When Student Incentives Don't Work
Evidence from a Field Experiment in Malawi

Jim Berry[1]   Hyuncheol Bryant Kim[2]   Hyuk Son[2]

[1]Delaware University
[2]Cornell University

December, 2019

Motivation
Impacts of Financial incentives in Education

1. Theoretically ambiguous.
   - Standard economic theory predicts that financial incentives may induce student effort.
   - Incentive may crowd out intrinsic motivation.
2. Empirical evidence is largely mixed.
   - Mixed impacts on both efforts and intrinsic motivation.
3. Understanding why incentive programs do and don't work remains an important open research area.

Motivation
Tournament schemes as an incentives

1. Individual tournament scheme received substantial research attention
   - It allows for the policy maker to set a fixed budget for the incentives.
   - It is incentive compatible to induce effort.
2. But,
   - it may induce effort only from top students.
   - bottom students may not be motivated to exert effort.

This Paper

1. Studies the impacts of two types of incentive programs(framed as scholarship schemes)

   - Incentive: MWK 4500 (USD 9.70)

   - *Standard* merit-based scholarship

     - Top 15% receive the incentive.

   - *Relative* merit-based scholarship

     - Grouped into bins by baseline test score.
     - Top 15% within a bin receive the incentive.
     - Based on "pay for percentile" scheme for teachers (Barlevy and Neal, 2012)
     - Similar to performance standards based on improvements.

Scholarship
Beneficiary

Merit-based

Scholarship
Beneficiary

Relative Merit-based
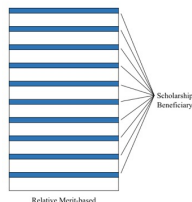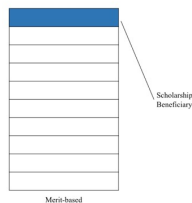
2. Examines complementarities of test result feedback and incentives

Contribution to the Literature
Financial incentives in Education

1. Evidence on these programs is generally mixed
   - Developing countries: Kremer, Miguel, and Thornton (2009); Sharma (2010); Behrman et al. (2015); Hirshleifer (2017)
   - Developed countries: Gneezy, Meier, and Rey-Biel (2011) for a review
2. Kremer, Miguel and Thornton (2009)
   - merit-based girls' scholarship program in rural Kenya (top 15%)
   - Average test scores improved by 0.19 standard deviation.
   - Improvement of test score among low-scoring girls and boys
   - Similar to our *Standard* incentive scheme
3. A few papers compare different incentive schemes:
   - Group vs. Individual: Li et al. (2014); Blimpo (2012)
   - Effort vs. Achievement : Hirschleifer (2016)
   - By targets: Teachers vs. Students Behrman, et al (2015), Parents vs. Children(Berry (2015))

Contribution to the Literature
Educational Incentives' influence on motivation and other non-cognitive skills

1. In psychology, no consensus on if incentives decrease motivation: Cameron and Pierce (1994); Deci, Koestner, and Ryan (1999)
2. In economics, evidence is also mixed.
   - Bettinger (2011): incentives for exam performance did not decrease survey-based intrinsic motivation.
   - Visaria et al. (2016): incentives for attendance among primary students in India decreased intrinsic motivation.

Contribution to the Literature
Feedback

1. Feedback may provide information:
   - Bandiera, Larcinese and Rasul (2015): providing feedback on prior test scores increases subsequent exam performance
2. Feedback may be motivating through a sense of competition:
   - Tran and Zeckhauser (2012), Azmat and Iriberri (2010): providing feedback on relative rank improves academic performance

1. TA Chimutu, Malawi
   - Rural area in Lilongwe District.
   - About 20km from capital city.
   - 2/3 employed in agriculture.
2. Education in TA Chimutu
   - Student-teacher ratio – 85:1
   - At the end of each semester, students take exams in six subjects.
   - Exam fee: about USD 0.5 to 1.
   - Passing the exams at the third semester is required to proceed.

