

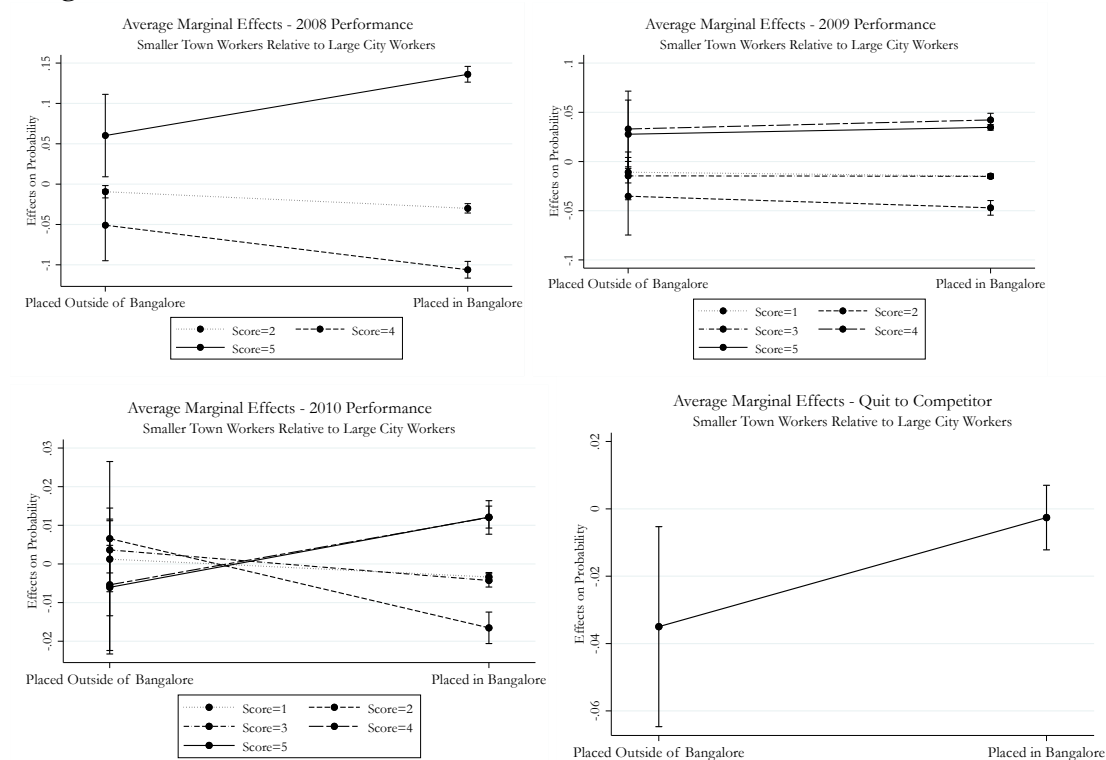
## **APPENDIX**

### **Table of Contents**

Appendix A – Additional Robustness Results for Main Paper	p.2-26
Appendix B – Additional Evidence for our Interpretation of the “Smaller Town” Construct	p.27-28
Appendix C – Employee Random Assignment Protocol	p.29
Appendix D – Human Capital Rents by Worker Origin and Placement Location	p.30-35

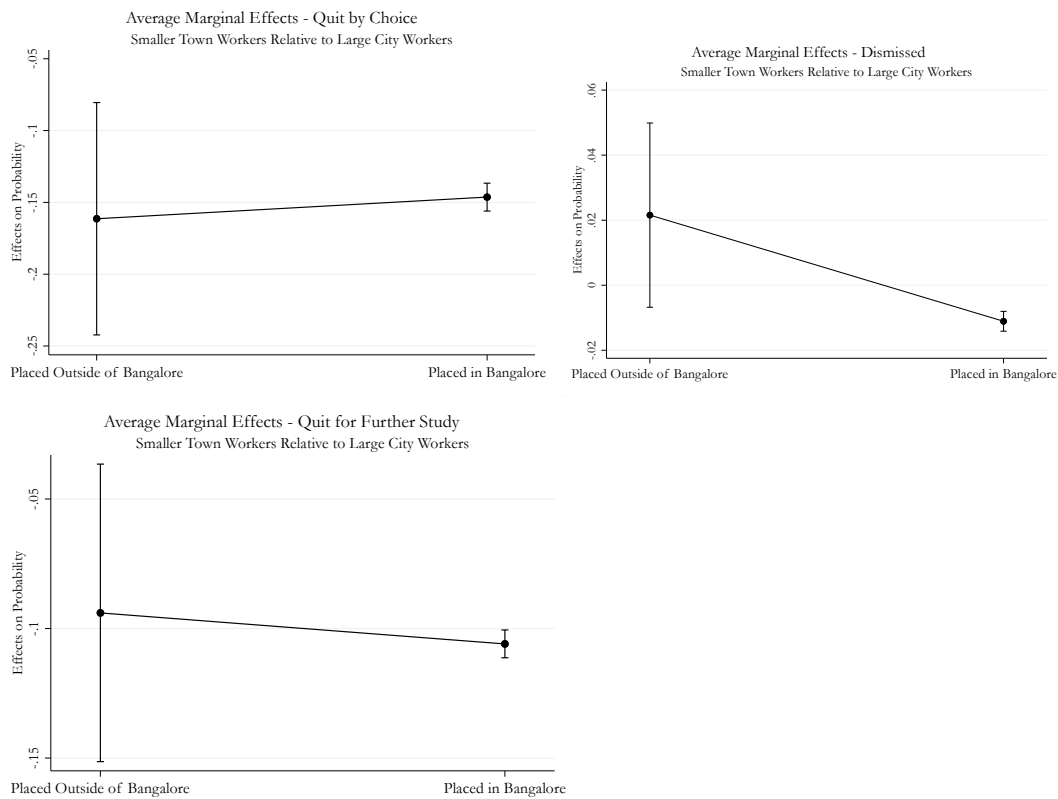
## Figures

**Figure A1**  
**Marginal Effects Plots for Table A4 – Performance and Turnover**



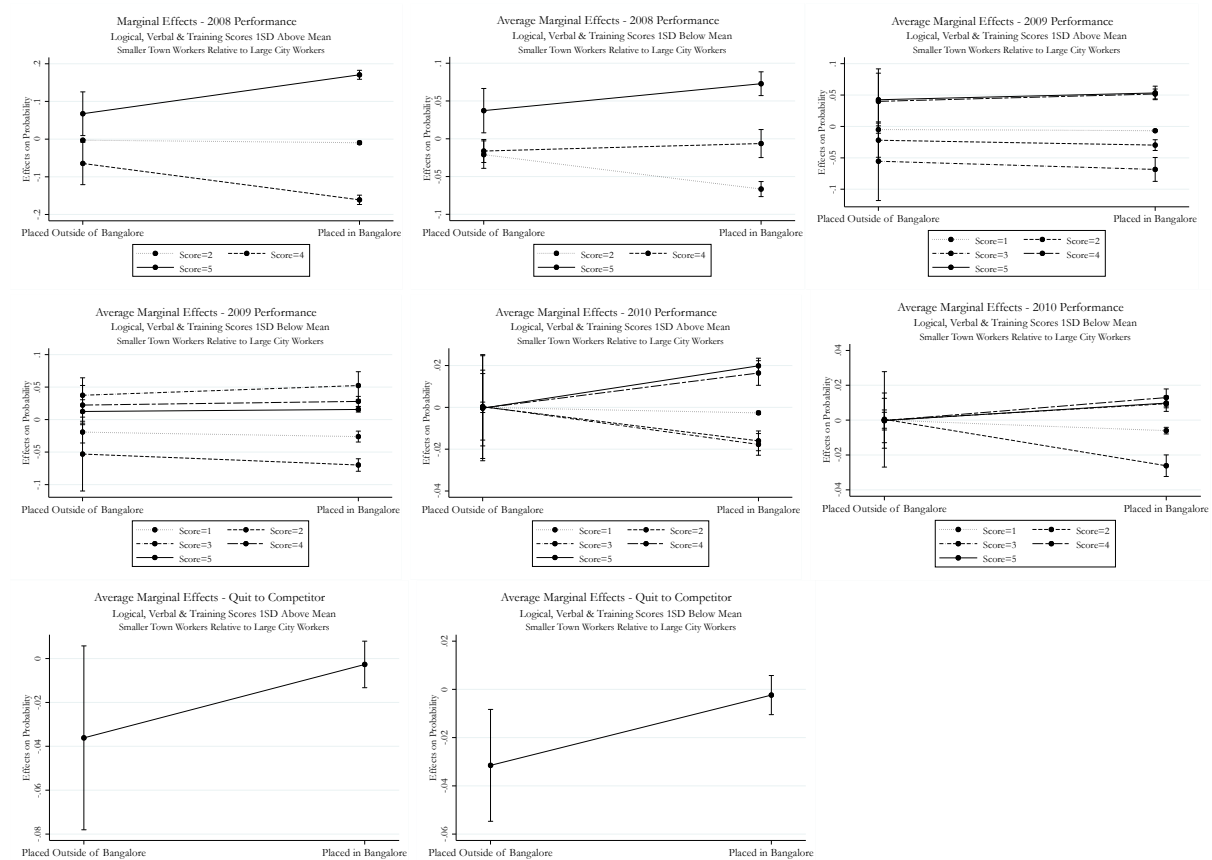
NOTES – Interpreting Marginal Effects: Here we plot the average marginal effects for each model interaction in Table A3. Each line represents the change in *relative* probability of a given outcome for workers from smaller towns compared to workers from large cities who are placed either outside of Bangalore or in Bangalore. Outcomes for *Performance* range between 1 and 5, with 5 being the highest. Outcomes for *Quit to Competitor* take the values of 0 or 1. We obtain these marginal effects using the STATA command “margins *Bangalore*, dydx(*From Smaller Town*)” following each regression. Note that the results remain qualitatively the same if we instead use “margins *Bangalore* # *From Smaller Town*, atmeans”. Different Number of Performance Levels in 2008: While outcomes range between 1 and 5, with 5 being the highest, for the 2007 incoming cohort that satisfied the “nine-month rule” and received a performance rating in 2008, no worker received a score of 1 or 3. We therefore show only the changes in probability of receiving each of the three other scores. For the ratings received in 2009 and 2010, all five score levels were given to workers.

