• Introduction to reproductive health • Introduction to child health • School health services
2	7: Anatomy and physiology	20	• Introduction to the human body • Musculoskeletal system • Cardiovascular system • Respiratory system • Digestive system • Urinary system • Special senses
3	8: Basic procedures	20	• Occupational safety and health • Lifting and moving patients • History and physical exam skills • Assessment of hygiene, nutrition, physical activity, pain, and vital signs • Wound care • Palliative care
	9: Common medical conditions	38	• Common conditions (malaria, diarrhea, respiratory infection, HIV, tuberculosis, anemia, etc.) • Oral health
4	10: Diagnostic procedures	4	• Malaria rapid diagnostic testing • HIV testing • Sputum collection for TN testing
	11: First aid	9	• Principles of first aid • Bandaging • Lifting and moving patients • Cardio-pulmonary resuscitation • Handling of selected emergencies (toxic ingestion, bites and stings, fractures, burns, drowning, foreign body ingestion)

Figure A.2: Control Group



NDOLA COMMUNITY HEALTH ASSISTANT TRAINING SCHOOL

Ndola 18th July 2011

Dear **Martha Banda,**

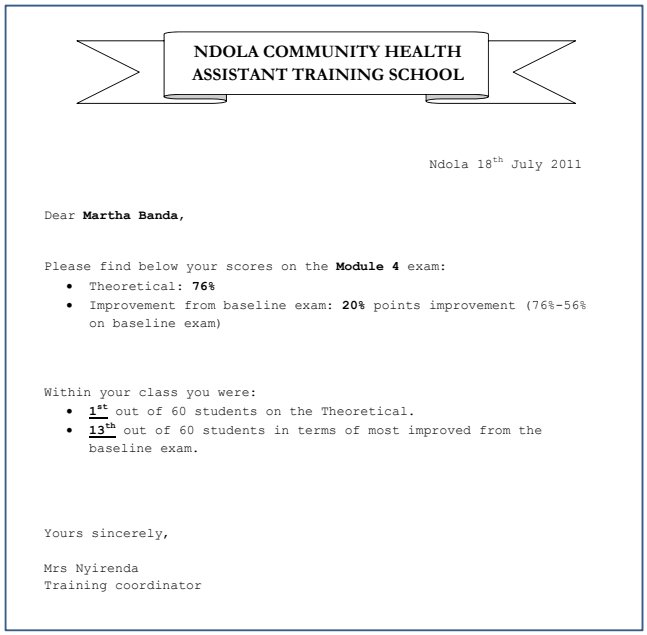
Please find below your scores on the **Module 4** exam:

- Theoretical: **76%**
- Improvement from baseline exam: **20%** points improvement (76%-56% on baseline exam)

Yours sincerely,

Mrs Nyirenda
Training coordinator

Figure A.3: Treatment #1



NDOLA COMMUNITY HEALTH ASSISTANT TRAINING SCHOOL

Ndola 18th July 2011

Dear **Martha Banda,**

Please find below your scores on the **Module 4** exam:

- Theoretical: **76%**
- Improvement from baseline exam: **20%** points improvement (76%-56% on baseline exam)

Within your class you were:

- **1st** out of 60 students on the Theoretical.
- **13th** out of 60 students in terms of most improved from the baseline exam.

Yours sincerely,

Mrs Nyirenda
Training coordinator

All names are fictional.

Figure A.4: Treatment #2

NDOLA COMMUNITY HEALTH ASSISTANT TRAINING SCHOOL

Ndola 18th July 2011

Dear **Martha Banda,**

Please find below your scores on the **Module 4** exam:

- Theoretical: **76%**
- Improvement from baseline exam: **20%** points improvement (76%-56% on baseline exam)

Within your class you were:

- **1st** out of 60 students on the Theoretical.
- **13th** out of 60 students in terms of most improved from the baseline exam.

Please note the following top performers:

- **Martha Banda** was 1st in this class on Theoretical score.
- **James Mwanza** was 2nd in this class on Theoretical score.
- **Peter Mwaba** was 1st in this class on most improved from the baseline exam.
- **Martha Chilima** was 2nd in this class on most improved from the baseline exam.

Yours sincerely,

Mrs Nyirenda
Training coordinator

Figure A.5: Treatment #3

NDOLA COMMUNITY HEALTH ASSISTANT TRAINING SCHOOL

Ndola 18th July 2011

Dear **Martha Banda,**

Please find below your scores on the **Module 4** exam:

- Theoretical: **76%**
- Improvement from baseline exam: **20%** points improvement (76%-56% on baseline exam)

Within your class you were:

- **1st** out of 60 students on the Theoretical.
- **13th** out of 60 students in terms of most improved from the baseline exam.

Please note the following top performers:

- **Martha Banda** was 1st in this class on Theoretical score.
- **James Mwanza** was 2nd in this class on Theoretical score.
- **Peter Mwaba** was 1st in this class on most improved from the baseline exam.
- **Martha Chilima** was 2nd in this class on baseline exam.