Experimental Design
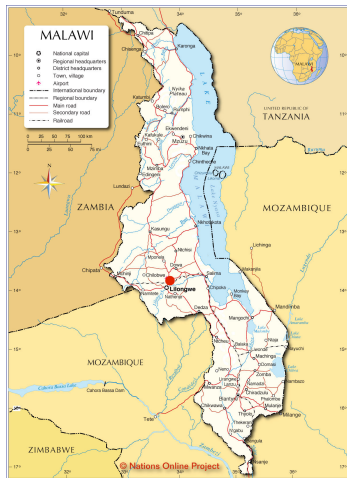
1. 7,386 5th-8th graders in 31 primary schools
2. Financial incentives: school-grade level randomization, 118 school-grades
   - *Standard* merit-based scholarship
     - Incentive condition: achieve top 15% within district at the final exam
   - *Relative* merit-based scholarship
     - Grouped into bins of 100 students by baseline test score within grade.
     - Incentive condition: achieve top 15% within bin at the final exam.
3. Incentives
   - 4,500 MWK (= USD 9.7)
   - 4 options - cash, shoes, school bags, and uniform
   - Malawi GDP per capita (2014): around USD 362.7
4. Rank information feedback : individual level randomization
   - No feedback: baseline rank (+reminder of scholarship scheme)
   - Feedback: baseline & midterm rank (+reminder of scholarship scheme)
   - Standard 8 was excluded due to academic schedule

Experimental Design
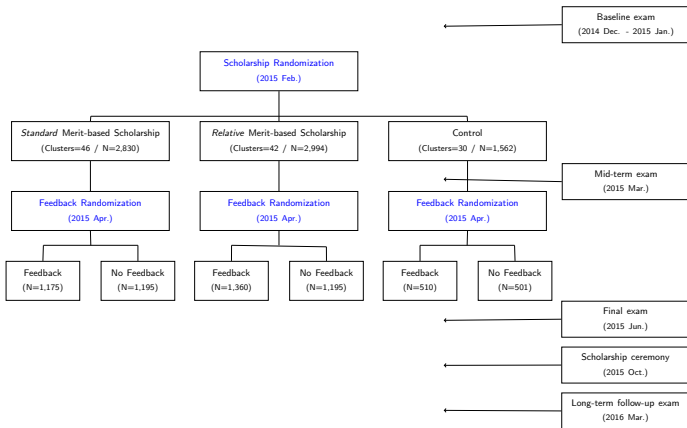
Data

1. District-level test score data
   - Baseline
   - Midterm: used for the feedback intervention
   - Final: used to measure school achievement and select scholarship recipients
     - 5th to 7th grade: Integrated exam on district level, Jul 2015
     - 8th grade: Primary School Leaving Certificate Exam, May 2015
   - Long-term follow-up exam: administered 9 months after the experiment
     - for 5th and 6th graders only.
2. Attendance data from random visits to schools (Nov 2014 - Jun 2015)
3. Baseline and follow-up surveys
   - Demographics, study effort, perceptions of teacher and parental support
   - Non-cognitive skills: level of motivation, self-esteem, grit scale, conscientiousness

Balance

	Scholarship Randomization				Feedback Rand.	
	Control Mean	*Standard* vs.Control	*Relative* vs.Control	N	Feedback vs. No Feedback	N
	(1)	(2)	(3)	(4)	(5)	(6)
Age	14.4 [3.60]	-0.366 (0.311)	-0.300 (0.280)	7385	0.199** (0.093)	6103
Male	0.486 [0.500]	-0.004 (0.019)	-0.028 (0.018)	7385	0.013 (0.013)	6103
Ethnic group: Chewa	0.914 [0.280]	-0.033 (0.035)	-0.036 (0.035)	7358	-0.003 (0.006)	6077
Household size	7.81 [1.66]	0.228 (0.361)	0.157 (0.328)	7385	0.038 (0.032)	6103
Asset index	-0.009 [1.88]	0.0006 (0.183)	0.012 (0.175)	7102	-0.090* (0.051)	5848
Baseline rank(%)	51.5 [27.3]	-0.284 (3.05)	1.89 (3.90)	7342	-0.246 (0.591)	6061
Baseline Score	51.5 [8.59]	-1.78 (1.28)	-1.93 (1.56)	7342	-0.075 (0.154)	6061

Balance(cont.)