**Figure A2**  
**Marginal Effects Plots for Table A6 – Attrition Types: Quit by Choice, Quit for Further Study and Dismissed**



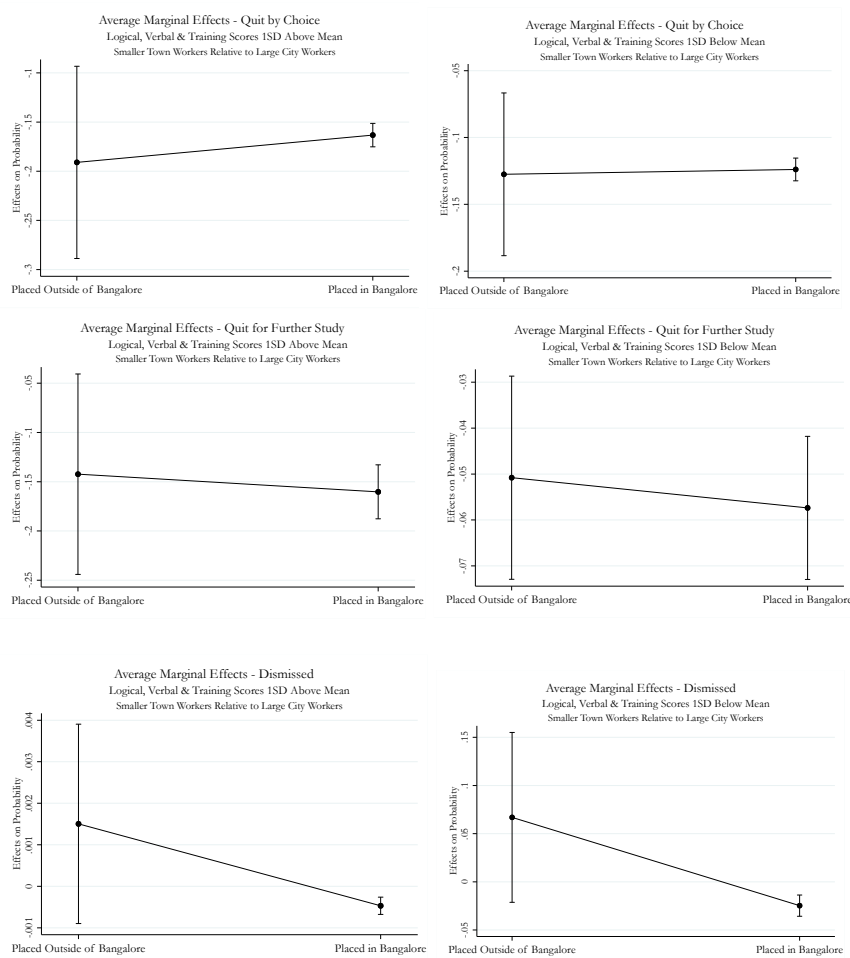
NOTES – Interpreting Marginal Effects: Here we plot the average marginal effects for each model interaction in Table A9. Each line represents the change in *relative* probability of a given outcome for workers from smaller towns compared to workers from large cities who are placed either outside of Bangalore or in Bangalore. We obtain these marginal effects using the STATA command “`margins Bangalore, dydx(From Smaller Town)`” following each regression. Note that the results remain qualitatively the same if we instead use “`margins Bangalore # From Smaller Town, atmeans`”.

**Figure A3**  
**Marginal Effects Plots for Table A4 – Performance and Turnover – At Alternative Estimates**



**NOTES – Interpreting Marginal Effects:** Here we plot the average marginal effects for each model interaction in Table A3. Each line represents the change in *relative* probability of a given outcome for workers from smaller towns compared to workers from large cities who are placed either outside of Bangalore or in Bangalore. Outcomes for performance range between 1 and 5, with 5 being the highest. However, in contrast to Figure A1, here we condition all estimates at Logical Scores, Verbal Scores and CGPA training at one standard deviation above and below the mean, respectively, keeping gender and hometown location at their mean values (resetting *Male* and *Hometown* to either 0 or 1 does not change the qualitative conclusions of these graphs). We obtain the marginal effects for 1 standard deviation above the mean using the STATA command “margins *Bangalore*, dydx(*From Smaller Town*) at(*Logical Score* = 8.292 *Verbal Score* = 8.279 *CGPA* = 4.886” following each regression. We obtain the marginal effects for 1 standard deviation below the mean using the STATA command “margins *Bangalore*, dydx(*From Smaller Town*) at(*Logical Score* = 1.588 *Verbal Score* = 0.312 *CGPA* = 4.146” following each regression. The raw means and standard deviations of our control variables are available in Table 2 in the main manuscript. As the results in these figures show, our conclusions remain the same as the marginal effects conditioned at the means in Figure A1. **Different Number of Performance Levels in 2008:** While outcomes range between 1 and 5, with 5 being the highest, for the 2007 incoming cohort that satisfied the “nine-month rule” and received a performance rating in 2008, no worker received a score of 1 or 3. We therefore show only the changes in probability of receiving each of the three other scores. For the ratings received in 2009 and 2010, all five score levels were given to workers.

**Figure A4**  
**Marginal Effects Plots for Table A6 – Attrition Types: Quit by Choice, Quit for Further Study and Dismissed – Alternative Estimates**



NOTES – Interpreting Marginal Effects: Here we plot the average marginal effects for each model interaction in Table A9. Each line represents the change in *relative* probability of a given outcome for workers from smaller towns compared to workers from large cities who are placed either outside of Bangalore or in Bangalore. We obtain the marginal effects for 1 standard deviation above the mean using the STATA command “margins *Bangalore*, dydx(*From Smaller Town*) at(*Logical Score* = 8.292 *Verbal Score* = 8.279 *CGPA* = 4.886” following each regression. We obtain the marginal effects for 1 standard deviation below the mean using the STATA command “margins *Bangalore*, dydx(*From Smaller Town*) at(*Logical Score* = 1.588 *Verbal Score* = 0.312 *CGPA* = 4.146” following each regression. The raw means and standard deviations of our control variables are available in Table 2 in the main manuscript. As the results in these figures show, our conclusions remain the same as the marginal effects conditioned at the means in Figure A2.

## Tables

**Table A1**  
**University Rankings and Location**

India Rank 2021	World University Rank 2021	University	City	City Tier	State
1	301–350	Indian Institute of Science	Bangalore	1	Karnataka
2	351–400	Indian Institute of Technology Ropar	Rupnagar	3	Punjab
3	401–500	Indian Institute of Technology Indore	Indore	2	Madhya Pradesh
4	601–800	Banaras Hindu University	Varanasi	2	Uttar Pradesh
4	601–800	Institute of Chemical Technology	Mumbai	1	Maharashtra
4	601–800	University of Delhi	Delhi	1	Delhi
4	601–800	Indian Institute of Science Education and Research, Pune	Pune	2	Maharashtra
4	601–800	Indian Institute of Science Education and Research Kolkata	Kolkata	1	West Bengal
4	601–800	Indian Institute of Technology Gandhinagar	Gandhinagar	2	Gujarat
4	601–800	Indian Institute of Technology Hyderabad	Hyderabad	1	Telangana
4	601–800	Indraprastha Institute of Information Technology Delhi	New Delhi	3	Delhi
4	601–800	Jamia Millia Islamia	Jamia Nagar	3	Delhi
4	601–800	Jawaharlal Nehru University	New Delhi	3	Delhi
4	601–800	King George's Medical University	Lucknow	2	Uttar Pradesh
4	601–800	Mahatma Gandhi University	Kottayam	3	Kerala
4	601–800	Panjab University	Chandigarh	2	Punjab
4	601–800	Savitribai Phule Pune University	Pune	2	Maharashtra
4	601–800	Thapar University	Patiala	2	Punjab
19	801–1000	Aligarh Muslim University	Aligarh	2	Uttar Pradesh
19	801–1000	Amrita Vishwa Vidyapeetham	Coimbatore	2	Tamil Nadu
19	801–1000	Bharathiar University	Coimbatore	2	Tamil Nadu
19	801–1000	Delhi Technological University	New Delhi	3	Delhi
19	801–1000	Guru Jambheshwar University of Science and Technology	Hisar	3	Haryana
19	801–1000	Indian Institute of Technology Bhubaneswar	Bhubaneswar	2	Odisha
19	801–1000	Indian Institute of Technology (Indian School of Mines) Dhanbad	Dhanbad	2	Jharkand
19	801–1000	Indian Institute of Science Education and Research Bhopal	Bhopal	2	Madhya Pradesh
19	801–1000	Jadavpur University	Kolkata	1	West Bengal
19	801–1000	Jamia Hamdard University	New Delhi	3	Delhi
19	801–1000	National Institute of Technology Rourkela	Rourkela	3	Odisha
19	801–1000	National Institute of Technology Silchar	Silchar	3	Assam
19	801–1000	Osmania University	Hyderabad	1	Telangana
19	801–1000	Sri Venkateswara University	Andhra	3	Andhra Pradesh
19	801–1000	VIT University	Vellore	2	Tamil Nadu
34	1001+	Acharya Nagarjuna University	Guntur	2	Andhra Pradesh
34	1001+	Amity University	Noida	3	National Capital Region
34	1001+	Andhra University	Visakhapatnam	2	Andhra Pradesh
34	1001+	Anna University	Chennai	1	Tamil Nadu
34	1001+	Annamalai University	Cuddalore	3	Tamil Nadu

## Online appendix for Firm-Induced Migration Paths and Strategic Human-Capital Outcomes

34	1001+	Banasthali University	Vanasthali	3	Rajasthan
34	1001+	Birla Institute of Technology and Science, Pilani	Pilani	3	Rajasthan
34	1001+	University of Calcutta	Kolkata	1	West Bengal
34	1001+	Cochin University of Science and Technology	Cochin	3	Kerala
34	1001+	GITAM University	Visakhapatnam	2	Andhra Pradesh
34	1001+	Indian Institute of Technology Patna	Bihta	3	Bihar
34	1001+	Jawaharlal Nehru Technological University Anantapur (JNTUA)	Ananthapuramu	3	Andhra Pradesh
34	1001+	KIIT University	Bhubaneswar	2	Odisha
34	1001+	KL University	Guntur	2	Andhra Pradesh
34	1001+	University of Lucknow	Lucknow	2	Uttar Pradesh
34	1001+	Maharaja Sayajirao University of Baroda	Vadodara	2	Gujarat
34	1001+	Manipal Academy of Higher Education	Manipal	3	Karnataka
34	1001+	University of Mumbai	Mumbai	1	Maharashtra
34	1001+	University of Mysore	Mysuru	3	Karnataka
34	1001+	National Institute of Technology, Tiruchirappalli	Tiruchirappalli	3	Tamil Nadu
34	1001+	Pondicherry University	Puducherry	3	Puducherry
34	1001+	PSG College of Technology	Coimbatore	2	Tamil Nadu
34	1001+	SASTRA University	Thanjavur	3	Tamil Nadu
34	1001+	Sathyabama Institute of Science and Technology	Chennai	1	Tamil Nadu
34	1001+	Saveetha University	Chennai	1	Tamil Nadu
34	1001+	Siksha 'O' Anusandhan	Bhubaneswar	2	Odisha
34	1001+	SRM Institute of Science and Technology	Kattankulathur	3	Tamil Nadu
34	1001+	Tamil Nadu Agricultural University	Coimbatore	2	Tamil Nadu
34	1001+	Tezpur University	Sonitpur	3	Assam
34	1001+	Visvesvaraya National Institute of Technology, Nagpur	Nagpur	3	Maharashtra

NOTES – While unfortunately no comprehensive ranking is available for all top universities in India at the time of our study, given the slow evolution of university rankings over time (Bejan, 2007), we feel confident relying on a 2021 ranking to examine the geographic distribution of universities. This ranking is provided by the Times Higher Education and is available for the top 63 internationally ranked universities and colleges India. This table reproduces the 2021 India and World University Ranking produced by Times Higher Education. The original data can be obtained at: <https://www.timeshighereducation.com/student/best-universities/best-universities-india>.