Yours sincerely,

Mrs Nyirenda
Training coordinator

+

Top performers also receive letter from MoH →

MINISTRY OF HEALTH
REPUBLIC OF ZAMBIA

Lusaka 18 July 2011

Dear Martha,

Congratulations for being the most improved student in theory for module 4 in your class. I am very happy with your achievement and hope you will continue working hard. I have noted your performance and look forward to congratulating you in person.

Warm regards,

Mrs Mutinta Musonda
Assistant Director of Human Resources
Ministry of Health

Figure A.6: Treatment #4

NDOLA COMMUNITY HEALTH ASSISTANT TRAINING SCHOOL

Ndola 18th July 2011

Dear **Martha Banda**,

Please find below your scores on the **Module 4** exam:

- Theoretical: **76%**
- Improvement from baseline exam: **20%** points improvement (76%-56% on baseline exam)

Within your class you were:

- **1st** out of 60 students on the Theoretical.
- **13th** out of 60 students in terms of most improved from the baseline exam.

Please note the following top performers:

- **Martha Banda** was 1st in this class on Theoretical score.
- **James Mwanza** was 2nd in this class on Theoretical score.
- **Peter Mwaba** was 1st in this class on most improved from the baseline exam.
- **Martha Chilima** was 2nd in this class on most improved from the baseline exam.

Yours sincerely,

Mrs Nyirenda
Training coordinator

+

One of top performers also featured in a newsletter sent to their community →

CHA NEWSLETTER | Issue 1 2

"Why I want to be a CHA"

By Martha Banda

LET'S GET TO KNOW MARTHA!

Martha Banda is a 27-year-old CHA student representing Lubanda Health Post, Itezhi-tezhi District, Southern Province. Her counterpart at this health post is Andrew Phiri.

She is married and a mother of two children. She completed secondary school at Kabulonga Girls High School in 2001. Before joining the CHA program, Martha was volunteering as an HIV/AIDS educator in her community.

Martha Banda, age 27 years

In this section, we wish to provide an opportunity for CHA students to be known to the community. ***Martha Banda*** *from Lubanda Health Post shares her thoughts on why she wants to be a Community Health Assistant.*

Zambia has faced many challenges in its health sector but I have become inspired by the fact that the government wants to train community health workers. I joined the program because of my passion for patients, my love for others, and because I want to reduce the gap between my people and the hospital. I plan to encourage my community to use healthy resources like boreholes for clean drinking water and community resources such as the health post to learn more about nutritious diets.

I like telling people about health issues and I want to be a Community Health Assistant so that I can help my friends in my community, and so we can work together. I like caring for people who are sick – cooking for them, washing them, and giving them words of encouragement.

My first step to being a Community Health Assistant, after finishing my training, will be to gain the respect of my community, first and foremost. I will start by talking with my village headman and, after getting his blessing, I will begin to call the community for meetings where I plan to give health talks on how to care for children under-five and educate the community on treating their water with chlorine prior to drinking it. I am looking forward to serving my community!

Table A.1: Quantile Treatment Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Quantile	10th	20th	30th	40th	50th	60th	70th	80th	90th
T1: Private Social Comparison	-0.504***	-0.392***	-0.282*	-0.320***	-0.304**	-0.256**	-0.184**	-0.0635	-0.0118
	(0.176)	(0.142)	(0.158)	(0.115)	(0.135)	(0.105)	(0.0875)	(0.117)	(0.124)
T2: T1 + Public Social Comparison	-0.704***	-0.445***	-0.292*	-0.428***	-0.381***	-0.329***	-0.275***	-0.174	-0.152
	(0.171)	(0.140)	(0.156)	(0.114)	(0.133)	(0.103)	(0.0857)	(0.114)	(0.115)
T3: T2+ Employer Recognition Award	0.0681	-0.00494	0.170	0.0644	-0.0805	-0.0786	-0.0341	0.0105	-0.0440
	(0.157)	(0.137)	(0.152)	(0.113)	(0.133)	(0.103)	(0.0862)	(0.117)	(0.117)
T4: T2+Social Visibility Award	-0.0257	0.00956	0.0694	0.117	0.0218	-0.0151	0.0161	0.136	0.0677
	(0.168)	(0.153)	(0.172)	(0.126)	(0.147)	(0.115)	(0.0965)	(0.130)	(0.135)
Net effect of Public Social Comparison (T2-T1)	-0.200	-0.053	-0.010	-0.108	-0.076	-0.074	-0.090	-0.110	-0.140
	(0.180)	(0.141)	(0.156)	(0.114)	(0.133)	(0.104)	(0.087)	(0.117)	(0.128)
Net effect of Employer Recognition Award (T3-T2)	0.772***	0.441***	0.462***	0.492***	0.300**	0.251**	0.241***	0.184	0.108
	(0.164)	(0.137)	(0.152)	(0.111)	(0.131)	(0.102)	(0.086)	(0.116)	(0.118)
Net effect of Social Visibility Award (T4-T2)	0.679***	0.455***	0.362**	0.545***	0.402***	0.314***	0.291***	0.309	0.220
	(0.172)	(0.149)	(0.169)	(0.124)	(0.145)	(0.114)	(0.097)	(0.131)	(0.138)
Number of observations (trainee-courses)	1149	1149	1149	1149	1149	1149	1149	1149	1149
Adjusted R-squared	0.1528	0.1392	0.1317	0.1335	0.1336	0.1341	0.1363	0.1283	0.1183