	Scholarship Randomization				Feedback Rand.	
	Control Mean	*Standard* vs.Control	*Relative* vs.Control	N	Feedback vs. No Feedback	N
	(1)	(2)	(3)	(4)	(5)	(6)
Attendance	0.863 [0.196]	-0.011 (0.018)	-0.021 (0.018)	7385	0.005 (0.005)	6103
Study hours per week	16.8 [16.4]	-1.00 (0.865)	-0.818 (0.871)	7308	0.163 (0.374)	6031
Motivation to study	4.53 [0.789]	-0.054 (0.065)	0.016 (0.055)	7374	-0.0003 (0.021)	6092
Self-esteem	2.67 [0.338]	-0.027 (0.023)	-0.019 (0.024)	7368	0.011 (0.007)	6087
Conscientious	3.58 [0.600]	-0.028 (0.068)	0.045 (0.066)	7370	0.002 (0.015)	6089
Grit	3.21 [0.450]	-0.050* (0.026)	-0.029 (0.028)	7368	0.021* (0.012)	6087
Teacher effort index	-0.003 [1.000]	0.112 (0.144)	0.202 (0.127)	7364	0.002 (0.023)	6083
Parental Effort Index	0.001 [1.00]	-0.070 (0.076)	-0.050 (0.065)	7281	0.052** (0.025)	6024

Attrition

	Sample: Grade 5-8			Sample: Grade 5-6				
	Mid-term	1st Follow-up		Mid-term	1st Follow-up		2nd Follow-up	
	Exam	Survey	Exam	Exam	Survey	Exam	Survey	Exam
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A								
Standard	0.020	-0.019	0.022	0.017	-0.012	0.023	0.043	0.036
	(0.021)	(0.017)	(0.015)	(0.032)	(0.019)	(0.014)	(0.032)	(0.039)
Relative	0.008	-0.025	0.029**	-0.003	-0.014	0.027*	0.025	0.043
	(0.023)	(0.017)	(0.014)	(0.035)	(0.021)	(0.015)	(0.034)	(0.033)
N	7385	7385	7385	4562	4562	4562	4393	4393
R-Squared	0.006	0.003	0.004	0.002	0.000	0.001	0.001	0.001
Mean of Dep. Var.	0.877	0.827	0.896	0.859	0.836	0.891	0.629	0.568
Panel B								
Feedback		0.003	-0.003		-0.005	-0.006	0.015	0.005
		(0.008)	(0.007)		(0.010)	(0.007)	(0.013)	(0.014)
N		6103	6103		4562	4562	4393	4393
R-Squared		0.000	0.000		0.000	0.000	0.001	0.000
Mean of Dep. Var.		0.836	0.889		0.836	0.891	0.629	0.568

◂ Lee-bound

Empirical strategy

$$Y_{igsz1} = \beta_0 + \beta_1 Standard_{gsz} + \beta_2 Relative_{gsz} + Y_{igsz0} + X_{igsz} + \eta_g + \gamma_z + \varepsilon_{igsz} \tag{1}$$

$$
\begin{aligned}
Y_{igsz1} = {} & \beta_0 + \beta_1 Standard_{gsz} + \beta_2 Relative_{gsz} + \beta_3 Top15_{igsz0} \\
& + \beta_4 Standard_{gsz} * Top15_{igsz0} + \beta_5 Relative_{gsz} * Top15_{igsz0} \\
& + Y_{igsz0} + \eta_g + \gamma_z + X_{igsz} + \varepsilon_{igsz}
\end{aligned} \tag{2}
$$