**Table A2**  
**Origin Town Size - Continuous Population Measures**

VARIABLES	(1) Average Performance OLS	(2) Average Performance OLS	(3) Average Performance OLS w/FE	(4) Average Performance OLS	(5) Average Performance OLS	(6) Average Performance OLS w/FE	(7) Quit to Competitor OLS	(8) Quit to Competitor OLS	(9) Quit to Competitor OLS w/FE	(10) Quit to Competitor OLS	(11) Quit to Competitor OLS	(12) Quit to Competitor OLS w/FE
Placed in Bangalore	-0.043 (0.050)	0.135* (0.059)		-0.063 (0.045)	0.028 (0.042)		0.045*** (0.008)	0.087*** (0.016)		0.032*** (0.004)	0.051*** (0.010)	
Avg Origin Population	-0.023* (0.011)	-0.013 (0.010)	-0.009 (0.011)				0.002 (0.003)	0.004 (0.003)	0.003 (0.003)			
Avg Origin Population * Bangalore		-0.044*** (0.011)	-0.049*** (0.012)					-0.010*** (0.002)	-0.009*** (0.003)			
Avg Origin Population (Relaxed)				-0.012 (0.008)	-0.006 (0.010)	-0.004 (0.010)				0.004 (0.002)	0.005 (0.003)	0.004 (0.003)
Avg Origin Population (Relaxed) * Bangalore					-0.025** (0.010)	-0.028** (0.011)					-0.005* (0.003)	-0.005 (0.003)
Score Logical	-0.002 (0.011)	-0.003 (0.010)	-0.004 (0.009)	-0.004 (0.008)	-0.004 (0.008)	-0.005 (0.007)	0.001 (0.004)	0.001 (0.004)	0.000 (0.004)	0.001 (0.003)	0.001 (0.003)	0.001 (0.003)
Score Verbal	0.007 (0.008)	0.008 (0.007)	0.009 (0.008)	0.001 (0.006)	0.001 (0.006)	0.003 (0.006)	-0.004* (0.002)	-0.004* (0.002)	-0.004* (0.002)	-0.005** (0.002)	-0.005** (0.002)	-0.005** (0.002)
CGPA	0.690*** (0.037)	0.687*** (0.036)	0.713*** (0.041)	0.619*** (0.050)	0.617*** (0.050)	0.637*** (0.048)	0.050 (0.038)	0.049 (0.038)	0.051 (0.039)	0.053 (0.030)	0.053 (0.029)	0.053 (0.030)
Male	0.116** (0.040)	0.113** (0.039)	0.123** (0.037)	0.096* (0.050)	0.097* (0.050)	0.103* (0.048)	0.015 (0.015)	0.014 (0.014)	0.012 (0.016)	0.026 (0.014)	0.026 (0.015)	0.025 (0.015)
Constant	0.181 (0.188)	0.158 (0.201)	0.045 (0.233)	0.505* (0.266)	0.494 (0.273)	0.395 (0.276)	-0.162 (0.182)	-0.166 (0.188)	-0.152 (0.193)	-0.184 (0.142)	-0.187 (0.146)	-0.177 (0.148)
Observations	842	842	842	1,187	1,187	1,187	1,032	1,032	1,032	1,448	1,448	1,448
R-squared	0.091	0.098	0.118	0.078	0.080	0.098	0.015	0.017	0.024	0.017	0.018	0.019
Location FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes

NOTES – Standard errors in parentheses are clustered at the production centre level. This table replicates Table 3 in the main Results section with two alternative measures of small-town worker origin. The first measure, *Avg. Origin Population*, calculates the average population size across the worker’s school, high school, and university towns. The second measure, *Avg. Origin Population (Relaxed)*, also calculates the average population size across the worker’s school, high school, and university towns, but allows for missing data for one of the worker’s origin locations.

\*p<0.1; \*\*p<.05; \*\*\*p<.01

**Table A3**  
**Non-linear Estimation of Performance and Turnover**

VARIABLES	(1) Performance 2008 Ordered Logit	(2) Performance 2008 Ordered Logit	(3) Performance 2008 BUC Ordered Logit	(4) Performance 2009 Ordered Logit	(5) Performance 2009 Ordered Logit	(6) Performance 2009 BUC Ordered Logit	(7) Performance 2010 Ordered Logit	(8) Performance 2010 Ordered Logit	(9) Performance 2010 BUC Ordered Logit	(10) Quit to Competitor Logit	(11) Quit to Competitor Logit	(12) Quit to Competitor Conditional Logit
Placed in Bangalore	-0.367*** (0.104)	-0.519*** (0.120)		-0.097 (0.072)	-0.128* (0.075)		-0.224*** (0.072)	-0.276*** (0.080)		0.576*** (0.115)	0.426*** (0.150)	
From Smaller Town	0.382*** (0.123)	0.284** (0.126)	0.267 (0.180)	0.358** (0.159)	0.335* (0.201)	0.287 (0.237)	-0.020 (0.080)	-0.060 (0.092)	-0.101 (0.075)	-0.479* (0.275)	-0.712** (0.307)	-0.670** (0.341)
From Smaller Town * Bangalore		0.410*** (0.125)	0.481*** (0.183)		0.107 (0.201)	0.063 (0.234)		0.199* (0.103)	0.232*** (0.082)		0.684** (0.272)	0.640** (0.303)
Received 2008 Rating				0.256*** (0.083)	0.256*** (0.084)	0.226** (0.093)	0.266*** (0.057)	0.265*** (0.057)	0.276*** (0.057)			
Received 2009 Rating							1.392*** (0.314)	1.416*** (0.324)	1.465*** (0.276)			
Logical Score	0.083*** (0.025)	0.084*** (0.026)	0.086*** (0.027)	-0.000 (0.029)	-0.000 (0.029)	-0.009 (0.032)	0.002 (0.025)	0.002 (0.025)	-0.004 (0.025)	0.002 (0.058)	0.002 (0.058)	-0.001 (0.057)
Verbal Score	0.002 (0.015)	0.003 (0.014)	0.001 (0.015)	-0.000 (0.023)	-0.000 (0.023)	0.009 (0.024)	-0.010 (0.016)	-0.010 (0.016)	-0.010 (0.016)	-0.056** (0.023)	-0.055** (0.023)	-0.056** (0.022)
CGPA Training	2.109*** (0.278)	2.114*** (0.276)	2.122*** (0.248)	1.946*** (0.257)	1.947*** (0.257)	2.115*** (0.246)	1.270*** (0.182)	1.272*** (0.181)	1.347*** (0.159)	0.780 (0.573)	0.779 (0.574)	0.822 (0.586)
Male	0.219 (0.208)	0.221 (0.208)	0.147 (0.212)	0.173 (0.185)	0.173 (0.185)	0.222 (0.174)	0.462*** (0.085)	0.463*** (0.084)	0.518*** (0.095)	0.338 (0.220)	0.335 (0.219)	0.324 (0.233)
Placed in Hometown	0.061 (0.176)	0.060 (0.175)	0.064 (0.178)	-0.296** (0.142)	-0.296** (0.142)	-0.321** (0.160)	0.006 (0.108)	0.007 (0.109)	0.000 (0.107)	-0.562 (0.358)	-0.558 (0.357)	-0.582 (0.355)
Constant										-6.028** (2.819)	-5.988** (2.797)	
Observations	511	511	1,014	933	933	3,681	903	903	3,563	1,208	1,208	1,177
Location FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes

NOTES – Standard errors in parentheses are clustered at the production centre level. Note on the differences in the number of observations for performance regressions: For new hires, workers’ training schedule affected whether they satisfied the “nine-month rule” in their first year on the job. For instance, in the 2007 sample, workers who started their training after September 2007 would not finish until early 2008. Most of those workers were not assigned to a project prior to March 2008, making them ineligible to receive a 2008 performance rating. As a result, the number of workers receiving the 2008 performance rating is lower than the number of workers receiving the 2009 performance rating. However, INDTECH’s decision to deploy a worker to a project did not depend on superior ability or observable and/or unobservable characteristics.<sup>1</sup> Interpreting Marginal Effects: Figure A1 in the Appendix plots the average marginal effects for each model interaction, obtained using the STATA command “margins *Bangalore*, dydx(*From Smaller Town*)”. Note that the results remain qualitatively the same if we instead use “margins *Bangalore* # *From Smaller Town*, atmeans”. All results reported here are robust to re-running all models with OLS with and without production center fixed effects.

\*p<0.1; \*\*p<.05; \*\*\*p<.01

<sup>1</sup> We also empirically validate that controlling for the training batch, logical and verbal scores, gender, and smaller town origin were not significant predictors of receiving a performance rating in 2008; only CGPA training was a positive predictor of receiving a performance rating in 2008, so we include this control in all further specifications.

**Table A4**  
**Worker Performance & Turnover to Competitors – Matched Sample Analysis – Coarsened Exact Matching – Automatic Coarsening**

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Average Performance	Average Performance	Average Performance	Quit to Competitor	Quit to Competitor	Quit to Competitor
	(z-score)	(z-score)	(z-score)			
	OLS	OLS	OLS	OLS	OLS	OLS
	CEM Automatic Bins	CEM Automatic Bins	CEM Automatic Bins	CEM Automatic Bins	CEM Automatic Bins	CEM Automatic Bins
From Smaller Town	0.271 (0.148)	0.236 (0.185)	0.178 (0.178)	-0.030 (0.020)	-0.049*** (0.010)	-0.049*** (0.010)
Placed in Bangalore	-0.005 (0.110)	-0.067 (0.172)		0.058*** (0.007)	0.027** (0.009)	
From Smaller Town * Bangalore		0.174 (0.194)	0.231 (0.186)		0.090*** (0.011)	0.090*** (0.011)
Logical Score	0.004 (0.019)	0.003 (0.019)	0.004 (0.019)	-0.001 (0.008)	-0.001 (0.008)	-0.001 (0.008)
Verbal Score	-0.001 (0.021)	-0.000 (0.021)	0.006 (0.021)	-0.005** (0.002)	-0.005** (0.002)	-0.005** (0.002)
CGPA Training	0.847*** (0.111)	0.845*** (0.112)	0.910*** (0.105)	0.038 (0.058)	0.037 (0.058)	0.038 (0.061)
Male	0.218 (0.225)	0.214 (0.223)	0.218 (0.222)	0.039 (0.034)	0.037 (0.033)	0.039 (0.034)
Placed in Hometown	0.094 (0.108)	0.094 (0.108)	0.067 (0.101)	-0.026 (0.036)	-0.025 (0.035)	-0.028 (0.035)
Constant	-4.165*** (0.463)	-4.136*** (0.442)	-4.458*** (0.495)	-0.088 (0.311)	-0.075 (0.305)	-0.074 (0.318)
Observations	569	569	569	651	651	651
R-squared	0.088	0.089	0.117	0.022	0.026	0.031
Location FE	No	No	Yes	No	No	Yes

NOTES – Standard errors in parentheses are clustered at the location level. Coarsened Exact Matching (CEM) matches workers on their pre-location assignment characteristics of Logical Scores, Verbal Scores, CGPA Training and Gender. However, the results are also generally robust to matching on Logical and Verbal Scores only. CEM Bins are determined with STATA’s in-built automatic identification of appropriate bins using Sturge’s Rule (i.e., the number of bins is equal to  $1+3.322\log(n)$  where n is the number of observations), and results in a multivariate L1 distance equal to 0. 581 with 136 matched strata comprising of 232 treated and 419 untreated observations.

\*p<0.1; \*\*p<.05; \*\*\*p<.01

**Table A5**  
**Worker Performance & Turnover to Competitors – Matched Sample Analysis – Coarsened Exact Matching – Natural Breakpoints**

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Average Performance (z-score)	Average Performance (z-score)	Average Performance (z-score)	Quit to Competitor	Quit to Competitor	Quit to Competitor
	OLS	OLS	OLS	OLS	OLS	OLS
	CEM Manual Natural Bins	CEM Manual Natural Bins	CEM Manual Natural Bins	CEM Manual Natural Bins	CEM Manual Natural Bins	CEM Manual Natural Bins
From Smaller Town	0.171 (0.139)	0.093 (0.148)	0.094 (0.152)	-0.052* (0.024)	-0.069** (0.021)	-0.071** (0.023)
Placed in Bangalore	-0.033 (0.060)	-0.173 (0.105)		0.064*** (0.012)	0.034 (0.019)	
From Smaller Town * Bangalore		0.381** (0.137)	0.379** (0.140)		0.082*** (0.022)	0.084*** (0.024)
Logical Score	0.002 (0.019)	0.001 (0.019)	0.000 (0.019)	-0.002 (0.009)	-0.002 (0.009)	-0.002 (0.009)
Verbal Score	0.015 (0.013)	0.015 (0.013)	0.017 (0.013)	-0.004 (0.003)	-0.004 (0.003)	-0.004 (0.003)
CGPA Training	0.988*** (0.114)	0.986*** (0.116)	1.018*** (0.113)	0.079 (0.055)	0.078 (0.055)	0.067 (0.059)
Male	0.180* (0.097)	0.189* (0.098)	0.208* (0.105)	0.038 (0.031)	0.040 (0.032)	0.039 (0.033)
Placed in Hometown	-0.068 (0.113)	-0.058 (0.111)	-0.065 (0.109)	-0.046 (0.036)	-0.044 (0.036)	-0.045 (0.035)
Constant	-4.683*** (0.477)	-4.639*** (0.481)	-4.848*** (0.446)	-0.254 (0.285)	-0.244 (0.281)	-0.184 (0.300)
Observations	454	454	454	526	526	526
R-squared	0.115	0.121	0.136	0.041	0.044	0.054
Location FE	No	No	Yes	No	No	Yes

NOTES – Standard errors in parentheses are clustered at the production center level. All models match workers on their pre-location assignment characteristics of Logical Scores, Verbal Scores, CGPA Training and Gender. However, the results are also generally robust to matching on Logical and Verbal Scores only. Coarsened Exact Matching (CEM) bins are determined based on natural breakpoints in the data within Logical and Verbal scores. Breakpoints for CGPA Training scores are determined with STATA’s in-built Sturge’s Rule (i.e., the number of bins is equal to  $1+3.322\log(n)$  where  $n$  is the number of observations). This approach leads to a multivariate L1 distance of 0.468, with 155 matched strata containing 209 treated and 317 untreated observations.