Notes: Table reports OLS estimates, weighted by course duration, with standard errors clustered at the trainee level in parentheses. Dependent variable is normalized exam score, normalized by the mean and standard deviation of the control group for each exam. Individual controls include: score in the baseline test for each of the four courses, English test score, gender, age, previous experience in the health sector, employment status at the time of application, district recruitment strategy.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Table A.2: Heterogeneous Treatment Effects by Baseline Test Score and Period

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Bottom tercile of baseline test score				Middle tercile of baseline test score				Top tercile of baseline test score			
		Heterogeneous effects by period				Heterogeneous effects by exam period				Heterogeneous effects by exam period		
	All periods	First period	Following periods	Test of equality: (2) = (3)	All periods	First period	Following periods	Test of equality: (6) = (7)	All periods	First period	Following periods	Test of equality: (10)=(11)
T1: Private Social Comparison	-0.672**	-0.578**	-0.737*	$p = 0.656$	-0.233	-0.119	-0.316	$p = 0.519$	0.010	-0.018	0.081	$p = 0.618$
	(0.298)	(0.263)	(0.394)		(0.271)	(0.286)	(0.328)		(0.213)	(0.233)	(0.234)	
T2: T1 + Public Social Comparison	-0.812***	-1.065***	-0.640**	$p = 0.124$	-0.179	-0.258	-0.126	$p = 0.562$	-0.113	0.059	-0.225	$p = 0.186$
	(0.260)	(0.261)	(0.314)		(0.199)	(0.256)	(0.209)		(0.203)	(0.226)	(0.230)	
T3: T2+ Employer Recognition Award	-0.280	-0.146	-0.372	$p = 0.335$	0.157	0.075	0.213	$p = 0.600$	-0.005	0.210	-0.154	$p = 0.060$
	(0.216)	(0.249)	(0.244)		(0.250)	(0.329)	(0.246)		(0.159)	(0.207)	(0.170)	
T4: T2+Social Visibility Award	-0.102	-0.102	-0.150	$p = 0.891$	0.204	-0.108	0.286	$p = 0.064$	-0.031	-0.203	0.028	$p = 0.294$
	(0.309)	(0.391)	(0.336)		(0.223)	(0.275)	(0.226)		(0.177)	(0.234)	(0.191)	
Net effect of Public Social Comparison (T2-T1)	-0.140	-0.487	0.097	$p = 0.115$	0.054	-0.138	0.190	$p = 0.303$	-0.153	0.076	-0.306	$p = 0.079$
	(0.334)	(0.298)	(0.428)		(0.243)	(0.243)	(0.320)		(0.214)	(0.229)	(0.247)	
Net effect of Employer Recognition Award (T3-T2)	0.532**	0.919***	0.268	$p = 0.007$	0.336	0.333	0.339	$p = 0.982$	0.107	0.152	0.071	$p = 0.705$
	(0.255)	(0.278)	(0.284)		(0.214)	(0.300)	(0.220)		(0.166)	(0.200)	(0.195)	
Net effect of Social Visibility Award (T4-T2)	0 .710**	0.963**	0.490	$p = 0.182$	0.383**	0.150	0.412**	$p = 0.223$	0.082	-0.261	0.253	$p = 0.032$
	(0.342)	(0.408)	(0.375)		(0.183)	(0.222)	(0.202)		(0.173)	(0.226)	(0.200)	
Trainee controls	yes		yes		yes		yes		yes		yes	
Number of clusters (trainees)	92		92		107		107		107		107	
Number of observations (trainee-courses)	350		350		401		401		398		398	
Adjusted R-squared	0.186		0.228		0.079		0.111		0.157		0.207	

Notes: Table reports OLS estimates, weighted by course duration, with standard errors clustered at the trainee level in parentheses. Dependent variable is normalized exam score, normalized by the mean and standard deviation of the control group for each exam. Individual controls include: score in the baseline test for each of the four courses, English test score, gender, age, previous experience in the health sector, employment status at the time of application, district recruitment strategy. Columns (2) and (3), (6) and (7), and (10) and (11) report coefficients estimated in the same regression where we include all treatments interacted with an indicator variable that takes value 0 in the first period and 1 thereafter.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.