- Y_{ijgk1}: The outcome of interest for student i of grade g in school s at school zone z
- $Standard$: $Standard$ merit-based scholarship indicator
- $Relative$: $Relative$ merit scholarship indicator
- $Top15$: Indicates baseline test score is within top 15%
- X: Baseline controls.
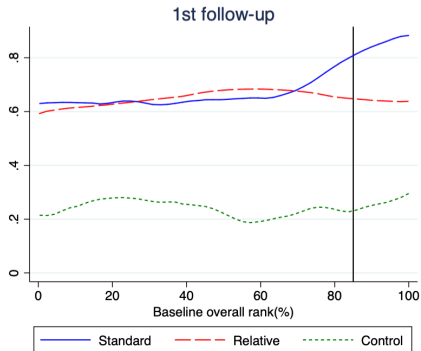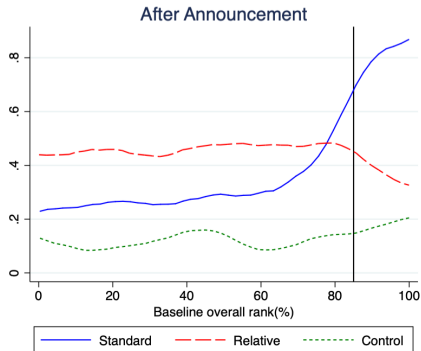- Standard errors are clustered at the school-grade level.
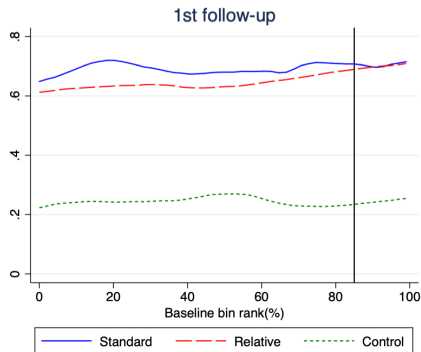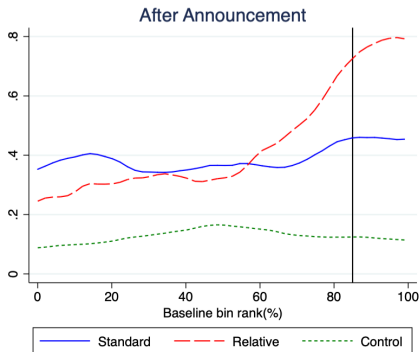
Understanding of Scholarship
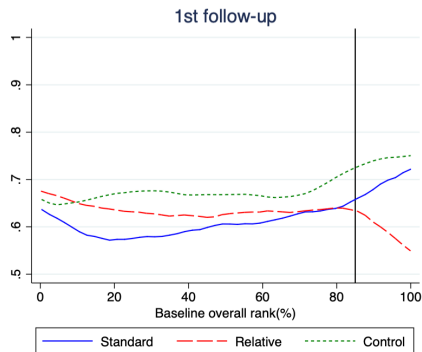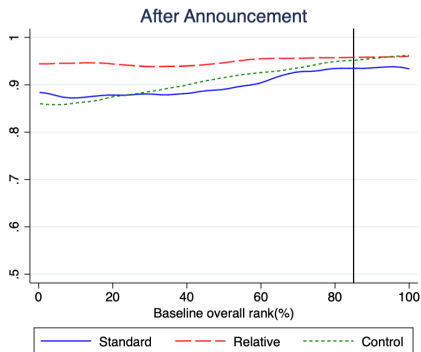
Expectation of Scholarship
Baseline Overall Rank

Expectation of Scholarship

Baseline Bin Rank

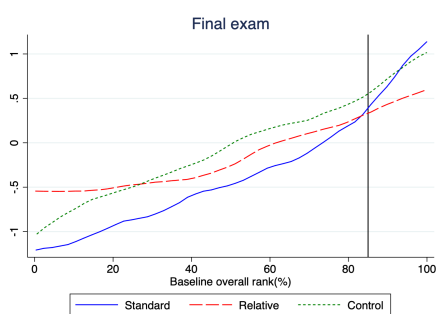Baseline overall rank

Baseline bin rank
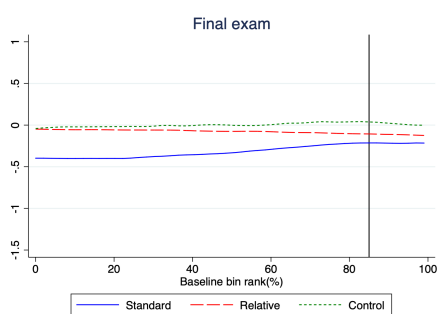
Test Results
Parametric Estimation results