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

**Table A6**  
**Worker Performance & Turnover to Competitors – Matched Sample Analysis – Coarsened Exact Matching – Decile Breakpoints**

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Average Performance	Average Performance	Average Performance	Quit to Competitor	Quit to Competitor	Quit to Competitor
	(z-score)	(z-score)	(z-score)			
	OLS	OLS	OLS	OLS	OLS	OLS
	CEM Manual Decile Bins	CEM Manual Decile Bins	CEM Manual Decile Bins	CEM Manual Decile Bins	CEM Manual Decile Bins	CEM Manual Decile Bins
From Smaller Town	0.138 (0.159)	0.010 (0.130)	0.038 (0.127)	-0.040 (0.025)	-0.054* (0.028)	-0.058* (0.031)
Placed in Bangalore	-0.099 (0.077)	-0.354*** (0.105)		0.075*** (0.015)	0.050* (0.023)	
From Smaller Town * Bangalore		0.618*** (0.126)	0.592*** (0.123)		0.064* (0.028)	0.068* (0.030)
Logical Score	0.010 (0.034)	0.010 (0.034)	0.014 (0.038)	-0.005 (0.005)	-0.005 (0.005)	-0.006 (0.006)
Verbal Score	-0.014 (0.027)	-0.014 (0.027)	-0.018 (0.027)	-0.001 (0.004)	-0.001 (0.004)	-0.001 (0.004)
CGPA Training	1.007*** (0.100)	0.982*** (0.110)	0.964*** (0.117)	0.062* (0.029)	0.062* (0.028)	0.068* (0.030)
Male	0.210* (0.111)	0.219* (0.111)	0.235* (0.115)	0.032 (0.032)	0.033 (0.032)	0.026 (0.032)
Placed in Hometown	-0.046 (0.140)	-0.047 (0.141)	-0.059 (0.139)	-0.020 (0.029)	-0.020 (0.029)	-0.019 (0.024)
Constant	-4.730*** (0.465)	-4.560*** (0.530)	-4.585*** (0.540)	-0.192 (0.126)	-0.184 (0.121)	-0.190 (0.127)
Observations	281	281	281	332	332	332
R-squared	0.129	0.144	0.178	0.040	0.042	0.057
Location FE	No	No	Yes	No	No	Yes

NOTES – Standard errors in parentheses are clustered at the production center level. All models match workers on their pre-location assignment characteristics of Logical Scores, Verbal Scores, CGPA Training and Gender. However, the results are also generally robust to matching on Logical and Verbal Scores only. Coarsened Exact Matching (CEM) bins are determined based on deciles within Logical, Verbal and CGPA Training scores. This approach leads to a multivariate L1 distance of 0.358, with 118 matched strata containing 141 treated and 191 untreated observations.

\*p<0.1; \*\*p<.05; \*\*\*p<.01

**Table A7**  
**Additional Measures of Performance and Turnover**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Dismissed	Dismissed	Dismissed	Quit for Further Study	Quit for Further Study	Quit for Further Study	Quit by Choice	Quit by Choice	Quit by Choice
VARIABLES	OLS	OLS	OLS w/FE	OLS	OLS	OLS w/FE	OLS	OLS	OLS w/FE
From Smaller Town	0.024 (0.017)	0.029 (0.021)	0.026 (0.021)	-0.094*** (0.025)	-0.091** (0.031)	-0.080** (0.028)	-0.157*** (0.035)	-0.160*** (0.043)	-0.146*** (0.041)
Placed in Bangalore	-0.024*** (0.004)	-0.018*** (0.005)		0.006 (0.019)	0.009 (0.025)		0.061** (0.021)	0.057* (0.030)	
From Smaller Town * Bangalore		-0.022 (0.020)	-0.019 (0.021)		-0.012 (0.030)	-0.024 (0.026)		0.015 (0.040)	0.001 (0.039)
Logical Score	-0.003 (0.002)	-0.003 (0.002)	-0.003 (0.002)	-0.002 (0.003)	-0.002 (0.003)	-0.002 (0.003)	-0.003 (0.005)	-0.003 (0.005)	-0.002 (0.005)
Verbal Score	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.013*** (0.002)	0.013*** (0.002)	0.012*** (0.002)	0.009*** (0.003)	0.009*** (0.003)	0.009*** (0.003)
CGPA Training	-0.296*** (0.044)	-0.296*** (0.044)	-0.293*** (0.044)	0.123** (0.040)	0.123** (0.041)	0.113** (0.037)	0.195*** (0.045)	0.196*** (0.045)	0.184*** (0.045)
Male	0.011 (0.009)	0.011 (0.009)	0.012 (0.009)	0.027** (0.010)	0.027** (0.010)	0.020 (0.011)	-0.027 (0.026)	-0.027 (0.026)	-0.036 (0.028)
Placed in Hometown	-0.007 (0.013)	-0.007 (0.013)	-0.009 (0.013)	0.000 (0.031)	0.000 (0.031)	0.006 (0.031)	-0.025 (0.027)	-0.025 (0.027)	-0.020 (0.026)
Constant	1.388*** (0.201)	1.387*** (0.202)	1.374*** (0.203)	-0.424** (0.165)	-0.424** (0.165)	-0.377** (0.161)	-0.505** (0.185)	-0.504** (0.184)	-0.440* (0.201)
Observations	1,208	1,208	1,208	1,208	1,208	1,208	1,208	1,208	1,208
R-squared	0.259	0.260	0.262	0.049	0.049	0.061	0.056	0.056	0.068
Location FE	No	No	Yes	No	No	Yes	No	No	Yes

NOTES – Standard errors in parentheses are clustered at the production center level. The variable *Dismissed* takes the value of 1 if INDTECH dismissed the focal worker from his/her job within the first three years of their employment and 0 otherwise. Dismissals are driven almost entirely by low performance. For instance, no workers receiving at least the middle performance rating were dismissed, but 70% of those receiving the lowest performance rating were dismissed. However, the dismissal outcomes are also available for workers who left INDTECH before they were eligible to receive their first performance rating. Overall, across the full sample of 1,665 workers, 5.3% were dismissed. *Quit by Choice* takes the value of 1 if the worker departed INDTECH by 2011 of their own volition and 0 otherwise. *Quit for Further Study* takes the value of 1 if the worker departed INDTECH by 2011 of their own volition and listed further study as the reason and 0 otherwise. Furthermore, all results are robust to re-estimating the models with logit.

\*p<0.1; \*\*p<.05; \*\*\*p<.01

**Table A8**  
**Additional Measures of Performance and Turnover – Non-Linear Models**

VARIABLES	(1) Dismissed Logit	(2) Dismissed Logit	(3) Dismissed Conditional Logit	(4) Quit for Further Study Logit	(5) Quit for Further Study Logit	(6) Quit for Further Study Conditional Logit	(7) Quit by Choice Logit	(8) Quit by Choice Logit	(9) Quit by Choice Conditional Logit
From Smaller Town	0.520 (0.396)	0.662 (0.405)	0.598 (0.421)	-0.831*** (0.211)	-0.809*** (0.261)	-0.722*** (0.232)	-0.774*** (0.168)	-0.808*** (0.209)	-0.751*** (0.200)
Placed in Bangalore	-0.962*** (0.202)	-0.528** (0.232)		0.030 (0.135)	0.047 (0.153)		0.276*** (0.096)	0.243* (0.127)	
From Smaller Town * Bangalore		-1.344*** (0.442)	-1.263*** (0.445)		-0.105 (0.235)	-0.185 (0.208)		0.147 (0.190)	0.092 (0.183)
Logical Score	-0.072** (0.032)	-0.073** (0.033)	-0.088*** (0.031)	-0.010 (0.020)	-0.010 (0.020)	-0.009 (0.019)	-0.011 (0.026)	-0.011 (0.026)	-0.010 (0.025)
Verbal Score	-0.012 (0.041)	-0.013 (0.041)	-0.012 (0.042)	0.087*** (0.014)	0.087*** (0.014)	0.084*** (0.013)	0.041*** (0.013)	0.041*** (0.012)	0.038*** (0.012)
CGPA Training	-4.615*** (0.390)	-4.672*** (0.385)	-4.514*** (0.384)	1.098*** (0.368)	1.098*** (0.369)	1.011*** (0.337)	0.999*** (0.234)	0.999*** (0.233)	0.953*** (0.235)
Male	0.257 (0.196)	0.253 (0.196)	0.239 (0.177)	0.204** (0.094)	0.204** (0.093)	0.153 (0.098)	-0.126 (0.121)	-0.127 (0.121)	-0.170 (0.127)
Placed in Hometown	0.196 (0.322)	0.204 (0.319)	0.124 (0.323)	0.019 (0.236)	0.019 (0.236)	0.065 (0.239)	-0.109 (0.126)	-0.108 (0.125)	-0.086 (0.126)
Constant	16.675*** (1.616)	16.870*** (1.610)		-6.948*** (1.587)	-6.951*** (1.585)		-5.037*** (0.997)	-5.032*** (0.988)	
Observations	1,208	1,208	1,157	1,208	1,208	1,208	1,208	1,208	1,208
Location FE	No	No	Yes	No	No	Yes	No	No	Yes

NOTES – Standard errors in parentheses are clustered at the production center level. Interpreting Marginal Effects: Figure A2 in the Appendix plots the average marginal effects for each model interaction, obtained using the STATA command “margins *Bangalore*, dydx(*From Smaller Town*)”. Note that the results remain qualitatively the same if we instead use “margins *Bangalore* # *From Smaller Town*, atmeans”. All results reported here are robust to re-running all models with OLS with and without production center fixed effects.