	Dependent Variable: Normalized exam score					
	Grade 5-8		Grade 5-6			
	Final		Final		Long-term f/u	
	(1)	(2)	(3)	(4)	(5)	(6)
Standard	-0.266*	-0.308**	-0.517**	-0.547**	-0.226	-0.217
	(0.145)	(0.152)	(0.247)	(0.266)	(0.143)	(0.148)
Relative	-0.129	-0.080	-0.375	-0.324	-0.319**	-0.326**
	(0.174)	(0.183)	(0.277)	(0.297)	(0.158)	(0.158)
Std. × Top 15%		0.238		0.182		-0.072
		(0.236)		(0.294)		(0.265)
Rel. × Top 15%		-0.275		-0.275		-0.003
		(0.260)		(0.325)		(0.278)
Top 15%		0.109		0.129		0.160
		(0.216)		(0.262)		(0.197)
N	6323	6323	3860	3860	2505	2505
R-Squared	0.323	0.329	0.317	0.322	0.201	0.202
Mean of Dep. Var.	-0.146	-0.146	-0.264	-0.264	0.045	0.045

Test Results

Parametric Estimation using each decile

$$
\begin{aligned}
Y_{ijgk} \;=\;& \beta_0 + \beta_1 Standard_{ij} + \beta_2 Relative_{ij} + \sum_{l \in D} \gamma 1_l \, Topl \\
& + \sum_{l \in D} \gamma 2_l Standard_{ij} \, Topl + \sum_{l \in D} \gamma 3_l Relative_{ij} \, Topl \qquad (3) \\
& + \eta_g + \gamma_k + X_{ij} + \varepsilon_{ijgk}, \text{ where } D = \{1,2,3,4,5,6,7,8,10\}
\end{aligned}
$$
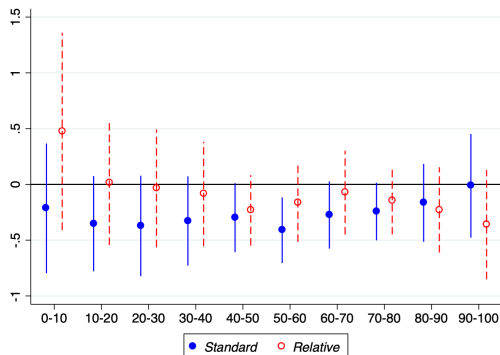
Students' Effort

	Attendance		Study Hours	
	(1)	(2)	(3)	(4)
Standard	0.024*	0.024*	-0.970	-0.961
	(0.013)	(0.013)	(1.036)	(1.121)
Relative	0.009	0.010	-1.562	-1.432
	(0.015)	(0.016)	(1.158)	(1.237)
Std. × Top 15%		-0.008		0.093
		(0.023)		(1.721)
Rel. × Top 15%		-0.021		-0.977
		(0.027)		(2.049)
Top 15%		0.043***		1.511
		(0.016)		(1.526)
N	7085	7085	5242	5242
R-Squared	0.193	0.194	0.076	0.076
Mean of Dep. Var.	0.756	0.756	14.526	14.526

Motivation and Non-Cognitive Skills

	Motivation to study hard		Self esteem		Conscientiousness		Grit	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Standard	-0.071**	-0.090**	-0.030*	-0.035*	-0.045	-0.059*	-0.034	-0.039*
	(0.035)	(0.038)	(0.017)	(0.018)	(0.032)	(0.031)	(0.023)	(0.023)
Relative	-0.036	-0.049	-0.028	-0.026	-0.027	-0.021	-0.027	-0.011
	(0.039)	(0.042)	(0.017)	(0.018)	(0.034)	(0.031)	(0.023)	(0.024)
Std. × Top 15%		0.116*		0.032		0.083		0.029
		(0.063)		(0.039)		(0.094)		(0.051)
Rel. × Top 15%		0.067		-0.016		-0.034		-0.098**
		(0.066)		(0.034)		(0.092)		(0.040)
Top 15%		-0.004		0.024		0.026		0.090***
		(0.050)		(0.030)		(0.083)		(0.029)
N	5754	5754	5842	5842	5844	5844	5842	5842
R-Squared	0.022	0.023	0.050	0.052	0.080	0.083	0.049	0.054
Mean of Dep. Var.	4.298	4.298	2.719	2.719	3.674	3.674	3.259	3.259

Teachers and parental response

	Teacher effort		Parental effort		Parents mention prgm	
	(1)	(2)	(3)	(4)	(5)	(6)
Standard	-0.045	-0.048	-0.037	-0.043	0.126**	0.081
	(0.102)	(0.106)	(0.085)	(0.090)	(0.064)	(0.067)
Relative	-0.059	-0.057	0.022	0.047	0.087	0.107
	(0.088)	(0.093)	(0.083)	(0.087)	(0.071)	(0.069)
Std. × Top 15%		0.019		0.030		0.278**
		(0.129)		(0.111)		(0.108)
Rel. × Top 15%		-0.023		-0.157		-0.054
		(0.125)		(0.111)		(0.120)
Top 15%		0.065		0.131		-0.230***
		(0.098)		(0.093)		(0.086)
N	5838	5838	5778	5778	5848	5848
R-Squared	0.085	0.086	0.044	0.046	0.038	0.042
Mean of Dep. Var.	-0.013	-0.013	-0.026	-0.026	3.409	3.409

Discussion

1. Controlling for non-cognitive skills reduces the impacts on test scores by about 11%. ▸ Mediation

2. No evidence of classroom environment being affected. ▸ Classroom environment

3. Incentives for cheating exists, no reports and doesn't explain the results well.

4. There are no long-term effects, despite the relatively large short-term effects.
   - Temporary decreases in effort and motivation while the incentives were in place did not persist after the incentives.
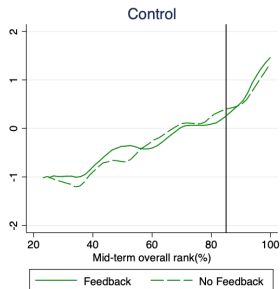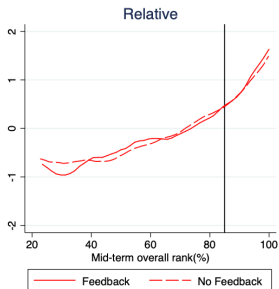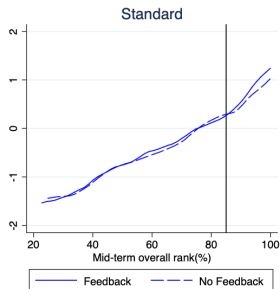
Test Results–Feedback
Non-parametric estimation

Final Exam

Conclusion
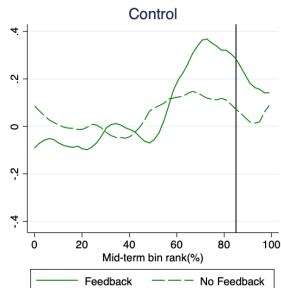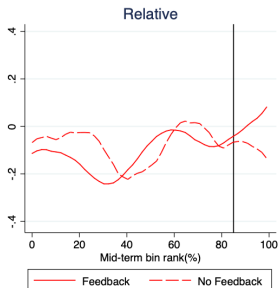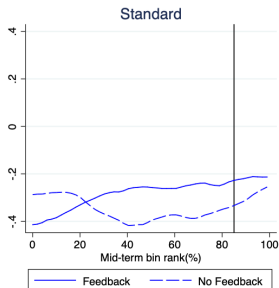
1. Financial incentives may not be successful in promoting educational achievement if such incentives have negative psychological effects
   - *Standard* merit-based scholarship significantly decreased test scores.
     - The largest decreases concentrated among those least likely to win.
     - The pattern corresponds to decreases in motivation to study among those least likely to win.
   - No such negative impacts among the *Relative* scholarship group.
2. Limited evidence that feedback on ranking may influences test scores.
   - It may increase test scores for initially high-performing students.
3. Tournament incentive schemes such as the standard scholarship may exacerbate inequality in education outcomes.
   - especially in environments where students knew their baseline ranking.