\*p<0.1; \*\*p<.05; \*\*\*p<.01

**Table A9**  
**Robustness Test for Table 3 - Error Clustering Validation**

	Average Performance (z-score)						Quit to Competitor					
	OLS		OLS		OLS with FEs		OLS		OLS		OLS with FEs	
	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
From Smaller Town	0.178	0.170	0.116	0.420	0.113	0.420	-0.028	0.130	-0.036	0.120	-0.033	0.140
Placed in Bangalore	-0.073	0.210	-0.155	0.140	-0.410	0.210	0.042	0.000	0.032	0.030	0.051	0.040
From Smaller Town * Bangalore			0.297	0.070	0.301	0.060			0.037	0.060	0.035	0.080
Logical Score	-0.004	0.800	-0.004	0.800	-0.005	0.780	0.000	0.980	0.000	0.970	0.000	0.980
Verbal Score	0.006	0.570	0.006	0.550	0.009	0.480	-0.004	0.060	-0.004	0.060	-0.004	0.060
CGPA Training	0.914	0.000	0.914	0.000	0.945	0.000	0.043	0.260	0.043	0.270	0.045	0.260
Male	0.122	0.150	0.123	0.150	0.139	0.070	0.021	0.330	0.021	0.330	0.020	0.350
Placed in Hometown	-0.023	0.790	-0.022	0.810	-0.027	0.730	-0.032	0.130	-0.032	0.130	-0.033	0.120
Constant	-4.251	0.000	-4.236	0.000	-4.140	0.000	-0.115	0.490	-0.113	0.490	-0.137	0.430
Observations	1,001		1,001		1,001		1,208		1,208		1,208	
R-squared	0.078		0.078		0.095		0.254		0.254		0.257	
Location FE	No		No		Yes		No		No		Yes	

NOTES – In addition to the robustness tests listed in the Results section, we specifically address the potential problem of our relatively small number of clusters (the number of INDTECH placement locations) in Table 3 by re-running these results with an OLS model with placement location fixed effects and standard errors estimated using the paired bootstrap-t clustered estimator with 1000 reps. As described in Cameron, Gelbach, and Miller (2008), this estimator is specifically designed to address the small number of clusters problem, by estimating p-values with the pairs cluster bootstrap-t procedure appropriate for data with small number of clusters (see Cameron, Gelbach & Miller (2008) for more details). As the results in this table indicate, our qualitative conclusions in Table 3 continue to hold. As before, the results show that the difference in smaller town workers’ performance across placements relative to their large city counterparts is larger and significant at the 10% level. Similarly, smaller town workers continue to be significantly more likely to move to competitors when placed in Bangalore, relative to their counterparts outside of Bangalore and relative to large city workers in Bangalore. Control variables also continue to have the same signs and similar levels of significance.

**Table A10**

**Worker Enrollment in Additional Coursework – Matched Sample Analysis – Coarsened Exact Matching – Automatic Coarsening**

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Number of Courses	Number of Courses	% Courses Passed	% Courses Passed	Number of English Courses	Number of English Courses	Level of English Courses	Level of English Courses
	OLS CEM Automatic Bins	OLS CEM Automatic Bins	OLS CEM Automatic Bins	OLS CEM Automatic Bins	OLS CEM Automatic Bins	OLS CEM Automatic Bins	OLS CEM Automatic Bins	OLS CEM Automatic Bins
From Smaller Town	1.095*** (0.253)	1.162*** (0.303)	1.603 (1.357)	0.971 (1.642)	-0.025 (0.035)	-0.047 (0.030)	0.220 (0.274)	0.284 (0.344)
Placed in Bangalore	-0.584*** (0.136)	-0.469** (0.159)	-0.330 (2.483)	-1.829 (3.405)	-0.040 (0.022)	-0.092*** (0.024)	0.709** (0.273)	1.013*** (0.261)
From Smaller Town * Bangalore		-0.332 (0.302)		3.683 (2.462)		0.127*** (0.025)		-0.461 (0.530)
Logical Score	-0.005 (0.065)	-0.004 (0.066)	-0.557 (0.303)	-0.569* (0.300)	0.007 (0.007)	0.006 (0.007)	0.047 (0.031)	0.056 (0.038)
Verbal Score	-0.060 (0.039)	-0.061 (0.039)	0.528* (0.275)	0.549* (0.277)	-0.010 (0.007)	-0.009 (0.008)	0.081 (0.069)	0.084 (0.070)
CGPA Training	-0.875 (0.742)	-0.871 (0.738)	16.386* (7.357)	16.285* (7.435)	0.069 (0.061)	0.066 (0.061)	0.429** (0.167)	0.471** (0.151)
Male	0.251 (0.229)	0.259 (0.228)	-6.027* (2.765)	-5.984* (2.765)	0.010 (0.044)	0.011 (0.045)	-0.416 (0.373)	-0.436 (0.385)
Placed in Hometown	0.205 (0.354)	0.203 (0.351)	-0.445 (2.816)	-0.377 (2.850)	0.012 (0.042)	0.015 (0.041)	0.047 (0.412)	0.065 (0.400)
Constant	7.401* (3.244)	7.352* (3.210)	10.597 (33.842)	11.273 (34.234)	-0.176 (0.286)	-0.153 (0.280)	-0.826 (1.194)	-1.098 (1.190)
Observations	651	651	410	410	410	410	55	55
R-squared	0.054	0.054	0.093	0.094	0.019	0.024	0.203	0.206
Location FE	No	No	No	No	No	No	No	No

NOTES – Standard errors in parentheses are clustered at the production center level. All models match workers on their pre-location assignment characteristics of Logical Scores, Verbal Scores, CGPA Training and Gender. However, the results are also generally robust to matching on Logical and Verbal Scores only. Coarsened Exact Matching (CEM) bins are calculated using STATA’s in-built identification of appropriate bins using Sturge’s Rule (i.e., the number of bins is equal to  $1+3.322\log(n)$  where n is the number of observations). These modes result in a multivariate L1 distance equal to 0.581 with 136 matched strata comprising of 232 treated and 419 untreated observations.

\*p<0.1; \*\*p<.05; \*\*\*p<.01

**Table A11**  
**Worker Enrollment in Additional Coursework – Matched Sample Analysis – Coarsened Exact Matching – Natural Breakpoints**

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Number of Courses	Number of Courses	% Courses Passed	% Courses Passed	Number of English Courses	Number of English Courses	Level of English Courses	Level of English Courses
	OLS CEM Manual Natural Bins	OLS CEM Manual Natural Bins	OLS CEM Manual Natural Bins	OLS CEM Manual Natural Bins	OLS CEM Manual Natural Bins	OLS CEM Manual Natural Bins	OLS CEM Manual Natural Bins	OLS CEM Manual Natural Bins
From Smaller Town	0.921* (0.416)	1.081* (0.479)	1.324 (1.885)	0.190 (1.885)	-0.048 (0.052)	-0.087* (0.044)	0.633 (0.418)	0.584 (0.400)
Placed in Bangalore	-0.950*** (0.207)	-0.673* (0.294)	-0.182 (2.154)	-2.927 (3.184)	-0.050 (0.033)	-0.144** (0.045)	0.280 (0.589)	-0.853 (0.862)
From Smaller Town * Bangalore		-0.767 (0.489)		6.841** (2.740)		0.235*** (0.033)		1.252** (0.491)
Logical Score	0.020 (0.067)	0.022 (0.068)	-1.150* (0.516)	-1.182** (0.507)	0.005 (0.006)	0.004 (0.006)	0.035 (0.036)	0.033 (0.035)
Verbal Score	-0.071 (0.051)	-0.072 (0.051)	0.681*** (0.173)	0.716*** (0.158)	-0.007 (0.007)	-0.005 (0.008)	0.073 (0.062)	0.081 (0.063)
CGPA Training	-0.410 (0.648)	-0.402 (0.642)	15.761* (7.161)	15.685* (7.235)	0.136 (0.077)	0.133 (0.081)	-0.023 (0.498)	-0.151 (0.578)
Male	0.151 (0.233)	0.137 (0.232)	-4.674 (3.332)	-4.423 (3.370)	-0.066* (0.029)	-0.058* (0.026)	-0.684** (0.282)	-0.706* (0.307)
Placed in Hometown	0.054 (0.196)	0.038 (0.194)	-2.311 (4.339)	-2.196 (4.399)	0.017 (0.049)	0.021 (0.049)	-0.505 (0.387)	-0.545 (0.436)
Constant	5.413* (2.908)	5.320 (2.872)	15.237 (31.197)	15.953 (31.544)	-0.425 (0.296)	-0.401 (0.306)	1.602 (2.319)	2.224 (2.744)
Observations	526	526	318	318	318	318	39	39
R-squared	0.051	0.053	0.097	0.101	0.040	0.057	0.257	0.265
Location FE	No	No	No	No	No	No	No	No

NOTES – Standard errors in parentheses are clustered at the production center level. All models match workers on their pre-location assignment characteristics of Logical Scores, Verbal Scores, CGPA Training and Gender. However, the results are also generally robust to matching on Logical and Verbal Scores only. Coarsened Exact Matching (CEM) bins are determined based on natural breakpoints in the data within Logical and Verbal scores. Breakpoints for CGPA Training scores are determined with STATA’s in-built Sturge’s Rule (i.e., the number of bins is equal to  $1+3.322\log(n)$  where  $n$  is the number of observations). This approach leads to a multivariate L1 distance of 0.468, but is also the most conservative of the three, with only 155 matched strata containing 209 treated and 317 untreated observations.

\* $p < 0.1$ ; \*\* $p < .05$ ; \*\*\* $p < .01$

**Table A12**  
**Worker Enrollment in Additional Coursework – Matched Sample Analysis – Coarsened Exact Matching – Decile Breakpoints**

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Number of Courses	Number of Courses	% Courses Passed	% Courses Passed	Number of English Courses	Number of English Courses
	OLS CEM Manual Decile Bins	OLS CEM Manual Decile Bins	OLS CEM Manual Decile Bins	OLS CEM Manual Decile Bins	OLS CEM Manual Decile Bins	OLS CEM Manual Decile Bins
From Smaller Town	1.140* (0.509)	1.231* (0.619)	1.224 (2.584)	-0.585 (2.104)	-0.039 (0.046)	-0.079* (0.035)
Placed in Bangalore	-1.064*** (0.258)	-0.895** (0.374)	1.570 (2.026)	-3.824 (2.906)	-0.025 (0.034)	-0.145** (0.044)
From Smaller Town * Bangalore		-0.434 (0.579)		11.183*** (2.480)		0.250*** (0.029)
Logical Score	0.181* (0.086)	0.182* (0.086)	-0.856 (0.476)	-0.872 (0.472)	-0.006 (0.012)	-0.006 (0.011)
Verbal Score	-0.161** (0.066)	-0.162** (0.066)	0.425 (0.285)	0.499 (0.277)	0.005 (0.009)	0.006 (0.009)
CGPA Training	0.700 (0.719)	0.704 (0.712)	17.104** (5.250)	16.289** (5.026)	0.180** (0.054)	0.162** (0.052)
Male	-0.435 (0.336)	-0.438 (0.338)	-2.779 (4.518)	-2.647 (4.469)	-0.092* (0.048)	-0.089* (0.047)
Placed in Hometown	-0.535* (0.257)	-0.533* (0.256)	-2.816 (4.564)	-2.978 (4.658)	0.046 (0.050)	0.043 (0.049)
Constant	0.143 (2.942)	0.091 (2.865)	8.378 (22.114)	12.802 (20.943)	-0.604** (0.240)	-0.505* (0.222)
Observations	332	332	200	200	200	200
R-squared	0.119	0.120	0.099	0.110	0.050	0.069
Location FE	No	No	No	No	No	No

NOTES – Standard errors in parentheses are clustered at the production center level. All models match workers on their pre-location assignment characteristics of Logical Scores, Verbal Scores, CGPA Training and Gender. However, the results are also generally robust to matching on Logical and Verbal Scores only. Coarsened Exact Matching (CEM) bins are determined based on deciles within Logical, Verbal and CGPA Training scores. This approach leads to the lowest multivariate L1 distance of 0.358, but is also the most conservative of the three, with only 118 matched strata containing 141 treated and 191 untreated observations.