When Student Incentives Don't Work
Evidence from a Field Experiment in Malawi

Jim Berry[1]   Hyuncheol Bryant Kim[2]   Hyuk Son[2]

[1]Delaware University
[2]Cornell University

December, 2019

hs924@cornell.edu

Sample Composition

Panel A: Scholarship Treatment (Grade 5-8)

Scholarship Assignment	School-Grades	Students
Standard scholarship	46	2830
Relative scholarship	42	2993
Control	30	1562
Total	118	7385

Panel B: Scholarship Treatment (Grade 5-6 with long-term follow-up)

Scholarship Assignment	School-Grades	Students
Standard scholarship	24	1869
Relative scholarship	24	2000
Control	13	693
Total	61	4562

Panel C: Feedback Treatment (Grade 5-7)

Scholarship Assignment	Feedback Assignment	Students
Standard scholarship	No Feedback	1175
	Feedback	1195
Relative scholarship	No Feedback	1360
	Feedback	1362
Control	No Feedback	510
	Feedback	501
Total		6103

◂ Design

Scholarship Randomization

1. Announcement (Feb. 2015)
   - All three treatment arms were explained to students.
   - Treatment assignment and the student's baseline exam score were distributed.
   - Individualized cards were used for the announcement.
   - Took a short quiz to measure understanding of the programs.

```
ID    XXXXXXX School XXX
STD  7            Name XXX
Group A
Current Position
      25% [759 out of 1928]
You can receive a present when you are reanked at:
      15%(455th) or above
```

Standard merit-based scholarship

```
ID    XXXXXXX School XXX
STD  5            Name XXX
Group B
Current Position
      75% [2286 out of 3037]
      86% [86 out of 100 learners with similar score]
You can receive a present when you are reanked at:
      15th or above among 100 learners of similar score
```

Relative merit-based scholarship

Feedback Randomization

1. 6,470 (87.6%) participated in the mid-term exam (March 2015)
2. Feedback on rank information was provided to feedback group.
   - Control & *Standard* merit-based treatments: midterm overall rank.
   - *Relative* merit-based treatment: midterm bin rank .

Lee (2009) Bounds of Main Test Score Estimates

	Exam Rank			Exam Score (Norm)		
	Main	Lower Bound	Upper Bound	Main	Lower Bound	Upper Bound
	(1)	(2)	(3)	(4)	(5)	(6)
Merit	-7.402**	-8.321**	-6.707*	-0.265*	-0.303**	-0.232*
	(3.671)	(3.671)	(3.671)	(0.138)	(0.138)	(0.138)
Relative merit	-2.516	-3.724	-1.374	-0.046	-0.123	0.003
	(4.730)	(4.730)	(4.730)	(0.187)	(0.187)	(0.187)
N	6586	6586	6586	6586	6586	6586

◂ Attrition

Understanding of Scholarship

	After Announcement		Follow-up	
	(1)	(2)	(3)	(4)
Standard	-0.009	-0.007	-0.021	-0.019
	(0.023)	(0.026)	(0.023)	(0.023)
Relative	0.036	0.041*	-0.028	-0.011
	(0.022)	(0.024)	(0.024)	(0.025)
Std. × Top 15%		-0.015		-0.018
		(0.025)		(0.035)
Rel. × Top 15%		-0.040*		-0.107***
		(0.022)		(0.029)
Top 15%		0.056***		0.091***
		(0.020)		(0.019)
N	5617	5617	5851	5851
R-Squared	0.038	0.047	0.092	0.098
Mean of Dep. Var.	0.924	0.924	0.636	0.636

Expectation of Scholarship

	After Announcement		Follow-up	
	(1)	(2)	(3)	(4)
Standard	0.301***	0.231***	0.442***	0.407***
	(0.057)	(0.059)	(0.043)	(0.046)
Relative	0.358***	0.386***	0.405***	0.409***
	(0.066)	(0.066)	(0.044)	(0.046)
Std. × Top 15%		0.485***		0.212***
		(0.084)		(0.045)
Rel. × Top 15%		-0.135		-0.028
		(0.083)		(0.054)
Top 15%		0.046		0.013
		(0.042)		(0.037)
N	5594	5594	5750	5750
R-Squared	0.097	0.157	0.135	0.145
Mean of Dep. Var.	0.356	0.356	0.579	0.579