\*p<0.1; \*\*p<.05; \*\*\*p<.01

**Table A13**  
**2008 Crime Rates, National Crime Records Bureau**

City	Incidence of total cognizable crimes	Population (in 100,000's) as per 2011 census	Rate of total cognizable crimes per 100,000 inhabitants
1 AGRA	4826	13.21	365.3
2 AHMEDABAD	18544	45.19	410.4
3 ALLAHABAD	2068	10.5	197
4 AMRITSAR	2327	10.11	230.2
5 ASANSOL	1676	10.91	153.6
6 BENGALURU	29664	56.87	521.6
7 BHOPAL	11515	14.55	791.4
8 CHENNAI	11829	64.25	184.1
9 COIMBATORE	4180	14.46	289.1
10 DELHI (CITY)	44573	127.91	348.5
11 DHANBAD	1302	10.64	122.4
12 FARIDABAD	4516	10.55	428.1
13 HYDERABAD	18567	55.34	335.5
14 INDORE	15430	16.39	941.4
15 JABALPUR	5128	11.17	459.1
16 JAIPUR	15407	23.24	663
17 JAMSHEDPUR	2685	11.02	243.6
18 KANPUR	8885	26.9	330.3
19 KOCHI	7956	13.55	587.2
20 KOLKATA	13005	132.17	98.4
21 LUCKNOW	11735	22.67	517.6
22 LUDHIANA	2847	13.95	204.1
23 MADURAI	2470	11.95	206.7
24 MEERUT	2765	11.67	236.9
25 MUMBAI	32770	163.68	200.2
26 NAGPUR	8661	21.23	408
27 NASIK	3813	11.52	331
28 PATNA	9014	17.07	528.1
29 PUNE	14467	37.56	385.2
30 RAJKOT	5525	10.02	551.4
31 SURAT	10741	28.11	382.1
32 VADODARA	5386	14.92	361
33 VARANASI	2734	12.12	225.6
34 VIJAYAWADA	5127	10.11	507.1
35 VISHAKHAPATNAM	5015	13.29	377.4
<b>TOTAL (CITIES)</b>	<b>347153</b>	<b>1078.8</b>	<b>321.8</b>

NOTES – This table reproduces the publicly available 2008 crime rate statistics for 35 cities in India provided by the National Crime Records Bureau, part of the Ministry of Home Affairs. These data are available at <https://ncrb.gov.in/en/crime-in-india-table-additional-table-and-chapter-contents?page=1>.

**Table A14**  
**Origin Location Characteristics – 30% Cut-off**

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Average Performance (z- score) OLS w/FE	Average Performance (z- score) OLS w/FE	Quit to Competitor OLS w/FE	Quit to Competitor OLS w/FE	Average Performance (z- score) OLS w/FE	Average Performance (z- score) OLS w/FE	Quit to Competitor OLS w/FE	Quit to Competitor OLS w/FE
Overall Score Bottom 30%	-0.293** (0.101)	-0.300* (0.132)	0.022 (0.047)	-0.007 (0.041)				
Overall Score bottom 30% * Bangalore		0.033 (0.144)		0.147** (0.047)				
Crime Rate Top 30%					-0.007 (0.159)	-0.096 (0.174)	0.046* (0.024)	0.035 (0.027)
Crime Rate Top 30% * Bangalore						0.480** (0.191)		0.066** (0.025)
Logical Score	-0.015 (0.018)	-0.015 (0.018)	-0.000 (0.004)	-0.000 (0.004)	-0.009 (0.019)	-0.008 (0.020)	-0.001 (0.003)	-0.001 (0.003)
Verbal Score	0.010 (0.011)	0.011 (0.012)	-0.007** (0.003)	-0.007* (0.003)	0.007 (0.008)	0.008 (0.009)	-0.006* (0.003)	-0.005* (0.003)
CGPA Training	1.022*** (0.078)	1.022*** (0.079)	0.047 (0.030)	0.051 (0.032)	0.895*** (0.115)	0.905*** (0.110)	0.038 (0.032)	0.038 (0.032)
Male	0.286** (0.099)	0.286** (0.100)	0.023 (0.014)	0.022 (0.013)	0.145** (0.058)	0.158** (0.060)	0.001 (0.017)	0.003 (0.017)
Hometown	0.080 (0.091)	0.081 (0.091)	-0.022 (0.025)	-0.023 (0.025)	0.068 (0.100)	0.061 (0.101)	-0.023 (0.023)	-0.024 (0.023)
Constant	-4.757*** (0.325)	-4.758*** (0.327)	-0.122 (0.138)	-0.136 (0.142)	-4.212*** (0.465)	-4.270*** (0.458)	-0.070 (0.151)	-0.073 (0.149)
Observations	470	470	579	579	513	513	648	648
R-squared	0.155	0.155	0.038	0.046	0.112	0.118	0.027	0.029
Location FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTES – This table replicates the analyses in Table 9 in the main manuscript but replaces the binary variables capturing the top (bottom) 20% of the distribution for workers' origin town characteristics with an alternative classification based on a cut-off of 30%. Standard errors in parentheses are clustered at the production centre level.

\*p<0.1; \*\*p<.05; \*\*\*p<.01

**Table A15**  
**Origin Location Characteristics – 10% Cut-off**

VARIABLES	(1) Average Performance (z- score) OLS w/FE	(2) Average Performance (z- score) OLS w/FE	(3) Quit to Competitor OLS w/FE	(4) Quit to Competitor OLS w/FE	(5) Average Performance (z- score) OLS w/FE	(6) Average Performance (z- score) OLS w/FE	(7) Quit to Competitor OLS w/FE	(8) Quit to Competitor OLS w/FE
Overall Score Bottom 10%	-0.192** (0.067)	-0.221** (0.077)	0.034 (0.061)	-0.021 (0.025)				
Overall Score bottom 10% * Bangalore		0.133 (0.095)		0.274*** (0.036)				
Crime Rate Top 10%					-0.301*** (0.087)	-0.344** (0.107)	0.100 (0.084)	0.034 (0.059)
Crime Rate Top 10% * Bangalore						0.166 (0.105)		0.292*** (0.065)
Logical Score	-0.015 (0.017)	-0.016 (0.017)	-0.000 (0.004)	-0.001 (0.004)	-0.009 (0.019)	-0.009 (0.019)	-0.001 (0.003)	-0.001 (0.003)
Verbal Score	0.010 (0.012)	0.011 (0.012)	-0.007** (0.003)	-0.006* (0.003)	0.007 (0.009)	0.007 (0.009)	-0.006* (0.003)	-0.005* (0.003)
CGPA Training	1.040*** (0.077)	1.037*** (0.077)	0.048 (0.030)	0.048 (0.030)	0.895*** (0.111)	0.889*** (0.117)	0.038 (0.031)	0.030 (0.027)
Male	0.268** (0.093)	0.268** (0.093)	0.023 (0.014)	0.023 (0.014)	0.152* (0.068)	0.155* (0.068)	-0.005 (0.017)	-0.001 (0.018)
Hometown	0.128 (0.087)	0.128 (0.087)	-0.025 (0.024)	-0.027 (0.023)	0.059 (0.112)	0.059 (0.113)	-0.026 (0.024)	-0.027 (0.023)
Constant	-4.898*** (0.343)	-4.884*** (0.342)	-0.120 (0.139)	-0.123 (0.133)	-4.194*** (0.497)	-4.168*** (0.521)	-0.061 (0.149)	-0.029 (0.126)
Observations	470	470	579	579	513	513	648	648
R-squared	0.145	0.145	0.039	0.055	0.117	0.118	0.031	0.044
Location FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTES – Standard errors in parentheses are clustered at the production centre level. This table replicates the analyses in Table 9 in the main manuscript but replaces the binary variables capturing the top (bottom) 20% of the distribution for workers' origin town characteristics with an alternative classification based on a cut-off of 10%.

\*p<0.1; \*\*p<.05; \*\*\*p<.01



**Table A17**  
**Time Spent on R&D – Matched Sample Analysis – Coarsened Exact Matching**

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	% Time on R&D OLS CEM Automatic Bins	% Time on R&D OLS CEM Automatic Bins	% Time on R&D OLS CEM Manual Natural Bins	% Time on R&D OLS CEM Manual Natural Bins	% Time on R&D OLS CEM Manual Decile Bins	% Time on R&D OLS CEM Manual Deciles Bins
From Smaller Town	0.405 (1.188)	-0.772* (0.402)	1.547 (1.349)	-0.056 (0.532)	4.116 (3.487)	0.228 (0.221)
Placed in Bangalore	4.520*** (0.681)	1.886** (0.759)	7.791*** (0.374)	4.071*** (0.368)	9.009*** (0.631)	0.603 (0.412)
From Smaller Town * Bangalore		7.682*** (0.742)		11.216*** (0.487)		24.828*** (1.157)
Logical Score	0.116* (0.059)	0.106 (0.057)	-0.054 (0.125)	-0.103 (0.157)	0.142 (0.188)	0.113 (0.140)
Verbal Score	0.369 (0.377)	0.332 (0.352)	0.539 (0.394)	0.541 (0.393)	0.423 (0.281)	0.146 (0.101)
CGPA Training	2.543 (2.944)	2.396 (2.812)	4.201 (3.492)	4.241 (3.554)	2.336 (1.568)	2.422 (1.574)
Male	0.212 (1.399)	0.454 (1.518)	1.661 (1.904)	2.174 (2.294)	1.332 (1.662)	0.997 (1.296)
Placed in Hometown	-2.365 (1.327)	-1.954 (1.059)	-2.654 (2.022)	-2.010 (1.521)	-3.395 (2.291)	-1.559 (1.137)
Constant	-12.337 (15.652)	-11.280 (14.687)	-22.106 (18.528)	-21.855 (18.305)	-15.144 (10.900)	-12.408 (8.837)
Observations	303	303	195	195	103	103
R-squared	0.056	0.077	0.143	0.176	0.175	0.329
Location FE	No	No	No	No	No	No

NOTES – Standard errors in parentheses are clustered at the location level except for Columns 7 and 8 where the errors are robust. Three versions of Coarsened Exact Matching (CEM) are presented, all of which match workers on their pre-location assignment characteristics of Logical Scores, Verbal Scores, CGPA Training and Gender. However, the results are also robust to matching on Logical and Verbal Scores only. The first in Columns 1 and 2 labelled Automatic, uses STATA’s in-built identification of appropriate bins using Sturge’s Rule (i.e., the number of bins is equal to  $1+3.322\log(n)$  where  $n$  is the number of observations), and results in a multivariate L1 distance equal to 0.526 with 81 matched strata comprising of 118 treated and 185 untreated observations. In the second approach in Columns 3 and 4 we use is a set of natural bins created based on the common Logical and Verbal scores achieved on the pre-entry exams and allowing STATA to select appropriate bins for CGPA training based on Sturge’s rule. This approach leads to an improved multivariate L1 distance equal to 0.433, but fewer matched strata – 71 – and fewer treated (89) and untreated (106) observations. Finally, the third, and most conservative approach in Columns 5 and 6 creates bins based on deciles within Logical, Verbal and CGPA Training scores. This approach leads to the lowest multivariate L1 distance of 0.227, but is also the most conservative of the three, with only 41 matched strata containing 47 treated and 56 untreated observations. Columns 7 and 8 present our results with Propensity Score Matching with the nearest neighbour algorithm with a single nearest neighbour, using the same pre-location assignment characteristics of Logical Scores, Verbal Scores, CGPA Training and Gender for the initial matching. However, note that since we are interested in the difference in the Bangalore placement effect between workers from smaller towns and large cities we perform the matching separately for workers from smaller towns and large cities and present the results for the treatment dummy on From Smaller Town. All observations end up on common support after this procedure. In unreported results we also confirm that balance is achieved on all variables in the matching process.