Mediation regression

	Dependent variable: Exam score			
	(1)	(2)	(3)	(4)
Standard	-0.234*	-0.298**	-0.245*	-0.299*
	(0.137)	(0.141)	(0.146)	(0.153)
Relative	-0.049	-0.001	-0.139	-0.098
	(0.190)	(0.198)	(0.177)	(0.185)
Std. × Top 15%		0.369		0.294
		(0.224)		(0.222)
Rel. × Top 15%		-0.250		-0.192
		(0.273)		(0.243)
Top 15%		-1.705**		-2.089***
		(0.708)		(0.622)
Non-cognitive cont.	Yes	Yes	Yes	Yes
N	5829	5829	5596	5596
R-Squared	0.269	0.284	0.334	0.342
Mean of Dep. Var.	-0.123	-0.123	-0.117	-0.117

Classroom environment

	Smart students help friends better		Willingness to help friends		Received help from friends	
	(1)	(2)	(3)	(4)	(5)	(6)
Standard	0.072	0.118	-0.038	-0.046	0.084	0.084
	(0.101)	(0.109)	(0.062)	(0.072)	(0.061)	(0.052)
Relative	-0.220	-0.250*	0.013	-0.008	-0.051	-0.047
	(0.134)	(0.148)	(0.060)	(0.070)	(0.064)	(0.060)
Std. × Top 15%		-0.300*		0.057		-0.001
		(0.162)		(0.101)		(0.144)
Rel. × Top 15%		0.078		0.113		-0.021
		(0.180)		(0.111)		(0.151)
Top 15%		0.239		-0.052		0.003
		(0.155)		(0.100)		(0.136)
N	2680	2680	2679	2679	2672	2672
R-Squared	0.086	0.090	0.018	0.019	0.022	0.022
Mean of Dep. Var.	3.755	3.755	4.074	4.074	3.888	3.888

Classroom environment (cont.)

	Provided help to friends		Asked for help to friends		Classroom competitiveness index	
	(1)	(2)	(3)	(4)	(5)	(6)
Standard	0.078	0.098	0.040	0.039	0.064	0.079
	(0.065)	(0.065)	(0.067)	(0.077)	(0.080)	(0.084)
Relative	0.008	0.048	-0.053	-0.064	-0.087	-0.093
	(0.067)	(0.065)	(0.066)	(0.068)	(0.080)	(0.086)
Std. × Top 15%		-0.129		0.020		-0.094
		(0.157)		(0.166)		(0.128)
Rel. × Top 15%		-0.223		0.087		0.016
		(0.191)		(0.196)		(0.155)
Top 15%		0.117		-0.151		0.035
		(0.153)		(0.137)		(0.123)
N	2674	2674	2680	2680	2682	2682
R-Squared	0.008	0.009	0.009	0.010	0.038	0.038
Mean of Dep. Var.	3.829	3.829	4.095	4.095	-0.010	-0.010

◂ Discussion

Test Results–Feedback
Parametric estimation results

	Dependent Variable: Final exam score (normalized)					
	All		Top 15%		Bottom 15%	
	(1)	(2)	(3)	(4)	(5)	(6)
Feedback	0.028	0.046	0.065	0.084	0.014	0.035
	(0.023)	(0.064)	(0.052)	(0.080)	(0.028)	(0.081)
Standard		-0.324		-0.185		-0.277
		(0.202)		(0.242)		(0.174)
Relative		-0.207		-0.109		-0.108
		(0.227)		(0.248)		(0.211)
Std. × FB		-0.015		-0.006		-0.018
		(0.072)		(0.102)		(0.090)
Rel. × FB		-0.027		-0.036		-0.031
		(0.073)		(0.134)		(0.090)
N	5159	5159	1057	1057	4102	4102
R-Squared	0.308	0.317	0.242	0.247	0.220	0.230
Mean Dep. Var.	-0.186	-0.186	0.846	0.846	-0.452	-0.452

◂ Back