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

**Table A18**  
**Time Spent on R&D by Workers Located in Chennai, Hyderabad, and Pune**

VARIABLES	(1) % Time on R&D OLS <i>Large City Sample</i>	(2) % Time on R&D OLS <i>Smaller Town Sample</i>	(3) % Time on R&D OLS <i>Large City Sample</i>	(4) % Time on R&D OLS <i>Smaller Town Sample</i>	(5) % Time on R&D OLS <i>Large City Sample</i>	(6) % Time on R&D OLS <i>Smaller Town Sample</i>
Placed in Chennai	-1.071** (0.344)	-1.191 (0.850)				
Placed in Hyderabad			1.110* (0.498)	-1.056 (1.326)		
Placed in Pune					-0.356 (0.552)	-1.024 (1.383)
Logical Score	-0.047 (0.207)	0.256 (0.201)	-0.037 (0.210)	0.251 (0.201)	-0.050 (0.207)	0.245 (0.194)
Verbal Score	-0.043 (0.132)	0.299 (0.194)	-0.042 (0.133)	0.304 (0.198)	-0.042 (0.133)	0.314 (0.210)
CGPA Training	1.413 (0.976)	1.131 (1.048)	1.440 (0.980)	1.197 (1.100)	1.502 (0.997)	1.245 (1.163)
Male	-0.926 (0.713)	1.177 (1.362)	-0.966 (0.715)	1.302 (1.477)	-0.919 (0.704)	1.281 (1.457)
Placed in Hometown	-0.880 (0.532)	-1.168 (0.795)	-0.870 (0.509)	-1.158 (0.779)	-0.875 (0.531)	-1.106 (0.697)
Constant	-3.451 (3.834)	-6.817 (6.201)	-3.941 (3.715)	-7.144 (6.421)	-3.890 (3.809)	-7.342 (6.684)
Observations	474	181	474	181	474	181
R-squared	0.011	0.049	0.012	0.050	0.009	0.050
Location FE	No	No	No	No	No	No

NOTES – Parentheses contain standard errors clustered at the production centre level. This table replicates the analyses in Table 5 in the main manuscript for each of the next three largest technology hubs where INDTECH has a production center: Chennai, Hyderabad and Pune.

\*p<0.1; \*\*p<.05; \*\*\*p<.01

**Table A19**  
**Worker Enrolment in Additional Coursework in Chennai, Hyderabad, and Pune**

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Number of Courses OLS with FEs	% Courses Passed OLS with FEs	Number of English Courses OLS with FEs	Level of English Courses OLS with FEs	Number of Courses OLS with FEs	% Courses Passed OLS with FEs	Number of English Courses OLS with FEs	Level of English Courses OLS with FEs	Number of Courses OLS with FEs	% Courses Passed OLS with FEs	Number of English Courses OLS with FEs	Level of English Courses OLS with FEs
From Smaller Town	0.771** (0.301)	1.223 (0.933)	-0.051 (0.031)	0.386 (0.256)	0.960*** (0.274)	2.434* (1.153)	-0.048 (0.030)	0.548** (0.199)	0.631*** (0.182)	2.649* (1.165)	-0.047 (0.032)	0.502** (0.212)
From Smaller Town * Chennai	0.555* (0.261)	5.365*** (1.013)	0.029 (0.030)	0.347 (0.326)								
From Smaller Town * Hyderabad					-0.795** (0.278)	-2.562* (1.256)	0.007 (0.031)	-0.737** (0.308)				
From Smaller Town * Pune									1.424*** (0.196)	-3.102** (1.257)	0.001 (0.034)	-0.196 (0.185)
Logical Score	0.043 (0.035)	-0.261 (0.347)	0.001 (0.005)	-0.035 (0.045)	0.044 (0.035)	-0.254 (0.347)	0.001 (0.005)	-0.034 (0.045)	0.043 (0.035)	-0.255 (0.348)	0.001 (0.005)	-0.035 (0.044)
Verbal Score	-0.076*** (0.019)	0.289** (0.123)	-0.005 (0.003)	0.086** (0.030)	-0.075*** (0.019)	0.290** (0.126)	-0.005 (0.003)	0.084** (0.029)	-0.078*** (0.018)	0.296** (0.121)	-0.005 (0.003)	0.084** (0.029)
CGPA Training	0.626 (0.446)	16.695*** (3.650)	-0.041 (0.040)	-0.984 (0.530)	0.625 (0.443)	16.647*** (3.608)	-0.041 (0.040)	-1.055* (0.517)	0.653 (0.444)	16.498*** (3.758)	-0.041 (0.040)	-1.024* (0.528)
Male	0.245 (0.141)	-4.289** (1.795)	0.011 (0.031)	-0.429 (0.254)	0.240 (0.143)	-4.311** (1.758)	0.011 (0.031)	-0.423 (0.245)	0.231 (0.137)	-4.313** (1.801)	0.011 (0.031)	-0.451 (0.246)
Placed in Hometown	0.114 (0.230)	-0.910 (1.948)	0.006 (0.031)	-0.272 (0.240)	0.106 (0.233)	-0.912 (1.927)	0.006 (0.031)	-0.265 (0.251)	0.088 (0.227)	-0.885 (1.936)	0.006 (0.031)	-0.241 (0.243)
Constant	0.232 (1.949)	7.387 (15.692)	0.353* (0.183)	6.249** (2.321)	0.234 (1.952)	7.561 (15.682)	0.354* (0.183)	6.562** (2.235)	0.148 (1.960)	8.211 (16.116)	0.354* (0.185)	6.439** (2.289)
Observations	1,208	687	687	96	1,208	687	687	96	1,208	687	687	96
R-squared	0.049	0.122	0.019	0.258	0.050	0.120	0.019	0.262	0.054	0.121	0.019	0.256
Location FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTES – This table replicates the analyses in Table 6 in the main manuscript for each of the next three largest technology hubs where INDTECH has a production center: Chennai, Hyderabad, and Pune. Standard errors in parentheses are clustered at the production centre level.

\*p<0.1; \*\*p<.05; \*\*\*p<.01

## References

Bejan, Adrian. "Why university rankings do not change: education as a natural hierarchical flow architecture." *International Journal of Design & Nature and Ecodynamics* 2.4 (2007): 319-327.

## Appendix B

### Additional Evidence for our Interpretation of the “Smaller Town” Construct

We scrape data from publicly available profiles of employees for the three largest Indian technology firms (TCS, Infosys, and Wipro), available on the social networking site LinkedIn. We reconcile differences in naming conventions for each of these firms on the LinkedIn profiles (e.g., Wipro was listed as “*Wipro*,” “*Wipro technologies*,” “*Wipro Ltd.*,” etc.) and clean and code the graduate school of each employee for 30,131 employees in this sample. We follow the same coding principle as in our definition of the *From smaller town* variable and code the graduate school as being located within a top six metropolitan city in India (i.e., New Delhi, Mumbai, Kolkata, Chennai, Bangalore, and Hyderabad) or elsewhere (i.e., in smaller towns). The fraction of employees who graduated from colleges located in smaller towns was 0.59, 0.62, and 0.62, respectively, for employees from TCS, Infosys, and Wipro.

To follow up, we conduct another survey at one of the top three Indian technology firms (not INDTECH). We hired a professional survey company to survey 1,054 employees at the firm’s Bangalore production center. 84 percent of employees surveyed indicated they had gone to school in a smaller town in India. We also find that employees from smaller towns had shorter tenures at the firm (difference in means = -1.12 years,  $t=-4.11$ ) and were more likely to be employed as a contractor (difference in means = 0.22,  $t=4.99$ ).

In the final step, we collected the CVs of the top 593 scientists in the government-owned research labs of India. India’s 42 state-owned national laboratories are organized under an autonomous umbrella organization, The Council of Scientific and Industrial Research (CSIR); collectively they have around 12,500 scientific and technical employees. The laboratories, covering all major scientific and engineering disciplines, were created in the 1940s and 1950s. We collected the CVs of the top 593 scientists across all 42 laboratories at the top three tiers of the organizational hierarchy. We then coded whether or not the individual was educated in a college located in a smaller town. The mean fraction of top CSIR scientists educated in smaller towns is 0.74, 0.71, and 0.57, respectively, for the top three organizational hierarchy levels of ‘Scientist-F,’ ‘Scientist-G,’ and ‘Scientist-H.’<sup>2</sup>

---

<sup>2</sup> Our secondary analyses benefit from looking at employees across the organizational hierarchy of these firms (by contrast, our main analyses look at entry-level employees at INDTECH) but have several limitations. For this broader sample of employees, we are unable to compare the performance of individuals educated in smaller towns vs. larger cities. We also do not know whether the Indian firms hired employees educated in smaller towns directly from their colleges or whether the individuals migrated to larger cities prior to being hired by these firms.

**Table B1**  
**Survey of Large City and Smaller Town Engineering Colleges**

	Large city colleges	Smaller town colleges
Average size of graduating class in computer science/IT (undergraduate and master's)	342	458
Average percentage of graduating class in computer science/IT hired by INDTECH (in 2011, 2012)	0.17%	0.06%
Average percentage of graduating class in computer science/IT hired by multinational technology firms IBM and Cognizant (in 2011, 2012)	9%	1%
Mean annual salary (Rupees Lakhs, 2011 and 2012 average)	6.20	2.70
N	7	4

NOTES – The researchers randomly selected 10 large city and 10 smaller town engineering colleges from the list of colleges from which INDTECH hires and contacted the colleges' representatives to ask them to participate in a telephone survey. The researchers were able to conduct interviews with representatives at seven out of the 10 large city colleges. These were the R.V. College of Engineering, Bangalore; M.S. Ramaiah Institute of Technology, Bangalore; MLR Institute of Technology, Hyderabad; Muffakham Jah College of Engineering and Technology, Hyderabad; Vasavi College of Engineering, Hyderabad; G. Narayanamma Institute of Technology & Science (GNITS), Hyderabad; and Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad. The researchers were also able to conduct interviews with representatives of four out of the 10 selected smaller town colleges. These included M.J.P. Rohilkhand University in Bareilly, Uttar Pradesh; Majhighariani Institute of Technology & Science, Rayagada Orissa; Bapatla Engineering College, Guntur, Andhra Pradesh; and Jaya Prakash Narayan College of Engineering, Mahabubnagar, Dharmapur, Telangana. The survey results indicated that the mean salaries for 2011 and 2012 were significantly higher for individuals hired from the large city colleges as compared to those from the smaller town colleges. We found this difference to be statistically significant, based on a t-test comparison of means. In addition, the survey revealed that multinational technology firms predominantly hire from large city colleges, while INDTECH follows the distinctive policy of hiring from both large city and smaller town colleges. Note that the results from the survey might be upward biased, given the small sample of colleges that participated in the survey. Rs.1 Lakh = Rs. 100,000

## Appendix C

### Employee Random Assignment Protocol

This text is based on field interviews and INDTECH internal documents. (Part of it is copied from INDTECH internal documents.)

INDTECH assigns its software engineer trainees to production centres based on a computer application called “Talent Planning,” which is part of the firm’s enterprise resource allocation software system. This application allocates trainees to a location based on the quarterly manpower budget released by Corporate Planning.

The “process life cycle steps” are:

- Collating the manpower budget and unit-wise requirements
- Trainee assignment (location)
- Communication with stakeholders

Talent Planning bases the assignment of employees on the following:

- Production center requirements: HR at each production center provides data on requirement for trainees trained in various technologies.
- Data from HR stationed at the training location: Two weeks prior to the completion of training batches, HR at the training location releases data on which employees are expected to complete training.

The two variables that the Talent Planning team considers while assigning employees to production centres, using the automated system, include the stream of training for the trainee and the estimated date of training completion. The prior background of the employee and the test scores of the employee are **not** considered in this decision. INDTECH communicates employee assignments through a centralized portal.

## **Appendix D**

### **Human Capital Rents by Worker Origin and Placement Location**

Our results raise the question of whether the decision to hire employees from smaller towns and place them in Bangalore rather than an alternative location creates economic value for INDTECH. In particular, on net, do workers from smaller towns who are placed in Bangalore create less value than their large city counterparts when we account for their higher mobility to competing firms? Similarly, what is the relative net return on placing workers from smaller towns and large cities outside of Bangalore? Unfortunately, answering this question requires detailed employee-level data comparing INDTECH's costs associated with hiring individuals from smaller towns versus large cities and the monetary returns to their employment, which we do not have access to. As a result, we could not run regressions using hiring costs. However, based on a set of assumptions gathered during field interviews, we are able to estimate at least the approximate net payoff of hiring individuals from smaller towns and large cities and placing them in different locations in their first year of employment at INDTECH. We outline here the steps in our estimation process and the results.

In the first step, we estimate the “revenues” generated by different types of workers. To do so, we begin by estimating the dollar value of productivity gains associated with hiring a worker from a smaller town and also of placing this individual in Bangalore. Our interviews suggest that at least in their first year on the job, as compared to those who achieve the highest performance rating in a given year, other workers need 35% more man-days to correct coding/testing/documentation errors. This is based on rough calculations with INDTECH HR managers on error rates and lost man-days due to coding/testing/documentation errors, and implies that workers achieving the top performance rating can complete the same amount of work as those receiving lower performance ratings in just under three quarters (0.741) the amount of time as lower-performing workers. We use the predicted probabilities of achieving the highest performance rating in 2008 for small-town versus large-city workers reported in Column (1) of Table A3 in the Appendix. Average marginal effects (estimated with the STATA command “margins i.from\_smaller\_town”) indicate that small-town workers receive the highest performance rating at a rate of 40.9 percent, while workers hired from large cities do so at a rate of 33.3 percent.

To arrive at a dollar-value estimate of the total value generated by each type of worker, however, we need to estimate each worker's contribution to firm revenues. While direct measures of the revenue

contributions of workers in our sample are not available, we are able to obtain an estimate of the likely upper and lower bounds for the average revenues that a worker at INDTECH generates. We obtain the upper bound by dividing publicly available data on INDTECH's total firm revenues in 2008 by the total number of workers at INDTECH in 2008. This calculation yields an average revenue of about \$50,000 per worker. Our reasoning is that while the workers in our sample work on the organization's core products and services, and are therefore likely to generate more revenue than workers in support functions at the firm, they are nonetheless entry-level workers and therefore unlikely to be contributing at a level above the mean revenue per worker within their first year on the job. We therefore use the average revenue of \$50,000 as our likely upper bound on the worker's productivity. We estimate the lower bound of productivity conservatively as the sum of the worker's annual wages and training costs, discussed below. However, please note, that while these figures ignore the substantial variation across workers and job types, our calculations of the relative rents from different types of workers are not sensitive (in relative rank) to the absolute value of average revenues chosen. We start by calculating the net returns using only the upper bound and then discuss the changes in the results with the lower bound.

Combining the relative performance differences across workers and the average upper bound of revenue contributions by all workers at INDTECH, we arrive at the following formula to calculate the total contribution to revenues by each worker type:

$$\text{Total Contribution to Firm Revenues} = \text{Probability of Achieving Top Performance Rating} * \text{Extra Revenue From Top Performers} * \$50,000 + (1 - \text{Probability of Achieving Top Performance Rating}) * \$50,000$$

Applying this formula to workers hired from smaller towns, we calculate that their total value generated for INDTECH in their first year of employment amounts to \$57,158 ( $=0.409*1.35*50,000+0.591*50,000$ ) and the total value generated by workers from large cities amounts to \$55,828 using the same formula. However, since we know that the *average* employee at INDTECH generates about \$50,000 in value, we rescale these figures to preserve this average for all workers, and arrive at the final value generated by smaller town workers of \$50,756 and large city workers of \$49,575 (the rescaling is achieved by assuming an equal number of workers from smaller towns and large cities and setting their average contributions to equal \$50,000; the resulting scaling factor is equal to 0.888, and we multiple this factor by each of the workers' contributions).

However, these figures are the averages for workers hired from smaller towns and large cities, regardless of placement location. If these workers are instead posted to headquarters in Bangalore, the relative performance differences grow between large city and smaller town workers, to 25.5% and 39.1% chances of receiving the top performance rating, respectively. Therefore, the relative re-scaled productivities of employees from smaller towns and large cities become \$50,920 and \$48,363, respectively.

In the second step, we estimate the costs of recruiting workers of different types. INDTECH's entry-level salaries are about \$5,870 per year (at 2008 U.S. Dollar to Rupee exchange rates) regardless of placement and worker productivity. Therefore, workers from smaller towns and large cities all receive the same salary of \$5,870. However, recruiting workers from smaller towns requires additional expenditures. Based on our discussions with INDTECH's recruiting managers, we estimate that there is a \$21 incremental cost of hiring a remote worker. This is based on several criteria: incremental travel costs for INDTECH executives involved in hiring from smaller towns, the additional search costs associated with trips to screen colleges and students from smaller towns, and the larger number of candidates who need to be interviewed in smaller towns as compared to large cities.

In addition, differences in attrition rates across different types of workers and placement locations create significant differences in the cost of training, turnover, and replacement. Specifically, INDTECH spends about \$3,500 to train each newly hired worker in the four-month training program. Since all hires go through this training, regardless of origin, the figure only enters our calculations through attrition rates. We therefore compare the overall attrition rates for workers from smaller towns and large cities, as well as their differential rates by placement location, using Columns (10) and (11) in Table A3. Average marginal analyses indicate that workers from smaller towns exit INDTECH at an average rate of 7.8% per year relative to workers from large cities, who do so at a rate of 13.0% per year.<sup>3</sup> This difference is reduced but remains large even when both types of workers are posted to Bangalore (9.7% annual attrition rate for workers from smaller towns and 14.6% annual attrition rate for workers from large cities). Multiplying these figures by the training (\$3,500) and recruitment (\$21) costs of smaller town workers yields a total

---

<sup>3</sup> Note that the attrition results in Table A3 are based on worker attrition after three years at INDTECH from the date of entry. We annualize these figures by assuming a constant rate of attrition each year, using the size of the entry cohort as the base with respect to which we calculate rates of departure. The annualized rates are therefore one-third of the total rates for the three years for each group – 39.1% for large city workers and 23.4% for workers from smaller towns.

recruitment, training, and replacement cost of \$275 each for workers from smaller towns and \$455 each for workers from large cities, regardless of placement location. When posted to Bangalore, these figures grow to \$342 and \$511 for smaller town and large city workers, respectively.

In sum, the expected costs of employing workers from smaller towns regardless of location amount to \$6,145 (\$5,870 in salary and \$275 in recruitment, training, and replacement costs) and employing workers from large cities amounts to \$6,325. In Bangalore, the respective estimated costs are \$6,212 and \$6,381 for smaller town and large city workers in their first year of employment.

Bringing together the estimates of revenue and costs for each worker type, we arrive at our back-of-the-envelope estimates for human capital rents captured by INDTECH each year from workers in our sample. Depicted as the “rent rectangles” in Figure D1, they show that workers from smaller towns placed outside of Bangalore generate the greatest amount of rents (\$44,724), followed closely by workers from smaller towns placed in Bangalore (\$44,708). The lowest rents derive from workers from large cities posted to Bangalore, whose higher rates of attrition and lower productivity generate only \$41,982 in rents, a difference of \$2,742 as compared to workers from smaller towns posted outside of Bangalore, which amounts to about 46.7% of the workers’ annual salary in 2008 of \$5,870.

Similarly, we calculate the lower bound of productivity using the same formula as above, but instead of scaling the returns by the average worker revenue of \$50,000, we scale it by the average lower bound of revenue in the worker’s first year on the job – the sum of the worker’s annual salary (\$5,870) and training costs (\$3,500) as the first year revenue, or \$9,370. Applying the same formula, we arrive at:

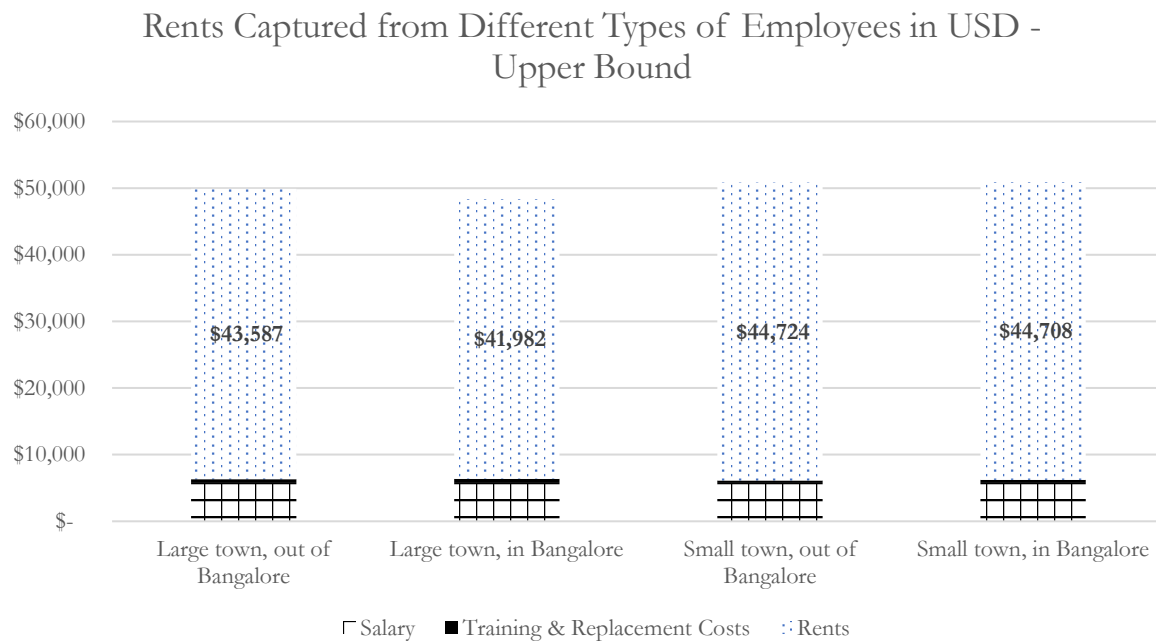
$$\text{Total Contribution to Firm Revenues} = \text{Probability of Achieving Top Performance Rating} * \text{Extra Revenue From Top Performers} * \$9,370 + (1 - \text{Probability of Achieving Top Performance Rating}) * \$9,370$$

For workers hired from smaller towns and placed in Bangalore, this calculation amounts to \$10,652 ( $=0.391*1.35*9,370+0.609*9,370$ ) and the total value generated by workers from large cities placed in Bangalore amounts to \$10,206 using the same formula. Rescaling these values to preserve the average yields \$9,459 and \$9,063, respectively. For workers placed outside of Bangalore, these figures are \$9,351 for workers from large cities and \$9,529 for workers from smaller towns. Subtracting from these values the average replacement and re-training costs for workers placed in Bangalore from smaller towns (\$340) and large cities (\$510), and outside of Bangalore from smaller towns (\$257) and large cities (\$443) as well as the

\$5,870 in wages, we arrive at net figures of \$3,249 for workers from smaller towns and \$2,683 for workers from large cities, placed in Bangalore. For large city workers outside of Bangalore, this figure is \$3,038 and for small town workers outside of Bangalore this figure is \$3,402. Therefore, as depicted in Figure D2, the difference between large city workers in Bangalore and small town workers in Bangalore amounts to \$719 per worker per year or 12.2% of annual worker salary.

However, given our data constraints, these calculations provide only *a very rough estimate* of the net payoff associated with hiring from smaller towns and have at least two important limitations. First, we do not have an estimate of sunk costs of investments that INDTECH made related to hiring from smaller towns. Nor do we have an estimate of operating costs per worker related to real estate, utilities, etc. Second, we cannot distinguish between the costs of losing workers to voluntary turnover for pursuing higher studies versus turnover to competing firms. The latter may have important observed and unobserved costs that affect the trade-off as estimated. For instance, additional (unreported) results indicate that top-performing workers from smaller towns are even more likely to leave to join competing firms and less likely to leave to take up further education when posted to Bangalore as compared to elsewhere and as compared to their large city counterparts, suggesting that such turnover may be even more costly than estimated here. However, while our data limit more accurate calculations, we hope that our approach will spur future research on the role of worker characteristics, not only in terms of value creation and capture, but also net rent generation for firms, since the latter is likely to be the true underlying driver of recruitment decisions.

**Figure D1**  
**‘Rent Rectangles’: INDTECH’s Value Creation and Appropriation by Employee Origin and Placement Location (in U.S. dollars) – Upper Bound**



**Figure D2**  
**‘Rent Rectangles’: INDTECH’s Value Creation and Appropriation by Employee Origin and Placement Location (in U.S. dollars) – Lower Bound**

